Charles University in Prague
Faculty of Mathematics and Physics

Habilitation Thesis

Spatiotemporal links and variability in the climate system: A regression analysis perspective

Jiří Mikšovský

2015

ACKNOWLEDGEMENTS

The results presented here could not have been achieved without the support of a large number of people and organizations. First, I would like to express my gratitude to all my collaborators, named in the author lists and acknowledgements of individual papers related to the topics presented. I am also deeply grateful to my co-workers at the Department of Atmospheric Physics (formerly Department of Meteorology and Environment Protection) of Faculty of Mathematics and Physics, Charles University in Prague for providing a pleasant and stimulating work environment for over a decade now. Financial support was granted by various institutions: In particular, I would like to thank the Grant Agency of Charles University (project 227/2002/B-GEO/MFF), Czech Science Foundation (grants 205/06/P181, P209/11/2405 and P209/11/0956), Ministry of Environment of the Czech Republic (project VaV/740/2/03), Ministry of Education of the Czech Republic (research plan MSM0021620860) and the European Commission (6th Framework Programme, project CECILIA). Finally, considering the nature of the climate research in general, and of time series-based studies in particular, our work would not have been possible without the effort of the many authors and providers of various datasets employed here and in the related contributions.

CONTENTS

CHAPTER 1

INTRODUCTION

Earth's climate system consists of a multitude of diverse components, active on a range of temporal and spatial scales, interrelated and subjected to external influences from the planetary interior or outer space as well as to the effects of human activity. The intricacy of the resulting structure marks it as one of the most challenging targets for study in – and beyond – the field of physics, and no current scientific technique is able to provide its complete, accurate description. Even so, much understanding about weather and climate can be gained through their simplified representations. Since analytical solutions do exist for only the most minimalistic embodiments of the related dynamics, numerical simulations have become the prime research tools in meteorology and climatology. Nevertheless, even the most sophisticated state-of-the-art models still fail to deliver a completely realistic reproduction of the climate system or its individual components. This applies not only to the prognostic simulations, limited in their ability to reliably forecast weather by the inherently chaotic nature of the atmosphere, but also to their climatic counterparts, struggling to provide a fully satisfactory approximation of the complex weave of the processes forming the Earth's climate. Consequently, many of the real-world features are misrepresented or absent in the simulated climates, or captured with substantial uncertainty. As illustrated (for instance) by the summary assessment by the Intergovernmental Panel on Climate Change (STOCKER ET AL. 2013), steady improvement of the performance of the climate models has been achieved over the past years, gradually alleviating many of their imperfections. Yet, even in their current advanced state, numerical simulations do still not offer a completely dependable picture of the climate and other approaches are needed to support, complement and validate them. This role is filled in a large part by statistical methods, ranging from basic descriptive and exploratory techniques to complex nonlinear algorithms for investigation of the variability patterns in multidimensional data.

A substantial part of the knowledge about the climate system comes from the study of its direct or indirect manifestations, recorded in the form of univariate or multivariate time series. The main role of statistical techniques then consists in extraction, refinement and interpretation of the information contained in such signals. Obviously, this brief thesis does not attempt to provide a full treatise of the extensive array of statistical methods used in the climatic research, or to deliver a comprehensive synopsis of their numerous applications to the observed and simulated data. Rather, it aims to highlight several specific topics pertaining to my past research in the field of statistical climatology, to deliver selected examples of the related results, and to connect them in a unifying frame.

The thesis has been created as summary, amalgamation and evolution of materials published in selected works authored or co-authored by me during my research career. Its core is built upon seven stand-alone publications with my major participation, provided in the appendices and dealing in a large part (though not exclusively) with various applications of regression mappings in the atmospheric and climatic research:

- **MIKŠOVSKÝ & RAIDL (2006)** → **(Appendix I, p. 46)**
  MIKŠOVSKÝ, J., AND A. RAIDL (2006), Testing for nonlinearity in European climatic time series by the method of surrogate data, *Theoretical and Applied Climatology*, *83*(1-4), 21-33, doi:10.1007/s00704-005-0130-7.
- **MIKŠOVSKÝ ET AL. (2008)** → **(Appendix II, p. 60)**
  MIKŠOVSKÝ, J., P. PIŠOFT, AND A. RAIDL (2008), Global Patterns of Nonlinearity in Real and GCM-Simulated Atmospheric Data, in *Nonlinear Time Series Analysis in the Geosciences: Applications in Climatology, Geodynamics and Solar-Terrestrial Physics* (Eds.: Donner, R. V., and S. M. Barbosa), *Lecture Notes in Earth Sciences*, *112*, 17-34, doi:10.1007/978-3-540-78938-3_2.
- **MIKŠOVSKÝ & RAIDL (2005)** → **(Appendix III, p. 79)**
  MIKŠOVSKÝ, J., AND A. RAIDL (2005), Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures, *Nonlinear Processes in Geophysics*, *12*(6), 979-991.
- **HUTH ET AL. (2015)** → **(Appendix IV, p. 93)**
  HUTH, R., J. MIKŠOVSKÝ, P. ŠTĚPÁNEK, M. BELDA, A. FARDA, Z. CHLÁDOVÁ, AND P. PIŠOFT (2015), Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature, *Theoretical and Applied Climatology*, *120*(3-4), 533-553, doi:10.1007/s00704-014-1190-3.
- **KRIŽAN ET AL. (2011)** → **(Appendix V, p. 115)**
  KRIŽAN, P., J. MIKŠOVSKÝ, M. KOZUBEK, W. GENGCHEN, AND B. JIANHUI (2011), Long term variability of total ozone yearly minima and maxima in the latitudinal belt from 20°N to 60°N derived from the merged satellite data in the period 1979-2008, *Advances in Space Research*, *48*(12), 2016-2022, doi:10.1016/j.asr.2011.07.010.
- **MIKŠOVSKÝ ET AL. (2014)** → **(Appendix VI, p. 123)**
  MIKŠOVSKÝ, J., R. BRÁZDIL, P. ŠTĚPÁNEK, P. ZAHRADNÍČEK, AND P. PIŠOFT (2014), Long-term variability of temperature and precipitation in the Czech Lands: an attribution analysis, *Climatic Change*, *125*(2), 253-264, doi:10.1007/s10584-014-1147-7.
- **BRÁZDIL ET AL. (2015B)** → **(Appendix VII, p. 136)**
  BRÁZDIL, R., M. TRNKA, J. MIKŠOVSKÝ, L. ŘEZNÍČKOVÁ, AND P. DOBROVOLNÝ (2015B), Spring-summer droughts in the Czech Land in 1805-2012 and their forcings, *International Journal of Climatology*, *35*, 1405-1421, doi:10.1002/joc.4065.

Additional materials have also been adapted from the following publications, not enclosed within the thesis:

- **BRÁZDIL ET AL. (2012A)**
  BRÁZDIL, R., M. BĚLÍNOVÁ, P. DOBROVOLNÝ, J. MIKŠOVSKÝ, P. PIŠOFT, L. ŘEZNÍČKOVÁ, P. ŠTĚPÁNEK, H. VALÁŠEK, AND P. ZAHRADNÍČEK (2012A), *Temperature and precipitation fluctuations in the Czech Lands during the instrumental period*, Masaryk University, Brno, 236 pp., ISBN 978-80-210-6052-4.
- **MIKŠOVSKÝ & PIŠOFT (2015)**
  MIKŠOVSKÝ, J., AND P. PIŠOFT (2015), Attribution of European temperature variability during 1882-2010: A statistical perspective, in *Global Change: A Complex Challenge* (Eds.: Urban, O., M. Šprtová, and K. Klem), Global Change Research Centre AS CR, Brno, 10-13, ISBN 978-80-87902-10-3 (in print).

Finally, to provide a more complete picture of some of the issues discussed, selected elements of yet unpublished analyses or those currently under preparation were also included (and they are designated as such in the text). To facilitate identification of the materials with my direct contribution (and with my explicit authorship or co-authorship), the respective references are followed by a superscript asterisk (*) in the rest of the text. Unless stated otherwise, my contribution to these publications was predominant regarding the primary focus of this thesis, i.e. implementation of the regression models and their application to the individual problems presented throughout this text.

While the topics covered here vary substantially in terms of methods employed, datasets examined, and even the overall purpose of the particular analyses, some joint themes can be highlighted. Besides the general subject of spatiotemporal relationships, and application of regression mappings for their characterization, the motif of manifestations of nonlinearity in the climatic data is particularly pervasive in my past research, from attempts to quantify the magnitude of nonlinear behavior in the univariate and multivariate series (MIKŠOVSKÝ & RAIDL 2005*, 2006*; MIKŠOVSKÝ ET AL. 2008*), to use of nonlinear functions for downscaling of large-scale data (MIKŠOVSKÝ & RAIDL 2005*; HUTH ET AL. 2015*) or application of regression models connecting the observed variability to various climate forcings (BRÁZDIL ET AL. 2012A*; MIKŠOVSKÝ ET AL. 2014*). The issue of attribution also permeates through much of my past work, whether focused on identification of the factors shaping the temporal variability of basic climatic variables such as temperature (BRÁZDIL ET AL. 2012A*; MIKŠOVSKÝ ET AL. 2014*; MIKŠOVSKÝ & PIŠOFT 2015*), assessment of temporal trends in the ozone series (KRIŽAN ET AL. 2011*), or imprints of climate forcings in drought indices (BRÁZDIL ET AL. 2015B*).

Despite the obvious topical diversity of the problems addressed here, and the resulting specificity of the conclusions reached, there are some general lessons to be learned from the results obtained. This unifying commentary is therefore not ordered by individual publications. Instead, the text is structured into

several topically focused (though still partly overlapping and interrelated) segments. Chapter 2 briefly illustrates the datasets employed to characterize the climate system, its dynamics and evolution. Chapter 3 shows selected representatives of linear and nonlinear regression mappings, as the primary methodological common point of the publications assembled within the thesis. The subsequent sections then summarize specific results pertaining to the three main categories of problems tackled here: Chapter 4 explores the manifestations of nonlinear behavior related to short-term prediction of atmospheric variables; Chapter 5 is devoted to description of spatial relationships within and between different datasets, with particular focus on the issues of temperature downscaling (Chap. 5.1) and an additional example demonstrating approximation of temperature data from other concurrently measured records (Chap. 5.2); Chapter 6 concentrates on assessment of trends in total ozone data (Chap. 6.1) and statistical attribution analyses targeting various temperature and precipitation series (Chap. 6.2) and series of drought indices (Chap. 6.3). Finally, summarizing and concluding remarks are provided in Chapter 7, along with the prospects of the related ongoing and future research by me and my collaborators.

CHAPTER 2
CLIMATIC DATA: OBSERVATIONS & SIMULATIONS

Various measured and simulated time series are the key source of information about the climate system and its evolution, but their origins and properties do vary substantially. To illustrate the range of datasets used in our past research, some of the prominent classes of climatic data are introduced in this section, and a brief mention is given to their specific representatives employed in the studies discussed in Chapters 4-6 (see individual papers for a more comprehensive overview of the data and additional details).

The basic – and most traditional – form of climatic records comes from the measurements taken at land-based stations, often established specifically for weather observations. The resulting series of meteorological variables such as temperature, precipitation totals or air pressure can span several decades, with the longest of them covering multiple centuries. Length of these signals makes them a valuable source for examining the climate variability at various time scales. On the other hand, records of this extent are also prone to presence of non-climatic breaks and inhomogeneities and they are often in need of quality control and homogenization (e.g. BRÁZDIL ET AL. 2012B). In the contributions within this thesis, numerous series of daily temperature and pressure from Czech weather stations were used, obtained from the observational network maintained by the Czech Hydrometeorological Institute (CHMI - http://www.chmi.cz/). Data for the downscaling tests targeting European daily temperatures in MIKŠOVSKÝ & RAIDL (2005*) were supplied from the European Climate Assessment & Dataset (ECA&D - http://eca.knmi.nl/; KLEIN TANK ET AL. 2002). Daily temperatures employed in HUTH ET AL. (2015*) were provided by various partners within the CECILIA project (Central and Eastern Europe Climate Change Impact and Vulnerability Assessment - http://www.cecilia-eu.org/). Monthly temperature and precipitation series from several secular Czech weather stations and their areal averages (BRÁZDIL ET AL. 2012A*,B) were studied in BRÁZDIL ET AL. (2012A*) and MIKŠOVSKÝ ET AL. (2014*), and they also served as a basis for calculation of the drought indices analyzed in BRÁZDIL ET AL. (2015B*).

While the nature of the records taken at individual weather stations makes them useful for assessing the local climate, they are not necessarily representative of a larger neighborhood of their site of origin. Furthermore, mutual comparability of the series of direct measurements may be compromised by technical factors, particularly by differences among the measuring and record keeping practices of individual data gatherers (such as national weather services). For these reasons, composite datasets are often created from the local measurements, through interpolation/extrapolation techniques supported by various quality-control and homogenization algorithms (e.g. ŠTĚPÁNEK ET AL. 2011). The resulting data are then typically provided in the form of spatiotemporal

fields, often on a regular longitude-latitude geographic grid. Several such gridded datasets were employed within this thesis. Gridded versions of daily minimum and maximum temperature created within the CECILIA project (ŠTĚPÁNEK ET AL. 2011) were used in HUTH ET AL. (2015*). Gridded monthly temperature anomalies from GISTEMP (HANSEN ET AL. 2010) and Berkeley Earth (ROHDE ET AL. 2013A,B) datasets were utilized in the attribution studies MIKŠOVSKÝ ET AL. (2014*) and MIKŠOVSKÝ & PIŠOFT (2015*), along with the series of their continental and global means.

As primarily physical disciplines, meteorology and climatology rely heavily on mathematical representations of their respective systems of interest, particularly on numerical simulations. Over the past decades, these have evolved from simple, low-resolution models into complex, multi-component structures, capturing much of the large-scale weather/climate dynamics and its responses to external forcings. The current generation of global climate models (GCMs) not only serves as the main tool for generating outlooks of climatic future, but provides valuable insights into past climate as well. While the GCM-type simulations do not follow the historical deterministic trajectory of the climate system, they are constructed to preserve its general statistical characteristics – at least in theory, as this goal is still just partly fulfilled, and even the best state-of-the-art simulations suffer from numerous deficiencies (e.g. STOCKER ET AL. 2013, CHAP. 9). Outcomes of the HadCM3 model (GORDON ET AL. 2000) were used as a source of the simulated geopotential height data for the analysis of nonlinear behavior in MIKŠOVSKÝ ET AL. (2008*).

Being inherently world-wide simulations, GCMs do generally provide outputs on a relatively coarse spatial grid. The resolution gap between GCM-generated data and fine-scale inputs needed in local-oriented studies can then be bridged by regional climate models (RCMs): High resolution simulations over a geographically limited area, embedded into a global model or other suitable source of boundary conditions (such as global reanalysis). Of the numerous RCMs in existence, outputs of the RegCM3 (HALENKA ET AL. 2006) and ALADIN-Climate/CZ (FARDA ET AL. 2010) models were used in our works, and subjected to the performance comparison with their statistical downscaling alternatives in HUTH ET AL. (2015*).

The direct climatic measurements (and their gridded versions) provide records of the actual climate variability, but are available for just some historical periods and locations. GCM simulations can deliver (almost) complete data coverage over their integration period, yet they do not track the deterministic trajectory of the real climate system, and they suffer from various systematic biases. Outcomes of atmospheric reanalyses can be considered a transitory form between these two types of data: By assimilating measurements into a numerical model-like framework, a reanalysis can provide a formally complete description of the state of the atmosphere, while still following the trajectory of past climate in a deterministic sense. To study various thermobaric characteristics of the atmosphere, two representatives of the modern-era reanalysis products were used in several entries to this thesis: NCEP/NCAR reanalysis (providing data

since the year 1948; KISTLER ET AL. 2001) and ERA-40 reanalysis (covering the period 1957-2002; UPPALA ET AL. 2005). Of particular interest for investigation of longer-term climate variations is also the relatively recent 20th Century Reanalysis (COMPO ET AL. 2011), providing data from the year 1871 on, though not without some notable deviations from the gridded temperature observations, as shown below and in Sect. 6.2.

The range of data characterizing past climate is obviously vast, regarding both the general type of the dataset and its specific representative. Often multiple options are available as potential analysis inputs when a particular problem is to be studied. In theory, data from different sources should conform to the same, historical, evolution of the climate system at all relevant spatial and temporal scales (or, in the case of GCM/RCM simulations, the general dynamical and statistical features should be captured in a realistic manner). In praxis, however, differences between individual datasets can be substantial, and so can be distinctions between results stemming from their use. Careful selection of the inputs and interpretation of the results with regard to the possible data-related biases and uncertainties are therefore paramount in the statistical analysis of climatic data.

A simple illustration of the possible contrasts among individual representatives of atmospheric variables is shown in Fig. 2.1. Temperature anomalies characterizing the area of the Czech Republic at monthly and annual time step are compared for a series derived directly from the local observations within the Czech Republic, two specimen of gridded temperature data and the 20th Century Reanalysis. All the signals show a similar (though not completely identical) structure at the monthly time scale over the years 1980-2010 (Fig. 2.1a). On the other hand, systematic differences appear in the long-term trends, with noticeable discrepancy detected especially between the reanalysis and the rest of the datasets (Fig. 2.1b). When match of the temperature series provided by various data sources is investigated globally, strong regional contrasts emerge – see the correlation-based comparison of a few temperature datasets in Fig. 2.2 and notice, for instance, their generally good agreement in Europe, and their rather loosened similarity in parts of Africa or South America. These distinctions may then translate into deviations between outcomes produced by otherwise identical analysis procedures applied to the data from different sources, as seen, for instance, from the attribution-focused example in Fig. 6.3.

In the frame of the topics addressed within this thesis, issues related to the problem of inter-dataset differences have been tackled to some (although admittedly limited) extent. Possible manifestations of nonlinearity in short-term prediction of the (pseudo)observed data were compared for direct meteorological measurements and their reanalysis-based counterparts (MIKŠOVSKÝ & RAIDL 2006*). Reanalysis data were also compared to the outputs of a global climate model, in terms of the geographical patterns of nonlinearity detectable from local multivariable systems (MIKŠOVSKÝ ET AL. 2008*). Consequences of gridding the station-based data were investigated in HUTH ET AL. (2015*), in the context of tests of various statistical and dynamical downscaling approaches. The effects of using

alternative versions of the input data were also considered in the statistical attribution analysis targeting the Czech climatic series (MIKŠOVSKÝ ET AL. 2014*), although only a very brief summary of the respective conclusions was then included in the paper itself. Some attention to the matter of inter-dataset contrasts was then paid in our attribution study MIKŠOVSKÝ & PIŠOFT (2015*), too, and this issue will be studied even more closely in the upcoming paper MIKŠOVSKÝ ET AL. (2015*). But even from the limited sample of results presented here, it should be clear that the problem of data-specific features and uncertainties needs to be treated with great care. Questions of whether and when directly measured climatic variables can be replaced by their gridded/reanalyzed/simulated counterparts (and which specific dataset should be used) must be carefully considered, and assessment of the effects of such choice is an important part of the studies dealing with spatiotemporal relations and variability in the climate system.

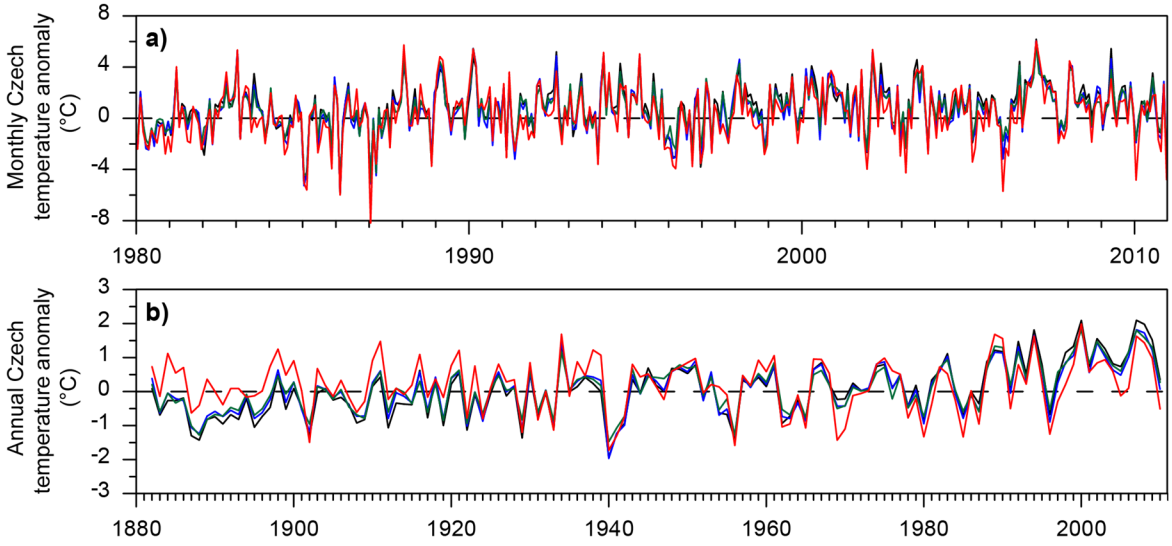
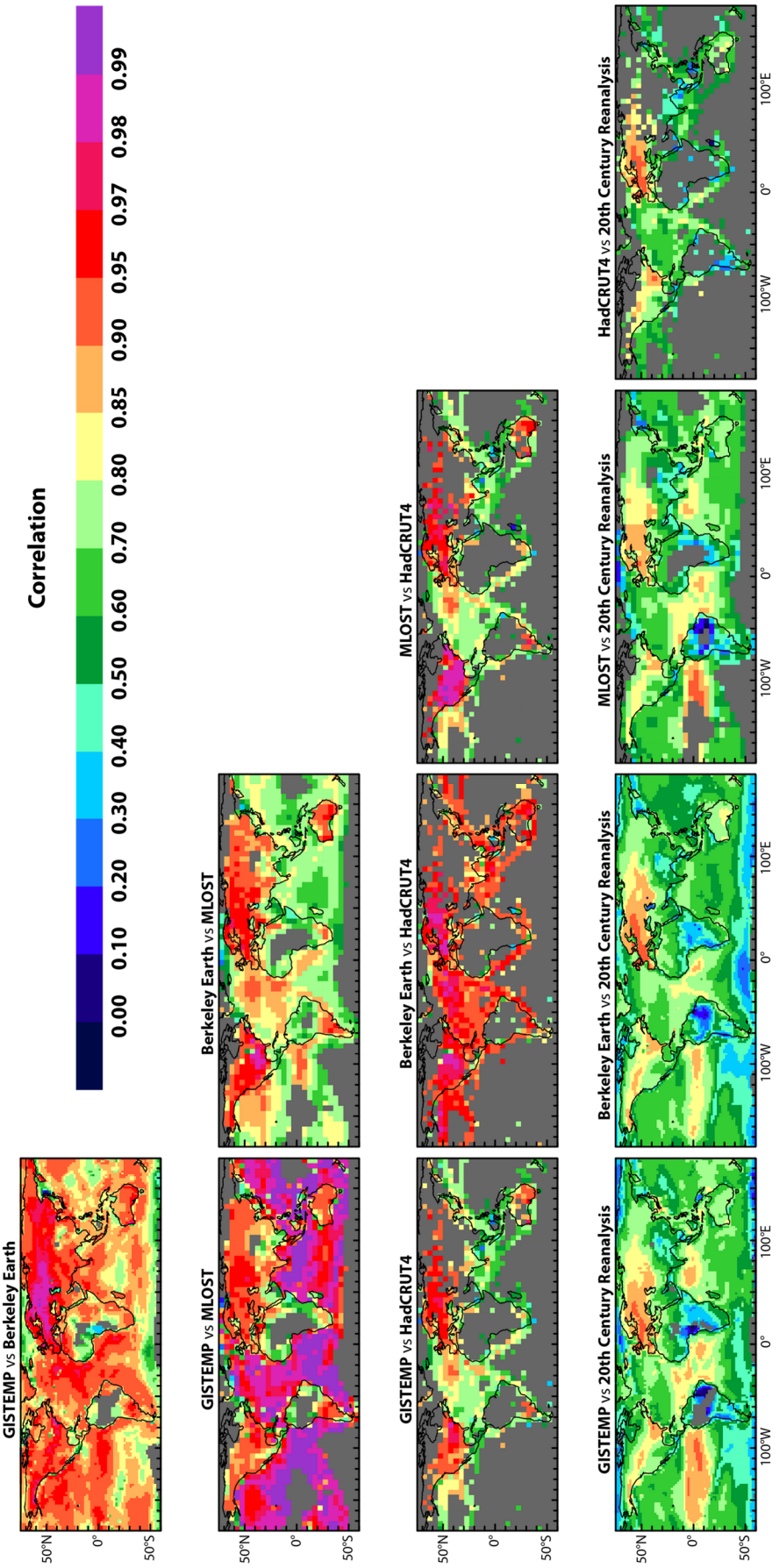


FIGURE 2.1: Time series of monthly **(a)** and annual **(b)** temperature anomalies for the area of the Czech Republic derived from data obtained from various sources: Mean areal temperature created from measurements at 10 Czech weather stations (**black:** BRÁZDIL ET AL. 2012A*); GISTEMP dataset (**green:** HANSEN ET AL. 2010); Berkeley Earth dataset (**blue:** ROHDE ET AL. 2013A,B); 20th Century Reanalysis (**red:** COMPO ET AL. 2011). The anomalies are expressed relative to the 1951-1980 period and shown for the years 1980-2010 (monthly series) and 1882-2010 (annual series).

FIGURE 2.2 (➔): Local values of Pearson correlation coefficient between time series of monthly temperature anomalies from selected global gridded datasets: GISTEMP (HANSEN ET AL. 2010); Berkeley Earth (ROHDE ET AL. 2013A,B); MLOST (SMITH ET AL. 2008); HadCRUT4 (MORICE ET AL. 2012); 20th Century Reanalysis (COMPO ET AL. 2011). The correlations were calculated over the 1901-2010 period; grey areas mark regions with insufficient amount of data available (more than 10% of missing temperature pairs in the analysis period). Adapted from materials to be included in the upcoming paper MIKŠOVSKÝ ET AL. (2015*).


Correlation
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.85 0.90 0.95 0.97 0.98 0.99
GISTEMP vs Berkeley Earth
Berkeley Earth vs MLOST
MLOST vs HadCRUT4
HadCRUT4 vs 20th Century Reanalysis
GISTEMP vs MLOST
Berkeley Earth vs HadCRUT4
MLOST vs 20th Century Reanalysis
GISTEMP vs HadCRUT4
Berkeley Earth vs 20th Century Reanalysis
GISTEMP vs 20th Century Reanalysis
50°N
0°
50°S
100°W
0°
100°E

CHAPTER 3
(NON)LINEAR REGRESSION TECHNIQUES

A wide range of statistical techniques was used to investigate individual problems presented throughout this text, from estimation of elementary descriptive statistics, to dimensionality reduction and clustering algorithms and an assortment of statistical significance tests. One particular topic, however, permeates through most of the analyses presented here: Application of various forms of linear and nonlinear regression, connecting a univariate predictand $y(t)$ to one or more predictors $p_i(t), i = 1, \dots, M$. Index t distinguishes between individual cases in the datasets studied (out of the total of N available), and it mostly pertains to time here. While straightforward in their basic purpose, regression mappings can be employed to fulfill various objectives, determined by the character of variables assigned to the role of predictand and predictors. Within the range of problems tackled here, regression was used for predictive tasks (i.e., predictand estimated from predictors preceding it in time), approximation of spatial relations (with concurrent predictand and predictors originating from different geographic locations), trend estimation (matching the target variable against time) or as a basis for attribution-seeking models (decomposing predictand into signals associated with explanatory variables representing various external and internal climate forcings). In this chapter, selected classes of regression models relevant to this thesis are very briefly outlined, with regard to their basic structure as well as some details concerning their implementation in our works.

A prominent (and historically dominant) place among the regression techniques is held by the multiple linear regression (MLR). The respective mapping between predictors and predictand takes a form of a simple weighted averaging formula,

$$y(t) = \hat{y}(t) + \varepsilon(t) = a_0 + \sum_{i=1}^{M} a_i p_i(t) + \varepsilon(t) \,, \qquad (1)$$

with regression coefficients a_i calculated to obtain a model of desired properties – typically one that minimizes the sum of squared regression residuals ε, calculated as differences between the actual values of y and their regression-based estimates $\hat{y}$. This so-called 'least squares method' of a_i calculation was employed in all applications of linear regression here; the specific implementational and preprocessing details are given in the respective publications.

While simple, fast and open to easy interpretation of its outcomes, linear regression suffers from an obvious limitation: In its basic form, it is only able to capture strictly linear links, embodying direct proportionality between the predictors and individual components in the predictand. However, it has been shown that linear mappings can be used to approximate dynamics of even strongly nonlinear systems, providing that linear models are applied locally for just small sections of the phase space or space of predictors (e.g., contributions in

OTT ET AL. 1994). This approach, dubbed method of local linear models (LLM) here, relies on calculation of the regression coefficients a_i individually for each instance of t. The coefficients can then no longer be considered globally valid constants, but rather t-dependent functions:

$$y(t) = \hat{y}(t) + \varepsilon(t) = a_0(t) + \sum_{i=1}^{M} a_i(t) p_i(t) + \varepsilon(t). \quad (2)$$

To achieve the local specificity of the regression coefficients, their calculation is carried out for just a limited number $L \ll N$ of cases from the calibration part of the data, representing situations with the closest resemblance to the one being processed (i.e., to the one pertaining to t). The similarity of individual cases can be measured by the distance of the respective M-dimensional vectors of predictors $\mathbf{p}(t) = \big(p_1(t), \ldots, p_M(t)\big)$, quantified by a suitable metric (often Euclidean). The optimum size and structure of the local neighborhood is subject to the specifics of the task investigated, including dimensionality of the system studied, type of time series involved and their eventual contamination by noise. Details on the design of the local linear models employed in our analyses are given in the individual papers in the appendices.

Over the past years, great popularity among the nonlinear regression techniques has been attained by various architectures of artificial neural networks (NNs) (see, e.g., HAYKIN 1999). The perhaps most prominent of them, multilayer perceptron (MLP), was employed in several of our studies, in a form containing a single hidden layer,

$$y(t) = \hat{y}(t) + \varepsilon(t) = b_0 + \sum_{i=1}^{L_{MLP}} b_i \, \varphi\left(b_{0i} + \sum_{j=1}^{M} b_{ji} \, p_j(t) \right) + \varepsilon(t), \quad (3)$$

where b_{ji} and b_i represent weights of connections between neurons in the input and hidden layer and in the hidden and output layer, respectively, and L_{MLP} denotes number of neurons in the hidden layer (and thus specifies complexity of the network). Of the possible forms of the (generally nonlinear) transfer function φ, either logistic function (MIKŠOVSKÝ & RAIDL 2005*, 2006*) or hyperbolic tangent (BRÁZDIL ET AL. 2012A*; MIKŠOVSKÝ ET AL. 2014*; BRÁZDIL ET AL. 2015B*; HUTH ET AL. 2015*) were applied in the examples here. The learning algorithms (i.e., procedures used to calculate weights b from the data available for calibration of the network) were based on error backpropagation, either in its basic form or in the quasi-Newtonian version.

An alternative type of neural networks built around radial basis functions (RBFs) (see, e.g., HAYKIN 1999) was also applied in some of our studies. The respective mapping can be captured by the formula

$$y(t) = \hat{y}(t) + \varepsilon(t) = c_0 + \sum_{i=1}^{L_{RBF}} c_i \, \rho(\|\mathbf{p}(t) - \mathbf{d}_i\|) + \varepsilon(t), \quad (4)$$

with M-dimensional vector $\mathbf{d}_i$ representing center of the radial function assigned to the i-th of L_{RBF} neurons in the hidden layer. In our analysis setups, Gaussian-style

RBFs were used, $\rho(\|\mathbf{p}(t) - \mathbf{d}_i\|) = \exp(-\|\mathbf{p}(t) - \mathbf{d}_i\|^2/2\sigma^2)$, with parameter σ controlling the width of the radial functions. Simple subsampling of the centers $\mathbf{d}_i$ from the training part of the datasets was typically employed, although more sophisticated methods (e.g. pre-processing through clustering algorithms) were also tested, but to little effect. The weights c_i were then calculated to minimize the sum of squared errors, in a fashion analogous to multiple linear regression.

The above introduced regression techniques do share a common purpose: to capture relations between the explanatory variables and the target signal. Intuitively, one might expect the nonlinear mappings to be more universal in their ability to approximate the respective links, and thus automatically superior to linear regression. In reality, such presumption often turns out to be unsupported: Despite the inherently nonlinear and deterministically chaotic nature of the Earth's climate system, deviations from purely linear behavior are not always detectable in the time series is spawns. Moreover, application of nonlinear algorithms typically comes with increased demands on computational power, more difficult interpretation of the regression outcomes and more complicated evaluation of their statistical significance. The question therefore remains how beneficial non-linear techniques really are and whether gain from their application outweighs the extra demands and interpretational challenges.

Even in the presence of nonlinearities strong enough to uphold the application of nonlinear regression, another important design choice has to be made: Selection of the most suitable form of the nonlinear mapping. The three examples above, embodied by Equations 2-4, represent different approaches to this problem. The method of local linear models builds upon an ensemble of individual, formally independent regression functions, pertaining to specific (and typically mutually overlapping) segments of the phase space. Multilayer perceptrons, on the other hand, can be considered a global mapping, without a specific link of individual neurons to particular states of the system (or vectors of the predictors). RBF-based networks form a middle ground between these two approaches: While the mapping is formally global, individual hidden neurons are associated with specific vectors in the space of predictors, and their activation is reduced for inputs more distant from their assigned centers. The general form of the regression function is not the only important factor determining the behavior of the nonlinear models: Their individuality is subject to the selection of the structure-defining descriptors (such as the complexity-controlling parameters L, L_{MLP} or L_{RBF} above), and finding the optimum setup is as critical as it is nontrivial. Some specific aspects of this problem are illustrated in the following chapters and in the respective publications in the appendices.

CHAPTER 4
NONLINEARITY IN PREDICTIVE MAPPINGS

Over the past decades, various methods have been developed for assessing the presence – and potentially magnitude – of nonlinear and chaotic behavior in univariate or multivariate time series. Numerous attempts have also been made to apply these techniques in the atmospheric and climatic sciences – see, for instance, the overview by SIVAKUMAR (2004) and the references discussed in MIKŠOVSKÝ ET AL. (2008*). The emergence of global- or continental-scale datasets of climatic data (particularly outcomes of various reanalysis projects) provided an opportunity for an even more systematic investigation of this problem, including the evaluation of geographic and seasonal patterns of nonlinearity. However, the variety of results in the existing studies also demonstrates that degree to which deviations from strictly linear behavior manifest depends on a number of factors, related to the datasets analyzed as well as tasks performed. Outcomes of nonlinearity tests are therefore subject to the choice of the testing criterion, reflecting the particular form of nonlinear interaction of interest. Prediction errors represent one of the natural choices of the discriminating statistic: Due to their relation to the information transfer between consequent states of the climate system, tests based on short-term predictive mappings can provide useful information about the local properties of the atmosphere, connected to its chaoticity and predictability. In this chapter, our experiments pertaining to this topic are outlined, published in the papers MIKŠOVSKÝ & RAIDL (2006* - APPENDIX I), MIKŠOVSKÝ ET AL. (2008* - APPENDIX II) and MIKŠOVSKÝ & RAIDL (2005* - APPENDIX III). Some of the earlier versions of the related materials were also previously included in my dissertation thesis (MIKŠOVSKÝ 2004*).

Our initial attempts at nonlinearity detection were focused on identification of rules governing the manifestations of nonlinear behavior in short-term forecasts of daily temperature and pressure, as documented in MIKŠOVSKÝ & RAIDL (2006*). The tests applied were built upon the method of surrogate data, employing the Iterative Amplitude Adjusted Fourier Transform (IAAFT) technique (SCHREIBER & SCHMITZ 1996, 2000). Implementation of the respective algorithms from the TISEAN software package was used (HEGGER ET AL. 1999; http://www.mpipks-dresden.mpg.de/~tisean/). Both univariable and multivariable time series were investigated for the presence of nonlinearities, using either the method of time delays (e.g. PACKARD ET AL. 1980) or the multivariate approach (e.g. KEPPENNE & NICOLIS 1989) to reconstruct the phase space of the local climate system (or, more accurately, to provide its approximate representation, and a set of predictors to enter the predictive regression mappings). Series of daily temperature (mean, minimum and maximum) and daily pressure measured at the weather station Prague-Ruzyně (Czech Republic) served as predictands, and they were complemented by their counterparts provided by the NCEP/NCAR reanalysis. The

reanalysis also supplied potential predictors for the multivariable analysis setups, with step-wise screening used to identify the best subset of the explanatory variables.

Figure 4.1 provides an illustrative sample of the outcomes of the surrogate data-based analysis in MIKŠOVSKÝ & RAIDL (2006*), comparing errors of prediction carried out by the method of local linear models for the original data and for an ensemble of their IAAFT-randomized versions. It was demonstrated that nonlinear behavior does indeed manifest in the predictive mappings, but only in some test configurations and in greatly varying degree. Just mild to no detectable nonlinearity (i.e., small difference between the prediction errors in the original data and in the surrogates) was indicated for the setups with predictors generated by the method of time delays. On the other hand, a distinct nonlinear component was typically uncovered in the predictive mappings employing multivariable predictors. Nonlinearity was generally stronger for longer signals (30-year-long series) than for their shortened (10-year-long) versions. It was also comparably most noticeable for the shortest-term prediction (lead time of 1 day), weakening and eventually disappearing as the lead time increased. Generally, our results suggested that nonlinear behavior manifests more strongly in setups with higher amount of information available within the data analyzed, provided that a deterministic link between predictand and predictors exists. The information content in individual scalar signals seemed insufficient to describe the complex dynamics of the local climate system beyond simple linear links, and application of nonlinear predictive mappings was thus largely baseless for the univariate settings (at least for the particular type of time series studied in our tests).

While the surrogate data-based tests can deliver statistically well founded conclusions about the presence of specific forms of nonlinearity, they are somewhat cumbersome and computationally demanding. From the perspective of applied time series analysis, a more direct question regarding nonlinear behavior may be of interest: What is the actual improvement achieved by application of a specific nonlinear mapping over its linear counterpart? This issue was only very briefly touched upon in MIKŠOVSKÝ & RAIDL (2006*), but we focused on it more specifically in MIKŠOVSKÝ & RAIDL (2005*). Comparison of the short-term predictive skill of linear regression and local linear models was carried out for daily temperatures across the European region, supplied from the NCEP/NCAR reanalysis. Multivariable predictors were used, arranged in a pre-defined geographic pattern. In addition to the method of local linear models, MLP and RBF neural networks were also applied, to assess the sensitivity of the results to the choice of the nonlinear model. Relatively strong nonlinear behavior (i.e., superiority of nonlinear methods over linear regression) was generally indicated, especially during boreal winter. Distinct geographic variations of nonlinearity were found, but just rudimentary explanation of their spatial patterns could be provided. Mostly insignificant differences between the predictive skills of individual types of nonlinear mappings were found.

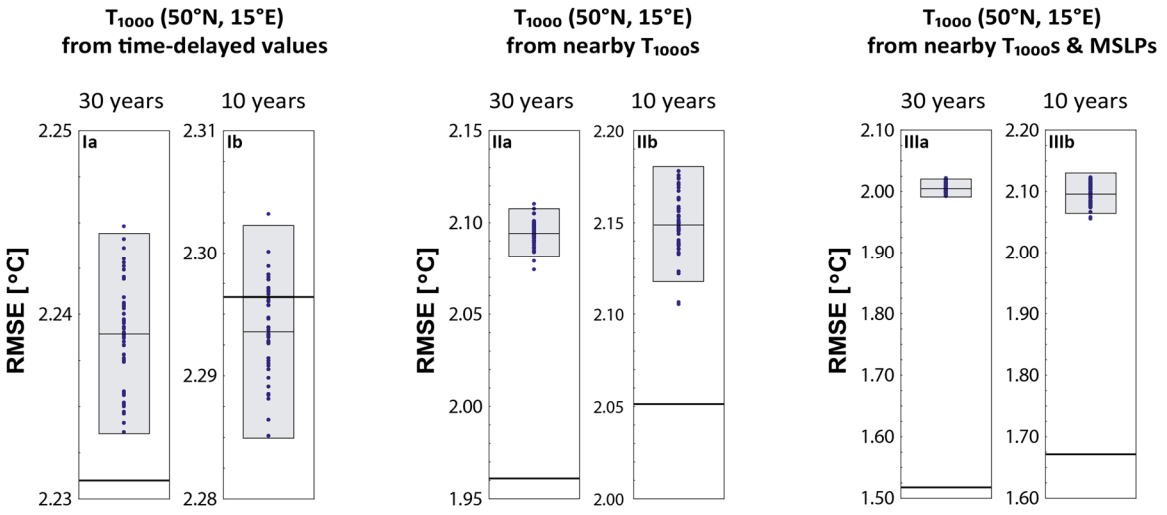


FIGURE 4.1: Manifestations of nonlinear behavior in univariable and multivariable time series. Root mean squared error (RMSE) of NCEP/NCAR daily temperature series (50°N, 15°E, 1000 hPa level) forecast 1 day ahead is shown, obtained by the method of local linear models for the original series (long horizontal line) and 49 instances of the corresponding IAAFT-generated surrogates (dots). Individual setups pertain to phase space reconstruction by the method of time delays **(I)**, multivariate reconstruction employing 1000 hPa temperatures from a region between 60°N, 0°E and 40°N, 30°E **(II)** and multivariate reconstruction employing 1000 hPa temperatures as well as mean sea level pressures from the same region **(III)**. Results are shown for approximately 30-year-long **(a)** and 10-year-long **(b)** versions of the series. The embedded rectangle with shorter inset horizontal line shows average RMSE for the surrogates and the matching 2σ range. Adapted from MIKŠOVSKÝ & RAIDL (2006*), where more details and other related results can be found.

In MIKŠOVSKÝ & RAIDL (2005*) and MIKŠOVSKÝ & RAIDL (2006*), we focused on nonlinearity manifestations within just a geographically limited region, and only (pseudo)observed time series were studied (either direct measurements or series originating from a reanalysis). In MIKŠOVSKÝ ET AL. (2008*), a global scope of the analysis was embraced, and outcomes of the HadCM3 global climate model were investigated along with data originating from the NCEP/NCAR reanalysis. The primary method of nonlinearity quantification in MIKŠOVSKÝ ET AL. (2008*) was based on direct comparison of the 1-day-ahead prediction error achieved by multiple linear regression and by the local linear models method, with multivariable predictors arranged in a regular pattern, centered on the location of the predictand (Fig. 4.2a). The role of predictand was filled by the relative topography of the 850-500 hPa layer (i.e., a quantity proportional to the average atmospheric temperature between the 850 and 500 hPa pressure levels) and by the geopotential height of the 850 hPa level.

The global nonlinearity patterns in the NCEP/NCAR data revealed a distinct contrast between relatively strong (and generally statistically significant) nonlinearities in the midlatitudes and largely negligible and insignificant improvement from application of a nonlinear predictive model in the equatorial regions

(Fig. 4.2b). Besides this basic latitudinal pattern, areas with the strongest manifestation of nonlinearity in the higher latitudes were identified and their possible link to the atmospheric zones with intensive synoptic activity was discussed. Our analysis also confirmed presence of distinct seasonal variations of the results, with nonlinearity typically intensified during the cold part of the year in the extratropical regions.

By comparing the nonlinearity patterns for the NCEP/NCAR reanalysis (approximating the actual historical variability of the climate system) and for the HadCM3 model (global numerical simulation, generating a trajectory uncorrelated with the historical one), we confirmed that the model is capable of reproducing the basic character of the observed nonlinearity patterns quite realistically, although differences appeared in both the finer details of the structures detected and in their magnitude (Fig. 4.2c). Our analysis thus served as an advanced validation tool of the GCM and suggested the ability of global climate models to replicate not only the elementary statistical characteristics of the climatic data, but also their properties related to the nonlinear and chaotic structures.

Finally, nonlinearity tests based on assessing the ratio between the prediction errors from the MLR and LLM methods were also compared to the approach employing surrogate data. Quite good match between the respective geographic patterns of nonlinearity was found (see Figs. 3a and 6 in MIKŠOVSKÝ ET AL. 2008[*]). This suggests that comparing errors from a linear and nonlinear mapping may be used as an alternative to the computationally more expensive surrogate-assisted testing (with some reservations, discussed in MIKŠOVSKÝ ET AL. 2008[*]). However, such conclusion should not be mistaken for invariance regarding the analysis setup: Choice of the specific form of the nonlinear model (and of its design parameters) can still affect the results to some extent, which needs to be taken into account when interpreting the outcomes of the nonlinearity tests.

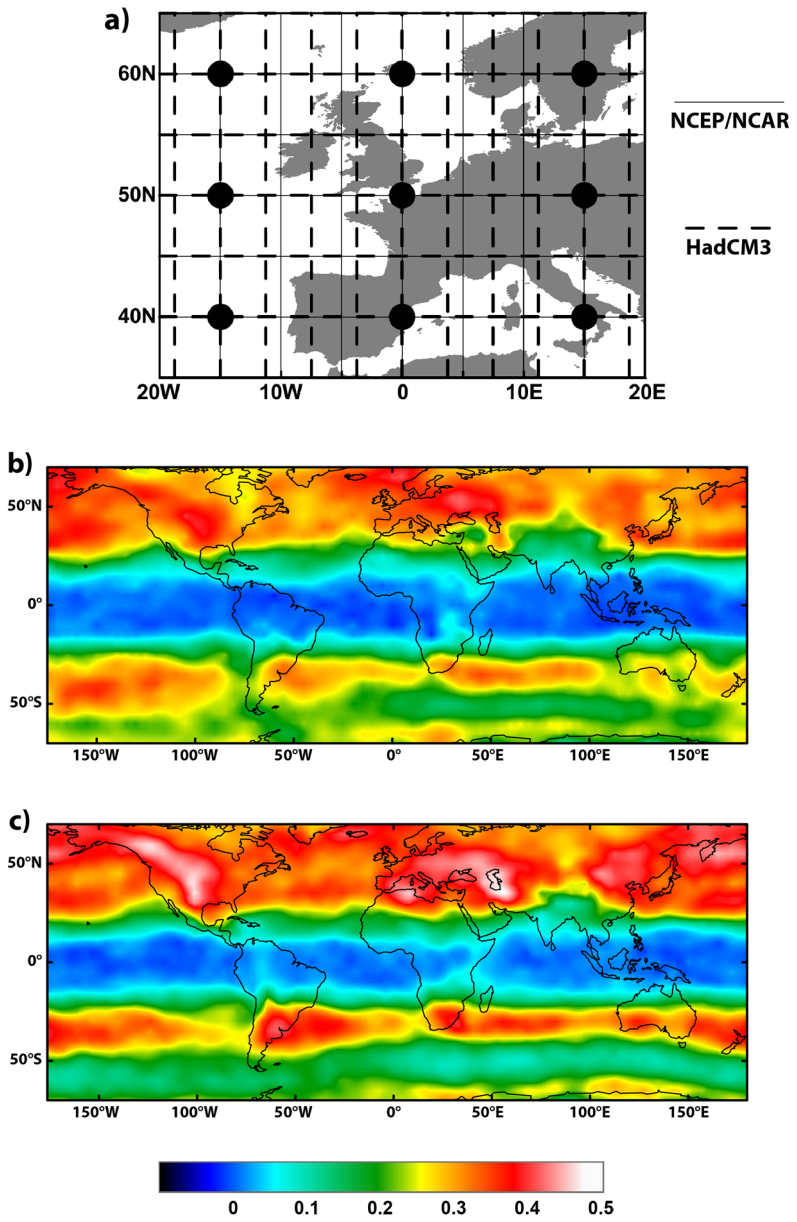


FIGURE 4.2: Global distribution of estimated regional magnitude of nonlinearity, associated with prediction of relative topography 850-500 hPa 1 day ahead. Multivariable vector of predictors was used, consisting of 9 values of relative topography 850-500 hPa and 9 values of geopotential height of the 850 hPa level, arranged in a pattern shown in **(a)** for the predictand series located at 50°N, 0°E. Nonlinearity was quantified by a skill score defined as $SS = 1 - (R_{LLM}/R_{MLR})^2$, with R_{LLM} and R_{MLR} representing root mean squared error of the forecast by the method of local models and multiple linear regression, respectively (by this definition, $SS = 0$ pertains to situations with both methods performing identically in terms of RMSE, and thus no detectable nonlinearity, while positive values of SS indicate nonlinear mapping outperforming its linear counterpart). Results are shown for the NCEP/NCAR reanalysis data **(b)** and for the outputs of the HadCM3 global climate model **(c)**, with the forecast mappings calibrated over the 1961-1990 period and validated for the years 1991-2000. Adapted from MIKŠOVSKÝ ET AL. (2008*), where more details and other related results can be found.

CHAPTER 5
SPATIAL RELATIONS IN CLIMATIC DATA

It is typical for climatic variables characterizing geographically close locations to share a portion of their temporal variability, and for the respective time series to be connected to some degree. These associations are often studied through simple linear correlations, but their nature may also be considerably more complex. Regression analysis techniques can be used to identify, extract and quantify the inter-variable dependencies; they can also help to reveal and describe connections between different datasets (for instance, to estimate station-specific series from large-scale data available from a reanalysis or global climate model). In this section, examples are given of our results related to approximation of spatial relations within and among various datasets of climatic data: Downscaling of large-scale atmospheric fields (Chap. 5.1; MIKŠOVSKÝ & RAIDL 2005* - APPENDIX III; HUTH ET AL. 2015* - APPENDIX IV) and estimation of temperature measurements from nearby concurrent records (Chap. 5.2).

5.1 STATISTICAL DOWNSCALING OF DAILY TEMPERATURES

As already mentioned in Chap. 2, spatial resolution of global climate models (as well as of global reanalyses) is often insufficient for local-oriented studies, and the resolution gap can be bridged by dynamical downscaling (i.e., by application of a high-resolution regional climate model embedded into the global simulation or reanalysis). As an alternative to such cascade of numerical simulations, statistical methods can also be used to approximate the connections between large-scale climatic fields and more site-specific data (such as observations at individual weather stations). Of the various techniques of statistical downscaling in existence, we focused on direct mappings between the large scale data (predictors) and local measurements or their gridded versions (predictands) in our works.

In MIKŠOVSKÝ & RAIDL (2005*), our main aim was to assess the suitability of different forms of empirical regression functions to provide downscaled versions of daily temperature. Using NCEP/NCAR reanalysis data as predictors, the four regression mappings introduced in Chap. 3 (MLR, LLM, MLP NN, RBF NN) were used to generate estimates of daily mean, minimum and maximum temperature, recorded at 25 sites across Europe and obtained from the ECA&D database (KLEIN TANK ET AL. 2002). A pre-defined pattern of predictors was employed (Fig. 5.1a). The regression models were calibrated using data from the 1961-1990 period and then validated for the years 1991-2000, separately for each location. Distinct differences between the temperature estimation errors for individual stations were found (see the example for daily maximum temperature in Fig. 5.1b,c, as well as figures and tables in MIKŠOVSKÝ & RAIDL 2005*). No clear geographic pattern of the error magnitudes was identified, suggesting a dominant

influence of the local specifics of each of the target sites. The analysis also highlighted a tendency for stronger nonlinearity during the boreal winter, though exceptions from this inclination were detected for some combinations of temperature type and location. Downscaling skill of the three nonlinear regression techniques (LLM, MLP NN, RBF NN) was found to be mutually similar.

The problem of daily temperature downscaling was later revisited in HUTH ET AL. (2015*), this time to provide a detailed comparison of the performance of various dynamical and statistical downscaling methods. The analysis utilized a high-resolution dataset of daily maximum and minimum temperature series, assembled within the CECILIA project (http://www.cecilia-eu.org/; ŠTĚPÁNEK ET AL. 2011) and providing both station-specific records and their versions interpolated onto a regular grid, for a geographically limited region along the joint borders of Austria, Czech Republic, Hungary and Slovakia. In addition to multiple linear regression and the three representatives of nonlinear regression (LLM, MLP NN, RBF NN), method of analogues (e.g. ZORITA & VON STORCH 1999) was also employed and compared to the other downscaling approaches. Predictors were supplied from the ERA-40 reanalysis and pre-selected through a step-wise screening procedure based on linear regression. Calibration of the regression mappings was carried out for the years 1961-1990, and their validation performed over the 1991-2000 period. The dynamical downscaling models were represented by the ERA-40-driven integrations of the RegCM3 (HALENKA ET AL. 2006) and ALADIN-Climate/CZ (FARDA ET AL. 2010) regional climate models.

In Fig. 5.2, performance of some of the downscaling techniques applied in HUTH ET AL. (2015*) is illustrated, through root mean squared errors of estimation of winter minimum daily temperature. Superiority of nonlinear regression over MLR was once again indicated, though exceptions were detected for some combinations of season, location and temperature type. Unlike in MIKŠOVSKÝ & RAIDL (2005*), however, RMSE did not serve as the primary validation criterion in HUTH ET AL. (2015*). Instead, emphasis was on evaluating the ability of the statistical and dynamical downscaling models to realistically reproduce the extreme quantiles of the statistical distributions, their higher moments (skewness, kurtosis), autocorrelation structures in the time series, spatial correlations between temperatures from different locations and long-term temporal trends in the series. As individual sections in HUTH ET AL. (2015*) show, no downscaling technique was found to be universally superior to the others. Depending on the type of temperature, location, season and validation criterion, the relative skill rank of individual downscaling approaches varied greatly: In some cases, statistical downscaling techniques out-performed the (arguably more popular) regional climate models, but the opposite was also occasionally true. Also, despite the relative superiority of nonlinear empirical models over linear regression in terms of RMSE, their advantage did not automatically extend to the above mentioned validation criteria related to statistical distributions or spatiotemporal correlations.

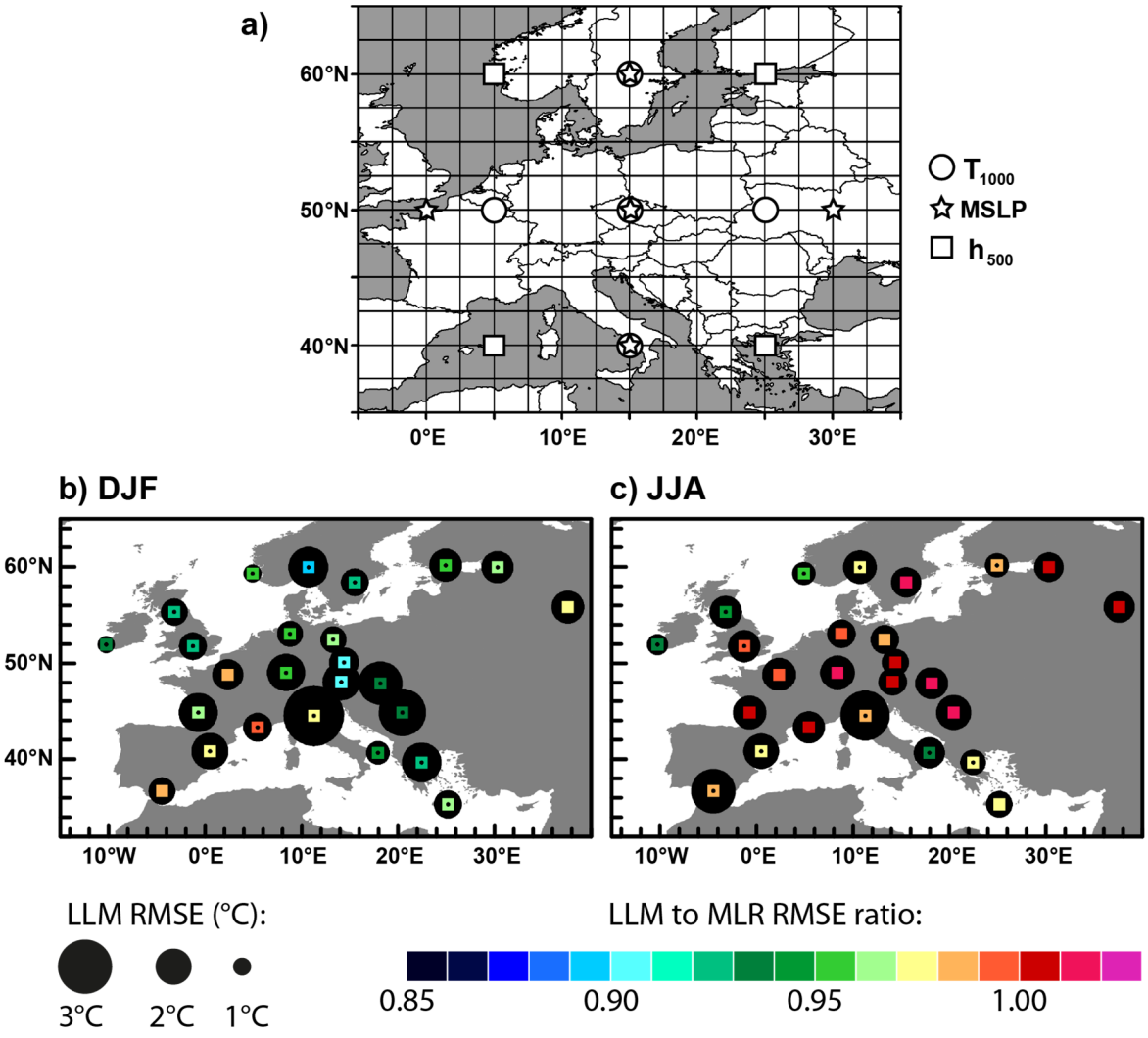


FIGURE 5.1: Results of maximum daily temperature downscaling for 25 European locations. A set of NCEP/NCAR reanalysis predictors consisting of the series of 1000 hPa level temperature (T_{1000}), mean sea level pressure (MSLP) and 500 hPa level geopotential height (h_{500}) was used. The predictors were arranged in a pre-defined pattern centered on the grid point closest to the target station, as illustrated in **(a)** for predictand located near coordinates 50°N, 15°E. Outcomes of the analysis are shown for boreal winter **(b)** and summer **(c)**. Root mean squared error (RMSE) of the temperature estimate is displayed through the size of the circle at the station's location, along with the ratio of RMSEs obtained by the method of local linear models (LLM) and multiple linear regression (MLR) (color of the embedded square). Presence of a central dot indicates statistically significant ($\alpha = 0.05$) difference between the series downscaled by LLM and MLR methods, according to the paired Wilcoxon test. Adapted from MIKŠOVSKÝ & RAIDL (2005*), where more details and other related results can be found.

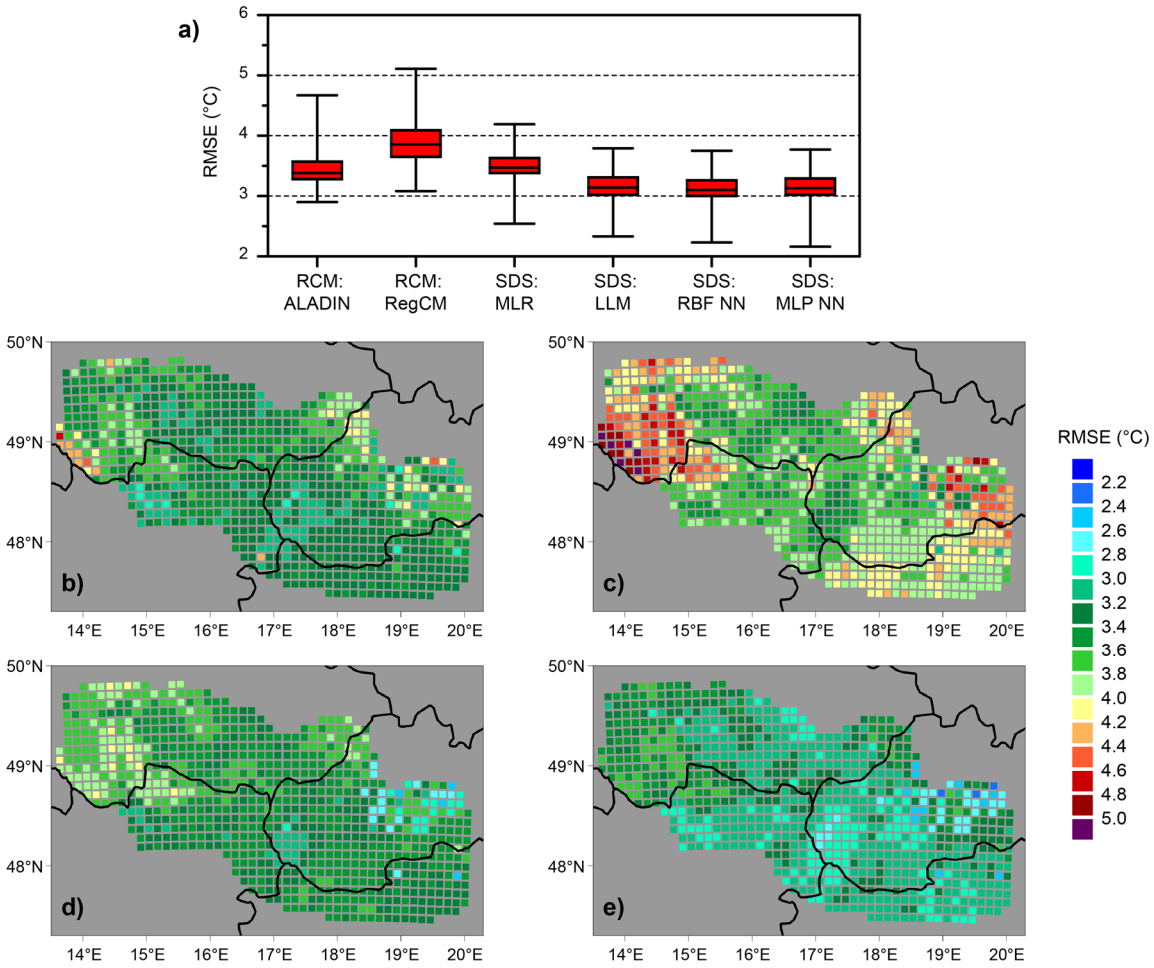


FIGURE 5.2: Root mean squared error (°C) of minimum daily temperature estimates in boreal winter (December, January, February), obtained by different methods of dynamical (RCM) and statistical (SDS) downscaling, using ERA-40 reanalysis data as inputs. Statistical distribution of errors within the target area is displayed in the form of boxplots, showing min-max range of the values, their inter-quartile range and median **(a).** Geographic pattern of the errors is visualized for the ALADIN climate model **(b)**, RegCM climate model **(c)**, statistical downscaling by multiple linear regression **(d)** and statistical downscaling by the method of local linear models **(e)**. Adapted from the outcomes of the analysis presented in HUTH ET AL. (2015*), where more details on the test setup and other related results can be found.

5.2 ESTIMATION OF DAILY TEMPERATURES FROM OTHER CONCURRENT RECORDS

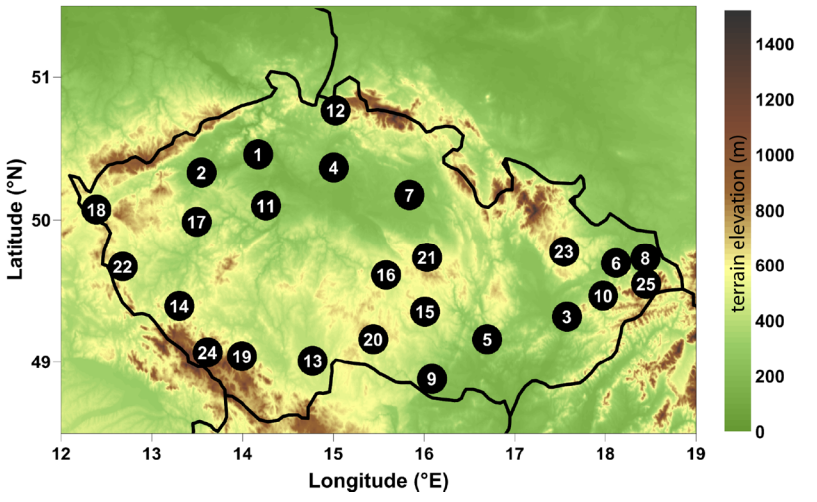
While the series of meteorological measurements from land-based weather stations represent one of the basic types of data in the atmospheric research, it is not uncommon for these records to be incomplete, interrupted by shorter or longer periods of missing values. Often, such gaps need to be filled before a subsequent analysis can be performed, and records from other nearby sites are used to do so. In this section, outcomes of my experiments with estimating daily temperature data from other concurrent measurements are briefly presented, with an emphasis again on comparing the performance of linear and nonlinear regression techniques. Although these results were not published as a stand-alone paper, their sample was included here to demonstrate yet another application of regression mappings for approximation of the spatial relations among climatic time series.

The tests were conducted on a dataset comprising daily mean, minimum and maximum temperature from 25 Czech weather stations (Fig. 5.3). Linear and nonlinear regression was used to generate estimates of each of these temperature series from the temperature records at the rest of the weather stations and from the temperatures and geopotential heights provided by the ERA-40 reanalysis. The regression mappings employed included multiple linear regression, method of local linear models and MLP and RBF neural networks, as introduced in Chap. 3. The pool of potential predictors consisted of mean, minimum and maximum temperature from the remaining 24 stations, as well as ERA-40 series of temperature and geopotential height at the 1000 hPa and 850 hPa levels from the area bounded by 40°N, 60°N, 0°E and 30°E. A step-wise screening procedure based on multiple linear regression was applied to identify the 20 most influential predictors, individually for each temperature type and location. These were then used as inputs for all four empirical models. The regression mappings were calibrated for the years 1961-1990 and validated for the 1991-2000 period. Other technical details of the tests were similar to those in HUTH ET AL. (2015*). The temperature estimates by different regression models were compared mutually and also to the outcomes of inverse distance weighting (IDW), one of the most common geostatistical interpolation techniques (e.g. JARVIS & STUART 2001).

Figure 5.4 summarizes root mean squared errors of the temperature estimates obtained for individual weather stations and temperature types. On average, all nonlinear models outperformed multiple linear regression. Gain from considering the nonlinear components of the spatial relations was generally strongest for the high-elevation weather stations, which can be considered atypical sites in their local geographic neighborhood. At locations with another station of similar character situated nearby, differences between outputs of linear and nonlinear mappings tended to be smaller, as did total error. Performance of RBF neural networks and of the method of local linear models was mutually comparable. Multilayer perceptrons, although no worse on average than the other

two nonlinear methods, have produced substantially greater dispersion of errors relative to linear regression, sometimes even giving less accurate temperature estimates than MLR. This intermittent performance loss did not seem to be related to MLP's sensitivity to the initialization of its training procedure (which was found to be quite low). Rather, it was traced to the greater vulnerability of the multilayer perceptrons to the inhomogeneities in the input data, present in some of the series of the station-based measurements. All regression techniques (including MLR) distinctly outperformed IDW interpolation. Experiments were also performed using ERA-40 series alone as the regression inputs, thus basically employing a downscaling-like setup (Fig. 5.4d). Unsurprisingly, the use of reanalysis-only predictors increased the error of temperature estimation. However, the loosening of the predictors-predictand links also resulted in generally greater relative improvement from the application of nonlinear regression models, which now outperformed multiple linear regression by an even greater margin.

While not shown here, regression models' ability to realistically reproduce statistical distributions of the temperature series was investigated as well. The validation statistics included standard deviation of the temperatures, as well as their higher statistical moments (skewness and kurtosis). Performance of all regression techniques was found to be generally good and mutually comparable in this regard, with only mild advantage occasionally indicated for the nonlinear mappings. On the other hand, all regression models, linear and nonlinear alike, displayed just very limited ability to realistically reproduce the individual long-term temporal trends in the temperature series.



Station no.	Station name	Longitude (°E)	Latitude (°N)	Elevation (m)
1	Doksany	14.17	50.46	158
2	Žatec	13.54	50.33	201
3	Holešov	17.57	49.32	224
4	Semčice	15	50.37	234
5	Brno-Tuřany	16.69	49.16	241
6	Mošnov	18.12	49.69	251
7	Hradec Králové	15.84	50.17	278
8	Lučina	18.44	49.73	300
9	Kuchařovice	16.09	48.88	334
10	Valašské Meziříčí	17.97	49.46	334
11	Praha-Ruzyně	14.26	50.1	364
12	Liberec	15.02	50.77	398
13	Třeboň	14.77	49.01	429
14	Klatovy	13.3	49.39	430
15	Velké Meziříčí	16.01	49.35	452
16	Havlíčkův Brod	15.58	49.61	455
17	Královice	13.49	49.99	468
18	Cheb	12.39	50.07	471
19	Husinec	13.99	49.04	536
20	Kostelní Myslová	15.44	49.16	569
21	Svratouch	16.03	49.74	737
22	Přimda	12.68	49.67	742
23	Červená	17.54	49.77	750
24	Churáňov	13.61	49.07	1118
25	Lysá Hora	18.44	49.54	1324

FIGURE 5.3: Locations of the Czech weather stations (maintained by the Czech Hydrometeorological Institute; http://www.chmi.cz/) providing data for the experiments with estimation of daily temperatures from other concurrent records. Numerical identifiers of the stations correspond to their ranks in the elevation-ordered list on the right.

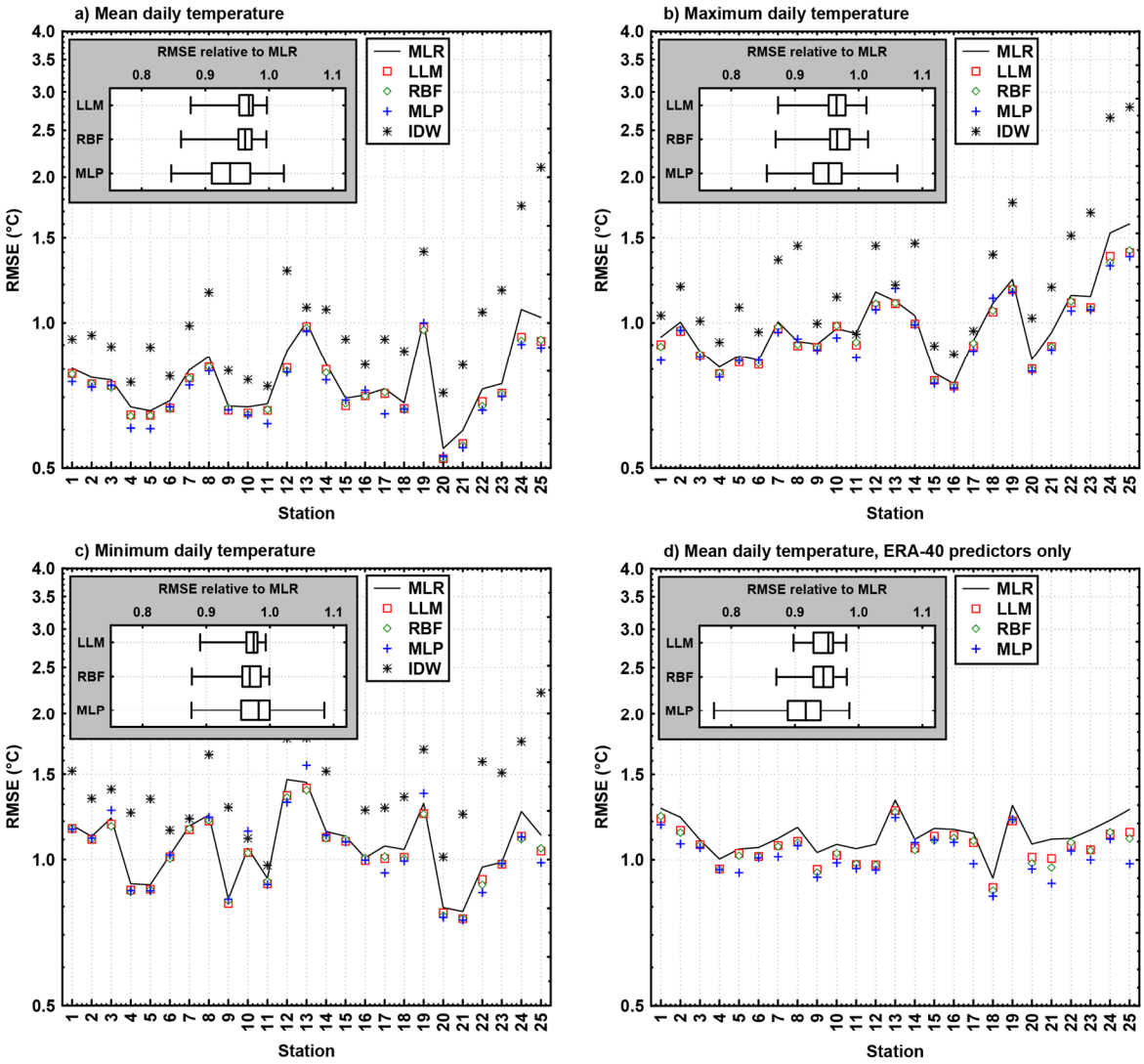


FIGURE 5.4: Root mean squared error (RMSE) of daily temperature estimation, carried out for 25 Czech weather stations by different regression techniques (multiple linear regression: MLR; local linear models: LLM; RBF neural network: RBF; multilayer perceptron neural network: MLP). The results are shown for daily mean **(a)**, maximum **(b)** and minimum **(c)** temperature computed using temperature series from the other observational sites and ERA-40 data as predictors, as well as for daily mean temperature computed from ERA-40 data only **(d)**. Identifiers of individual stations correspond to their numerical IDs in Fig. 5.3. The embedded boxplots show distributions of RMSE achieved by the nonlinear regression methods relative to linear regression in the set of all 25 stations: whiskers represent min-max range of the values, the box encloses values between lower and upper quartile, the central line corresponds to the median. Errors of temperature estimation by inverse distance weighting interpolation (IDW) are also shown.

CHAPTER 6

TREND AND ATTRIBUTION ANALYSIS

Among the problems studied by contemporary climatology, a prominent place belongs to detection of spatiotemporal variability within the climate system and its attribution, i.e. identification of the factors responsible. Particular attention has lately been devoted to investigation of the long-term components in the climatic series, and distinguishing between their natural and anthropogenic causes. In this chapter, several examples of our contributions to the related problems are given, including detection of trends in the total ozone data (Chap. 6.1; KRIŽAN ET AL. 2011* - APPENDIX V), attribution of temporal variability in annual and monthly temperature and precipitation series (Chap. 6.2; BRÁZDIL ET AL. 2012A*; MIKŠOVSKÝ ET AL. 2014* - APPENDIX VI; MIKŠOVSKÝ & PIŠOFT 2015*) and attribution of temporal variability in the series of drought indices (Chap. 6.3; BRÁZDIL ET AL. 2015B* - APPENDIX VII).

6.1 TRENDS IN TOTAL OZONE SERIES

Due to the intense interest in long-term changes in the climate system, estimation of magnitude and significance of trends in various observed and simulated series has become one of the most common climatological tasks. The problem is often approached by simple linear regression, matching the variable of interest against time. However, not all long-term components can be treated as strictly linear, and more complex characterization may be preferable, or outright necessary. One of the possible examples of a more intricate behavior is the evolution of total ozone amounts, shaped not only by natural forcings, but also by the impacts of human activity. In particular, the effect of man-produced halocarbons has been shown to be the cause of the rapid weakening of the ozone layer since the 1970s, with a follow-up recovery occurring since the mid-1990s due to the measures introduced by the Montreal Protocol. In KRIŽAN ET AL. (2011*), we investigated the presence of trends in the series of annual amplitude of total ozone over the 1979-2008 period (obtained from the TEMIS database: VAN DER A ET AL. 2010), with special attention to their geographical patterns across the northern midlatitudes. Piecewise, two-segmented linear regression was employed for trend detection, with break point in the year 1995. Existence of decreasing trends prior to 1995 and their reversal afterwards was confirmed in much of the target area, and discussed with regards to the behavior of annual minima and maxima of the ozone amounts. A relation was also suggested between the temporal evolution of the ozone characteristics and changes in the Brewer-Dobson circulation.

6.2 ATTRIBUTION OF TEMPORAL VARIABILITY OF TEMPERATURE AND PRECIPITATION

Of variables defining the state of the climate system, air temperature is perhaps the most intensely studied. Yet, despite the concentrated attention aimed at various thermal characteristics of the atmosphere, their behavior still remains only partly understood. Even variability in a single series of local temperature can be quite complicated (as illustrated in Fig. 2.1 for the Czech temperature series), and this complexity becomes even more overwhelming when spatial structures are taken into account. Identifying and quantifying the effects of individual climate-forming agents (and their eventual interactions) is a process often approached by statistical methods, including various forms of regression mappings (see, e.g., the introductory section in MIKŠOVSKÝ ET AL. 2014[*]). Several examples of our efforts in this field are presented here, demonstrating the application of statistical attribution analysis to various local and global temperature series and the insights that can obtained about the role of individual climate forcings.

Our first take on the issue of statistical attribution of temperature variability was aimed at the series of mean annual Czech temperature over the 1860-2008 period. The results were published as a part of the monograph BRÁZDIL ET AL. (2012A[*]). The temperature series investigated was created from measurements gathered at 10 Czech weather stations, quality-controlled and subjected to a homogenization procedure (BRÁZDIL ET AL. 2012A[*],B). Motivated by the prior attribution studies concerned with identification of the imprints of external and internal forcings in the temperature data (particularly by SCHÖNWIESE ET AL. 2010), we used multiple linear regression and multilayer perceptron neural network to detect temperature components related to the concentration of greenhouse gases (GHG), amounts of sulfate aerosols and solar activity, as well as to the phases of the Southern Oscillation (SO) and the North Atlantic Oscillation (NAO). Relatively prominent slow-variable components correlated with greenhouse gases concentration and sulfate amounts were found, along with a weaker imprint of solar activity. NAO proved to be an important driver of the inter-annual temperature variability, whereas the component attributed to SO was substantially weaker. Attention was also paid to the possibility of nonlinearities in the relations studied: Application of MLP neural network instead of basic linear regression resulted in just about 2% decrease of total RMSE (using the same set of predictors for both regression techniques), and the respective regression-based temperature estimates were found to be almost identical (Fig. 6.1).

The conclusions in BRÁZDIL ET AL. (2012A[*]) highlighted some possible connections between Czech temperature and climate forcings, both external and internal. However, the underlying analysis was somewhat rudimentary, and it neglected potentially critical aspects of the attribution problem such as assessment of the statistical significance of the relations, possibility of time-delayed responses or seasonal specifics of the links. In MIKŠOVSKÝ ET AL. (2014[*]), we therefore revisited the matter of regression-based attribution analysis in more depth. Monthly series of temperature were studied alongside their annual means,

and results for the Czech Republic were also compared to their counterparts derived from pan-European and global land temperature series (supplied from the Berkeley Earth dataset: ROHDE ET AL. 2013A,B). Statistical significance of the regression coefficients was tested by the moving block bootstrap (e.g. FITZENBERGER 1998). Our results confirmed the existence of a strong formal match between the long-term warming trends in temperature and concentration of greenhouse gases, and a weaker (and typically statistically non-significant) cooling tendency associated with sulfate aerosols. Only weak effects of solar activity were detected in any of the temperature series investigated. We also found no clear imprint of volcanic activity in the Czech (or European) temperatures, in contrast to a distinct temporary post-eruption cooling in global land temperature. Of the internal climatic oscillations, NAO was confirmed to be one of the dominant sources of shorter-term variability in the European region, whereas contributions from the Southern Oscillation, albeit noticeable, were only borderline statistically significant. A weakly significant component in the Czech temperature was also detected for the phase of the Atlantic Multidecadal Oscillation (AMO). Considering the approximate 70-year periodicity of AMO's main cycle, this result brings some interesting implications regarding the longer-term oscillations in the temperature signals, as further discussed in the concluding remarks in Chap. 7.

In addition to the temperature data, Czech precipitation series and their possible relations to the climate forcings were also investigated. The respective results in BRÁZDIL ET AL. (2012A*) and MIKŠOVSKÝ ET AL. (2014*) demonstrated that, unlike for temperatures, only very small fraction of total variance could be explained by any form of the regression model (7% or less, compared to up to 53 % for Czech annual temperature and 20% for Czech monthly temperature). NAO phase was found to be the only predictor contributing a statistically significant component to the precipitation series.

In MIKŠOVSKÝ ET AL. (2014*), too, nonlinear regression models were applied in addition to multiple linear regression, in an attempt to reveal (and hopefully isolate and quantify) the eventual nonlinear interactions among the predictors and temperature/precipitation. None of the nonlinear mappings did, however, clearly outperform multiple linear regression in terms of the fraction of variance explained. Hence, while some of the climate responses to the forcing factors may be inherently nonlinear, their combined manifestations in the monthly and annual series of temperature seem to be approximated fairly well by purely linear superposition. This finding facilitates the application of statistical attribution analysis to the gridded temperature datasets: The computational costs of nonlinear regression models would become quite prohibitive for predictands numbering thousands or tens of thousands (especially when combined with computationally intensive techniques for estimation of statistical significance such as bootstrap), but the analysis is still manageable using basic linear regression. In our follow-up work, we therefore turned our attention to various forms of gridded temperature data, using multiple linear regression alone as the tool for separating components associated with individual climate forcings.

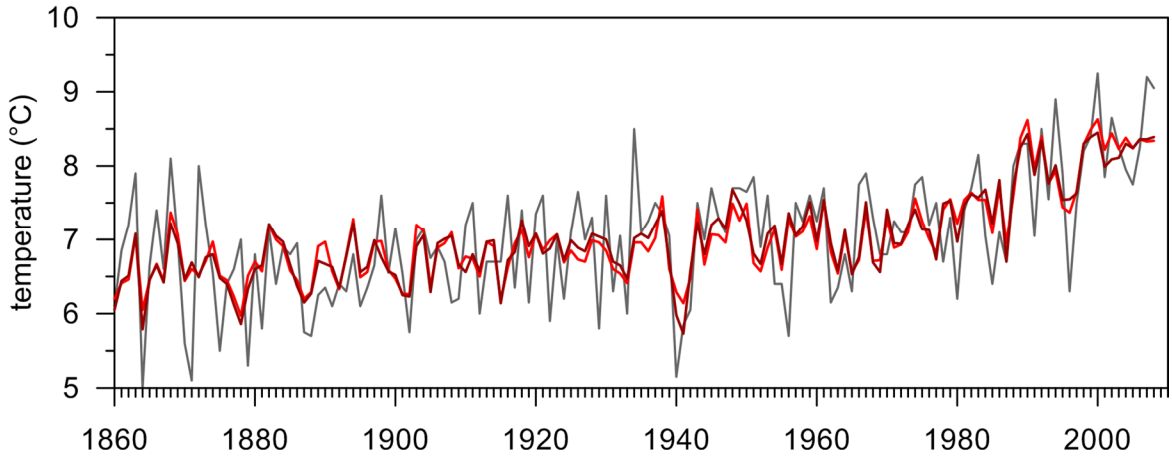


FIGURE 6.1: Annual mean areal Czech temperature in the 1860-2008 period, observed (**grey line**) and approximated by multiple linear regression (**lighter red line**) and multilayer perceptron neural network (**darker red line**) from a set of explanatory variables representing various external and internal climate forcings, described in BRÁZDIL ET AL. (2012A*) and based on the setup by SCHÖNWIESE ET AL. (2010). Adapted from BRÁZDIL ET AL. (2012A*).

In MIKŠOVSKÝ & PIŠOFT (2015*), we investigated the presence of imprints of various climate forcings in the series of gridded monthly temperature anomalies throughout the European region, supplied from the GISTEMP (HANSEN ET AL. 2010) and Berkeley Earth (ROHDE ET AL. 2013A,B) datasets. Multiple linear regression in its basic form (Eq. 1) was applied to identify the links between local temperature anomalies and selected predictors with established or suspected influence on the European weather and climate (Fig. 6.2). Statistical significance of the regression coefficients was assessed by moving block bootstrap. While not included in MIKŠOVSKÝ & PIŠOFT (2015*), the tests were also carried out for temperatures provided by the 20th Century Reanalysis (COMPO ET AL. 2011). This allowed for a comparison of the forcing fingerprints in the gridded observations and in a reanalysis dataset (which, in the particular case of the 20th Century Reanalysis, does not use temperatures from the land-based stations as inputs). In Fig. 6.3, our results are summarized in the form of responses of temperature to the pre-selected characteristic variations of the predictors, specified in the caption of Fig. 6.2. Some of the previously established effects of the climate forcings have been confirmed by our analysis. These included the universally strong, yet locally variable correlation between the GHG concentrations and the long-term temperature component, or the presence of a distinct response pattern related to the North Atlantic Oscillation. Some interesting outcomes regarding the effects of external forcings or teleconnections projected by internal variability modes also appeared (such as the association between the Pacific Decadal Oscillation and temperatures in Scandinavia). Furthermore, our analysis highlighted some of the uncertainties potentially stemming from the choice of the target temperature dataset. Of particular interest was the noticeable difference between the compo-

nents attributed to the greenhouse gases concentration in the GISTEMP and Berkeley Earth datasets and in the 20th Century Reanalysis. This contrast, symptomatic of the potential problems with the long-term temperature trends in the 20th Century Reanalysis (COMPO ET AL. 2013), serves as a cautionary example of the specifics of the reanalysis-type data, frequently employed as representatives of the real climate, yet often carrying a distinct signature of the numerical model involved in their creation and of the selection of its inputs. In our ongoing and future research, we will study these issues in greater depth, as discussed in the concluding remarks in Chap. 7.

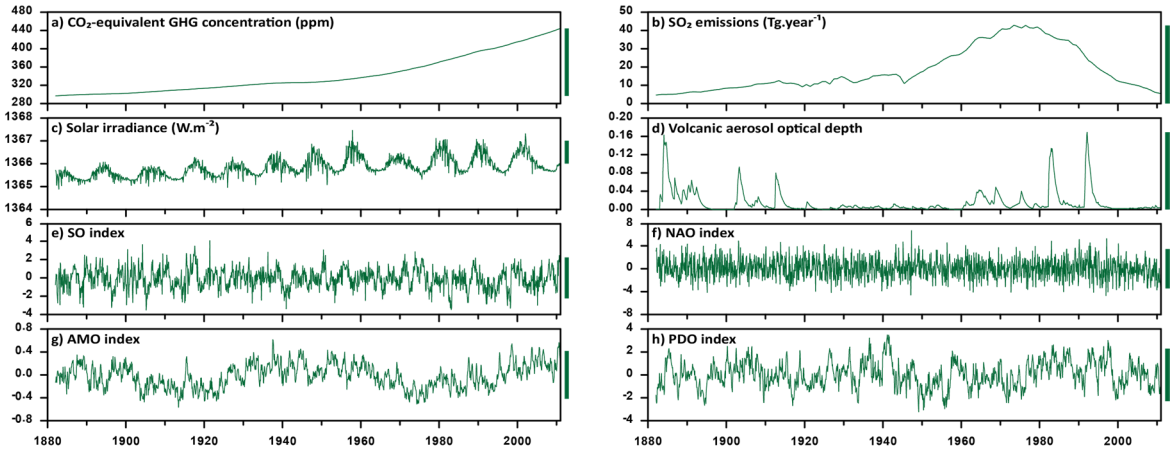


FIGURE 6.2: Time series of the explanatory variables employed in the attribution analysis in MIKŠOVSKÝ & PIŠOFT (2015*): CO_2-equivalent concentration of Kyoto protocol-controlled greenhouse gases (GHG), obtained from http://www.pik-potsdam.de/~mmalte/rcps/ (MEINSHAUSEN ET AL. 2011) **(a)**; European SO_2 emissions adapted from the data by SMITH ET AL. (2011) as a proxy for the amounts of anthropogenic sulfate aerosols **(b)**; monthly solar irradiance from http://climexp.knmi.nl/data/itsi_wls_mon.dat (WANG ET AL. 2005) **(c)**; volcanic aerosol optical depth from http://data.giss.nasa.gov/modelforce/strataer/ (SATO ET AL. 1993) **(d)**; Southern Oscillation (SO - e.g. TRENBERTH ET AL. 2002) index **(e)** and North Atlantic Oscillation (NAO - e.g. HURRELL ET AL. 2003) index **(f)** from CRU (http://www.cru.uea.ac.uk/cru/data/pci.htm); Atlantic Multidecadal Oscillation (AMO - e.g. ENFIELD ET AL. 2001) index from NOAA (http://www.esrl.noaa.gov/psd/data/timeseries/AMO/) **(g)**; Pacific Decadal Oscillation (PDO - e.g. ZHANG ET AL. 1997) index from KNMI Climate Explorer (http://climexp.knmi.nl/data/ipdo_erssta.txt) **(h)**. Green bars to the right of individual panels illustrate the size of the characteristic variation Δp_i of the predictor, used for calculation of the temperature responses shown in Fig. 6.3: Increase of the CO_2-equivalent GHG concentration between 1882 and 2010 (+148 ppm); peak value of the European SO_2 emissions (43 $Tg.year^{-1}$); increase of the solar irradiance by 1 $W.m^{-2}$; Mt. Pinatubo-sized volcanic event; increase of SO, NAO, AMO and PDO indices by four times their standard deviation. Modified from MIKŠOVSKÝ & PIŠOFT (2015*).

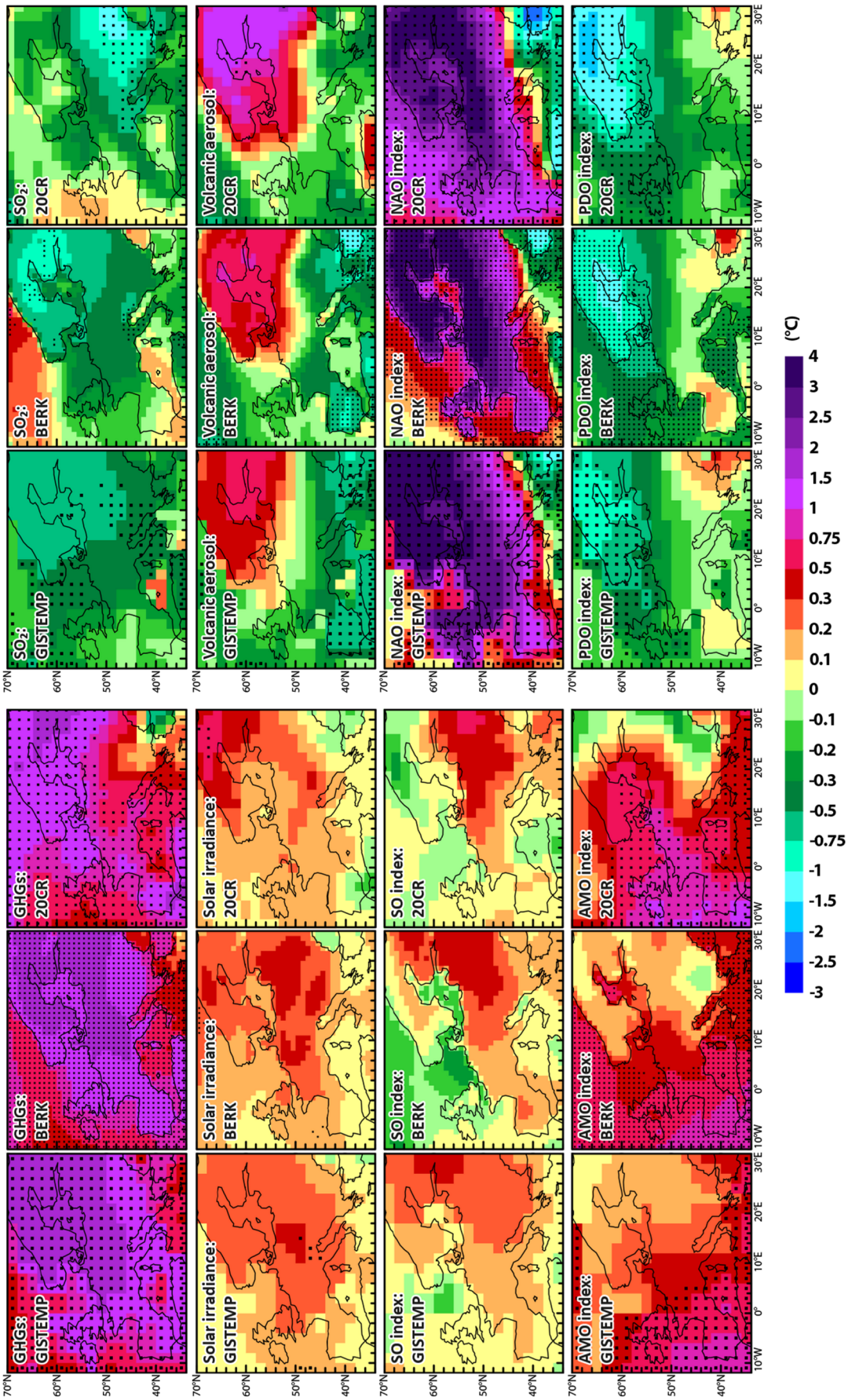

GHGs: GISTEMP
GHGs: BERK
GHGs: 20CR
SO2: GISTEMP
SO2: BERK
SO2: 20CR
Solar irradiance: GISTEMP
Solar irradiance: BERK
Solar irradiance: 20CR
Volcanic aerosol: GISTEMP
Volcanic aerosol: BERK
Volcanic aerosol: 20CR
SO index: GISTEMP
SO index: BERK
SO index: 20CR
NAO index: GISTEMP
NAO index: BERK
NAO index: 20CR
AMO index: GISTEMP
AMO index: BERK
AMO index: 20CR
PDO index: GISTEMP
PDO index: BERK
PDO index: 20CR
70°N
60°N
50°N
40°N
10°W
0°
10°E
20°E
30°E
(°C)
4
3
2.5
2
1.5
1
0.75
0.5
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.5
-0.75
-1
-1.5
-2
-2.5
-3

FIGURE 6.3 (🡐): Geographic patterns of local temperature response (°C) associated with various explanatory variables, calculated as a product of the regression coefficient a_i (computed individually for each grid point by multiple linear regression) and the characteristic variation Δp_i of the respective predictor (specified in Fig. 6.2). Monthly temperature anomalies from the GISTEMP (HANSEN ET AL. 2010), Berkeley Earth (BERK: ROHDE ET AL. 2013A,B) and 20th Century Reanalysis (20CR: COMPO ET AL. 2011) datasets were analyzed, for the 1882-2010 period. Statistical significance of the components associated with individual predictors was evaluated by moving-block bootstrap (e.g. FITZENBERGER 1998) – black dots mark grid points with response statistically significant at the 99% confidence level. Modified from MIKŠOVSKÝ & PIŠOFT (2015*) and expanded with the 20CR-based results.

6.3 ATTRIBUTION OF TEMPORAL VARIABILITY OF DROUGHTS

While temperature and precipitation totals are among the most basic – and most intensely studied – climatic descriptors, more intricate composite characteristics are often used to capture effects of the processes within the climate system. Drought indices, constructed to describe the degree of wet or dry conditions within some locality of interest, are one particular class of such impact-focused quantities. In BRÁZDIL ET AL. (2015B*), we examined selected aspects of various short- and long-term indices quantifying meteorological droughts in the Czech Republic. The analysis focused on spring and summer as the seasons most relevant to the drought impacts on agriculture. As a part of this assessment, we investigated presence of links between the time series of individual drought indices and several explanatory variables related to climate forcings. Attention was paid to the possible manifestations of the man-induced changes to the atmospheric composition (particularly increasing concentrations of the greenhouse gases) and to the effects of natural forcings (solar and volcanic activity). Existence of components correlated with the phase of the North Atlantic Oscillation, Southern Oscillation or Atlantic Multidecadal Oscillation was also assessed. Substantial and statistically significant connection between droughts and anthropogenic forcing was found for the indices involving temperature or evapotranspiration in their definition. Links to NAO were generally strong and a tendency towards drier conditions has been confirmed for the positive NAO phase. Possible influence of the Southern Oscillation was found as well, though weaker and only statistically significant for some of the indices and seasons. Once again, nonlinear regression techniques were used alongside multiple linear regression, and, just as it was the case for annual and monthly temperature and precipitation, only minor gain from nonlinear attribution models was indicated. Feasibility of the linear approach to the attribution analysis was therefore upheld in this case, too.

CHAPTER 7
CONCLUDING REMARKS AND FUTURE PROSPECTS

The individual pieces of analysis shown throughout this text have demonstrated a few examples of the wide range of possible applications of the regression mappings in the atmospheric and climatic sciences. The results achieved, diverse in their aims, methods used and datasets involved, can obviously not be summarized by a simple, all-encompassing conclusion. There are, however, a few points worthy of mentioning, related to the individual topics here as well as their common aspects. Furthermore, some details on the future directions of the related research are given in this concluding section, connecting the results presented with the expected outcomes of other ongoing studies with my participation.

First, our experiments have affirmed that nonlinearity, while inherent to the climate system, manifests with varying level of intensity in the climatic time series. It would be too daring to try to formulate specific guidelines pinpointing scenarios suitable for application of nonlinear regression techniques. Our results as well as those of various other studies devoted to this topic have highlighted numerous factors potentially affecting the level of observable nonlinearity. Additional ambiguity can be brought by technical specifics or imperfections of the data themselves, such as the presence of non-climatic inhomogeneities. Still, there seem to be some general factors tied to the superiority of the nonlinear approach (or lack thereof). In our analyses, nonlinear mappings proved best suited for setups with intermediate complexity and enough information provided in the signals analyzed to extract the respective relationships. In the absence of nontrivial, low-dimensional links between predictors and predictand (for instance when forecasting daily temperatures more than a few days ahead), no benefit stemmed from the use of nonlinear models. On the other hand, diminished degree of detectable nonlinearity was also characteristic of setups with very tight linear correlation between the predictand and one or more predictors, when most of the information could be transmitted through a purely linear function, leaving only small fraction of total variance unexplained and available for the extra contribution from a nonlinear mapping. Of the tasks studied in this thesis, superiority of nonlinear regression was often indicated for short-term prediction of climatic variables at daily time scales, downscaling of daily temperature from large-scale reanalysis data or daily temperature estimation from records taken at nearby observational sites. Even then, the gain was not automatically guaranteed for each individual test configuration. This means that the key problem – identification of setups suitable for the application of the nonlinear approach – needs to be treated on a case-by-case basis. Occasionally, pointers are available that can help to make the decision. In particular, presence of characteristic markers of low-dimensional deterministic chaos in the data is often an indicator of capturable nonlinear links. Reliable detection of relevant low-dimensional chaotic behavior

from the (pseudo)observed climatic time series can however be quite challenging, at least at the spatiotemporal scales investigated here. Practical conclusions about suitability of nonlinear techniques are therefore often easiest to reach by direct tests focused on the particular performance criteria of interest, sometimes basically resorting to a 'brute force' approach. The same then also applies to the selection of the best performing type of the regression mapping and its optimum configuration.

While the respective research may be somewhat overshadowed in volume by the topics related to the effects of the climate change, interest in manifestations of nonlinear and chaotic behavior in the climatic data still persists. In addition to the studies focused on region- or phenomena-specific problems, attention is being paid to the broader issues of links within the climate system and their geographic structures. Elements of our past research in this area were shown here (such as the nonlinearity patterns from MIKŠOVSKÝ ET AL. 2008*). Some more recent contributions to the related issues then operate with the concept of climate networks, allowing for study of generalized climatic teleconnections, and possible evaluation of their nonlinearity (e.g. HLINKA ET AL. 2013 and the sources within). Attention is also still being devoted to the assessment and quantification of the atmospheric chaoticity and predictability (e.g. LI & DING 2011; BADIN & DOMEISEN 2014). With J. Skořepa, my Ph.D. student, we currently investigate a related issue, with foreseeable results about to bring new insights into the matters of global and local chaoticity measures, their geographic patterns and their representation in various types of climatic data.

As shown in Chap. 5.1 for various types of daily temperature, statistical downscaling tasks do often benefit from the nonlinear approach to the construction of the regression functions. However, the gain is not automatically assured, and some validation statistics show no systematic improvement from the use of a nonlinear regression model, as we demonstrated (HUTH ET AL. 2015*) and as was also suggested by other studies with similar focus (e.g. HUTH ET AL. 2008). There are, however, additional matters in need of attention. Besides the critical – yet occasionally neglected – issue of the predictor selection (e.g. HUTH 2004), stability of the downscaling mappings must be carefully verified. This becomes particularly crucial when GCM outputs for the future time periods are downscaled, with predictors possibly falling outside the range of values typical for the past climate, and thus unprecedented in the data employed for calibration of the downscaling models. Bearing these caveats in mind, statistical downscaling by regression mappings – regardless of the specific methodology – remains a valuable tool for bypassing the resolution gap between global climate simulations and local-scale data. With further improvement of the spatial step of global climate models, and reduction of their still considerable biases, need for statistical downscaling (and also for statistical postprocessing, reducing the systematic errors in the GCM/RCM simulations – e.g. DÉQUÉ 2007; THEMEßL ET AL. 2012) may eventually disappear. Today, however, we are still far from such future, and statistical downscaling models do currently represent a viable alternative to their dynamical counterparts.

In our attribution-centered analyses (BRÁZDIL ET AL. 2012A*; MIKŠOVSKÝ ET AL. 2014*; BRÁZDIL ET AL. 2015B*; Chap. 6.2), searching for components associated with particular climate forcings in the climatic series at monthly and annual time step, only minor gain from application of nonlinear mappings was detected. Even in the prior studies concerned with similar problems, and reporting presence of nonlinearities, the magnitude of the nonlinear components was rather variable and inferiority of linear models not guaranteed for every test setting (PASINI ET AL. 2006; SCHÖNWIESE ET AL. 2010). The lack of distinct nonlinear components seems to also extend to more general links among the climatic time series (e.g. HLINKA ET AL. 2014), indirectly implying a predominantly linear character of some of the teleconnections considered in the attribution analyses. GCM-based experiments, too, have suggested that linear superposition of the forcings may often suffice to approximate their combined effects, at least for the external forcing factors (e.g. SHIOGAMA ET AL. 2013). Altogether, these pieces of evidence indicate that approaching the problem of attribution analysis via linear methods is not unreasonable, and linear approximation can be applied without incurring excessive oversimplification. That said, further research into the related matters is still highly desirable, especially regarding the potential intricacies of the interactions of internal climate variability modes.

Regarding the problem of attribution, several additional issues ought to be taken into account. First of all, statistical attribution analysis only considers formal similarities among the series analyzed, oblivious of the underlying physical dependencies (or their absence). Dangers of misrepresentation of the outcomes of the statistical approach to attribution have been highlighted in the past (e.g. BENESTAD & SCHMIDT 2009); we only very briefly touched upon this subject in MIKŠOVSKÝ ET AL. (2014*) and BRÁZDIL ET AL. (2015B*), but it is definitely worthy of – and in need of – much deeper attention in the future. Another related challenge then consists in selection of suitable explanatory variables, which should represent all relevant climate-forming factors, but without unnecessarily increasing dimensionality of the regression model, and without imposing redundancies. Besides the anthropogenic forcings (related especially to the amounts of greenhouse gases, but also to the effects of various aerosols – e.g. SKEIE ET AL. 2011; STOCKER ET AL. 2013, CHAP. 8) and natural external forcings (especially variations of solar and volcanic activity – e.g. SCHMIDT ET AL. 2011), internal variability modes in the climate system do also shape its temporal evolution and should be considered in the attribution analysis. Of particular interest regarding the mid- to long-term climatic trends is the natural multidecadal variability, manifesting for instance through Atlantic Multidecadal Oscillation (AMO – e.g. ENFIELD ET AL. 2001) or Pacific Decadal Oscillation (PDO – e.g. ZHANG ET AL. 1997). The respective oscillatory phenomena, related mutually as well as to other variability modes in the climate system (e.g. WU ET AL. 2011; WYATT ET AL. 2012), have been implicated as possible major contributors to longer-term temperature variability. In our analyses (MIKŠOVSKÝ ET AL. 2014*; MIKŠOVSKÝ & PIŠOFT 2015*), as well as various other recent studies (e.g. ROHDE ET AL. 2013B; ZHOU & TUNG 2013; CHYLEK ET AL. 2014), a distinct formal link between the AMO phase and

local or global temperature has been shown. This relationship needs to be interpreted carefully though – the question of AMO's involvement in the climate variability is still open and intensely discussed, regarding the origin of the variations observed (e.g. BOOTH ET AL. 2012; ZHANG ET AL. 2013), optimum way of their characterization and relations to other forcings (e.g. CANTY ET AL. 2013) and implications for the magnitude of the observed trends of temperature (e.g. VAN DER WERF & DOLMAN 2014). Final conclusions in these matters are still far from established and unlikely to be reached by statistical means alone (due to, in part, limited extent of the available records compared to the approximately 70-year periodicity of AMO). Numerical modelling perspective may eventually help to provide a clearer picture in the future. However, the ability of the current generation of climate models to capture the multidecadal oscillations (including AMO) is still limited and quite variable among the individual simulations (e.g. BA ET AL. 2014). Several of our current and planned research activities hope to further contribute to these intensely researched topics. In our upcoming paper MIKŠOVSKÝ ET AL. (2015*), a more complex approach to the issue of attribution will be embraced, with particular attention to the world-wide spatial fingerprints of external and internal forcings in the (pseudo)observed and reanalyzed temperature data. In the wake of this study, manifestations of the forcings will also be investigated in the historical runs of global and regional climate models, to assess the ability of these numerical simulations to reproduce the observation-based imprints of individual forcings. We also continue the investigation of the factors affecting the occurrence of droughts in central Europe, further building upon the foundations laid in BRÁZDIL ET AL. (2015B*). Outcomes of the next stage of this analysis will appear as a part of the drought-focused volume BRÁZDIL ET AL. (2015A*). Results derived from drought indices specific to individual locations within the Czech territory will be presented, for all seasons of the year, and complemented by a more detailed assessment of statistical significance of individual forcing-attributed components. We also recently revisited the issue of attribution of temporal variability in the middle atmosphere, with particular focus on regression-based identification of the imprints of solar activity in various reanalysis datasets. The resulting paper (KUCHAŘ ET AL. 2015*) delivers a detailed study of the solar forcing effects on the stratospheric temperature, ozone amounts and circulation characteristics (though it should be noted that my contribution was rather minor in its preparation, and the paper was therefore not included as a core part of this thesis, despite the topical compatibility).

In Chapter 2, the importance of carefully considering the specific properties of the datasets analyzed was emphasized. Here, I would like to accentuate this issue once more: Even from the few examples given throughout this text, it should be evident that substantial variation in the outcomes (and conclusions) of a given analysis may stem from the selection of the input data. This impact may be particularly severe in the case of the attribution-focused tasks: Not only do the target variables (such as temperature) exist in alternative versions, but multiple – and sometimes relatively disparate – options are also often available as representatives of the explanatory variables (such as the estimates of the levels of solar or

volcanic activity – e.g. SCHMIDT ET AL. 2011, 2012). In the aforementioned upcoming paper MIKŠOVSKÝ ET AL. (2015*), the issue of the specifics tied to different versions of the input data will be investigated more thoroughly, particularly regarding various representatives of past temperature. Special attention will be devoted to the behavior of the 20th Century Reanalysis – the increasingly popular source of climatic data, spanning more than a century, which however also exhibits some noteworthy deviations from the datasets based on gridded observations (as illustrated in Figs. 2.2 or 6.3). In MIKŠOVSKÝ ET AL. (2015*), these contrasts will be studied in more detail and on a global scale, including a comprehensive assessment of the specifics tied to the effects of individual climate-forming factors.

REFERENCES

BA, J., ET AL. (2014), A multi-model comparison of Atlantic multidecadal variability, *Climate Dynamics*, *43*(9-10), 2333-2348, doi:10.1007/s00382-014-2056-1.

BADIN, G., AND D. I. V. DOMEISEN (2014), A Search for Chaotic Behavior in Northern Hemisphere Stratospheric Variability, *Journal of the Atmospheric Sciences*, *71*(4), 1494-1507, doi:10.1175/jas-d-13-0225.1.

BENESTAD, R. E., AND G. A. SCHMIDT (2009), Solar trends and global warming, *Journal of Geophysical Research-Atmospheres*, *114*, D14101, doi:10.1029/2008jd011639.

BOOTH, B. B. B., N. J. DUNSTONE, P. R. HALLORAN, T. ANDREWS, AND N. BELLOUIN (2012), Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability, *Nature*, *484*(7393), 228-U110, doi:10.1038/nature10946.

BRÁZDIL, R., M. BĚLÍNOVÁ, P. DOBROVOLNÝ, J. MIKŠOVSKÝ, P. PIŠOFT, L. ŘEZNÍČKOVÁ, P. ŠTĚPÁNEK, H. VALÁŠEK, AND P. ZAHRADNÍČEK (2012A), *Temperature and precipitation fluctuations in the Czech Lands during the instrumental period*, Masaryk University, Brno, 236 pp., ISBN 978-80-210-6052-4.

BRÁZDIL, R., ET AL. (2015A), *Sucho v českých zemích: minulost, současnost, budoucnost (Drought in the Czech Lands: Past, Present and Future)*, Global Change Research Centre ASCR, Brno, 406 pp., ISBN 978-80-87902-11-0 (in print, in Czech with English summary).

BRÁZDIL, R., M. TRNKA, J. MIKŠOVSKÝ, L. ŘEZNÍČKOVÁ, AND P. DOBROVOLNÝ (2015B), Spring-summer droughts in the Czech Land in 1805–2012 and their forcings, *International Journal of Climatology*, *35*, 1405-1421, doi:10.1002/joc.4065.

BRÁZDIL, R., P. ZAHRADNÍČEK, P. PIŠOFT, P. ŠTĚPÁNEK, M. BĚLÍNOVÁ, AND P. DOBROVOLNÝ (2012B), Temperature and precipitation fluctuations in the Czech Republic during the period of instrumental measurements, *Theoretical and Applied Climatology*, *110*(1-2), 17-34, doi:10.1007/s00704-012-0604-3.

CANTY, T., N. R. MASCIOLI, M. D. SMARTE, AND R. J. SALAWITCH (2013), An empirical model of global climate - Part 1: A critical evaluation of volcanic cooling, *Atmospheric Chemistry and Physics*, *13*(8), 3997-4031, doi:10.5194/acp-13-3997-2013.

CHYLEK, P., J. D. KLETT, G. LESINS, M. K. DUBEY, AND N. HENGARTNER (2014), The Atlantic Multidecadal Oscillation as a dominant factor of oceanic influence on climate, *Geophysical Research Letters*, *41*(5), 1689-1697, doi:10.1002/2014gl059274.

COMPO, G. P., ET AL. (2011), The Twentieth Century Reanalysis Project, *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1-28, doi:10.1002/qj.776.

COMPO, G. P., P. D. SARDESHMUKH, J. S. WHITAKER, P. BROHAN, P. D. JONES, AND C. MCCOLL (2013), Independent confirmation of global land warming without the use of station temperatures, *Geophysical Research Letters*, *40*(12), 3170-3174, doi:10. 1002/grl.50425.

DÉQUÉ, M. (2007), Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, *Global and Planetary Change*, *57*(1-2), 16-26, doi:10.1016/ j.gloplacha.2006.11.030.

ENFIELD, D. B., A. M. MESTAS-NUNEZ, AND P. J. TRIMBLE (2001), The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US, *Geophysical Research Letters*, *28*(10), 2077-2080, doi:10.1029/2000gl012745.

FARDA, A., M. DEQUE, S. SOMOT, A. HORANYI, V. SPIRIDONOV, AND H. TOTH (2010), Model ALADIN as regional climate model for Central and Eastern Europe, *Studia Geophysica Et Geodaetica*, *54*(2), 313-332, doi:10.1007/s11200-010-0017-7.

FITZENBERGER, B. (1998), The moving blocks bootstrap and robust inference for linear least squares and quantile regressions, *Journal of Econometrics*, *82*(2), 235-287, doi:10.1016/ s0304-4076(97)00058-4.

GORDON, C., C. COOPER, C. A. SENIOR, H. BANKS, J. M. GREGORY, T. C. JOHNS, J. F. B. MITCHELL, AND R. A. WOOD (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Climate Dynamics*, *16*(2-3), 147-168, doi:10.1007/s003820050010.

HALENKA, T., J. KALVOVA, Z. CHLADOVA, A. DEMETEROVA, K. ZEMANKOVA, AND M. BELDA (2006), On the capability of RegCM to capture extremes in long term regional climate simulation - comparison with the observations for Czech Republic, *Theoretical and Applied Climatology*, *86*(1-4), 125-145, doi:10.1007/s00704-005-0205-5.

HANSEN, J., R. RUEDY, M. SATO, AND K. LO (2010), GLOBAL SURFACE TEMPERATURE CHANGE, *Reviews of Geophysics*, *48*, 29, doi:10.1029/2010rg000345.

HAYKIN, S. (1999), *Neural Networks: A Comprehensive Foundation (2nd ed.)*, Prentice Hall, Upper Saddle River, 842 pp., ISBN 0-13-273350-1.

HEGGER, R., H. KANTZ, AND T. SCHREIBER (1999), Practical implementation of nonlinear time series methods: The TISEAN package, *Chaos*, *9*(2), 413-435, doi:10.1063/1.166424.

HLINKA, J., D. HARTMAN, M. VEJMELKA, D. NOVOTNA, AND M. PALUS (2014), Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity, *Climate Dynamics*, *42*(7-8), 1873-1886, doi:10.1007/s00382-013-1780-2.

HLINKA, J., D. HARTMAN, M. VEJMELKA, J. RUNGE, N. MARWAN, J. KURTHS, AND M. PALUS (2013), Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information, *Entropy*, *15*(6), 2023-2045, doi:10.3390/e15062023.

HURRELL, J. W., Y. KUSHNIR, G. OTTERSEN, AND M. VISBECK (EDS.) (2003), *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, American Geophysical Union, Washington, DC, 279 pp., ISBN 9780875909943.

HUTH, R. (2004), Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors, *Journal of Climate*, *17*(3), 640-652, doi:10.1175/1520-0442(2004)017<0640:soldtc>2.0.co;2.

HUTH, R., S. KLIEGROVA, AND L. METELKA (2008), Non-linearity in statistical downscaling: does it bring an improvement for daily temperature in Europe?, *International Journal of Climatology*, *28*(4), 465-477, doi:10.1002/joc.1545.

HUTH, R., J. MIKŠOVSKÝ, P. ŠTĚPÁNEK, M. BELDA, A. FARDA, Z. CHLÁDOVÁ, AND P. PIŠOFT (2015), Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature, *Theoretical and Applied Climatology*, *120*(3-4), 533-553, doi:10.1007/s00704-014-1190-3.

JARVIS, C. H., AND N. STUART (2001), A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: The interaction between number of guiding variables and the type of interpolation method, *Journal of Applied Meteorology*, *40*(6), 1075-1084, doi:10.1175/1520-0450(2001)040<1075:acasfi>2.0.co;2.

KEPPENNE, C. L., AND C. NICOLIS (1989), Global properties and local structure of the weather attractor over Western Europe, *Journal of the Atmospheric Sciences*, *46*(15), 2356-2370, doi:10.1175/1520-0469(1989)046<2356:gpalso>2.0.co;2.

KISTLER, R., ET AL. (2001), The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, *Bulletin of the American Meteorological Society*, *82*(2), 247-267, doi:10.1175/1520-0477(2001)082<0247:tnnyrm>2.3.co;2.

KLEIN TANK, A. M. G., ET AL. (2002), Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, *International Journal of Climatology*, *22*(12), 1441-1453, doi:10.1002/joc.773.

KRIŽAN, P., J. MIKŠOVSKÝ, M. KOZUBEK, W. GENGCHEN, AND B. JIANHUI (2011), Long term variability of total ozone yearly minima and maxima in the latitudinal belt from 20°N to 60°N derived from the merged satellite data in the period 1979-2008, *Advances in Space Research*, *48*(12), 2016-2022, doi:10.1016/j.asr.2011.07.010.

KUCHAŘ, A., P. ŠÁCHA, J. MIKŠOVSKÝ, AND P. PIŠOFT (2015), The 11-year solar cycle in current reanalyses: A (non)linear attribution study of the middle atmosphere, *Atmospheric Chemistry and Physics*, *15*, 6879-6895, doi:10.5194/acp-15-6879-2015.

LI, J. P., AND R. Q. DING (2011), Temporal-Spatial Distribution of Atmospheric Predictability Limit by Local Dynamical Analogs, *Monthly Weather Review*, *139*(10), 3265-3283, doi:10.1175/mwr-d-10-05020.1.

MEINSHAUSEN, M., ET AL. (2011), The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, *Climatic Change*, *109*(1-2), 213-241, doi:10.1007/s10584-011-0156-z.

MIKŠOVSKÝ, J. (2004), *On some meteorological applications of nonlinear time series analysis methods*. Ph.D. Thesis, Charles University, Prague, 87 pp.

MIKŠOVSKÝ, J., R. BRÁZDIL, P. ŠTĚPÁNEK, P. ZAHRADNÍČEK, AND P. PIŠOFT (2014), Long-term variability of temperature and precipitation in the Czech Lands: an attribution analysis, *Climatic Change*, *125*(2), 253-264, doi:10.1007/s10584-014-1147-7.

MIKŠOVSKÝ, J., E. HOLTANOVÁ, AND P. PIŠOFT (2015), Imprints of climate forcings in global gridded temperature data (in preparation).

MIKŠOVSKÝ, J., AND P. PIŠOFT (2015), Attribution of European temperature variability during 1882-2010: A statistical perspective, in *Global Change: A Complex Challenge* (Eds.: Urban, O., M. Šprtová, and K. Klem), Global Change Research Centre AS CR, Brno, 10-13, ISBN 978-80-87902-10-3 (in print).

MIKŠOVSKÝ, J., P. PIŠOFT, AND A. RAIDL (2008), Global Patterns of Nonlinearity in Real and GCM-Simulated Atmospheric Data, in *Nonlinear Time Series Analysis in the Geosciences: Applications in Climatology, Geodynamics and Solar-Terrestrial Physics* (Eds.: Donner, R. V., and S. M. Barbosa), *Lecture Notes in Earth Sciences*, *112*, 17-34, doi:10.1007/978-3-540-78938-3_2.

MIKŠOVSKÝ, J., AND A. RAIDL (2005), Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures, *Nonlinear Processes in Geophysics*, *12*(6), 979-991.

MIKŠOVSKÝ, J., AND A. RAIDL (2006), Testing for nonlinearity in European climatic time series by the method of surrogate data, *Theoretical and Applied Climatology*, *83*(1-4), 21-33, doi:10.1007/s00704-005-0130-7.

MORICE, C. P., J. J. KENNEDY, N. A. RAYNER, AND P. D. JONES (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *Journal of Geophysical Research-Atmospheres*, *117*, D08101, doi:10.1029/2011jd017187.

OTT, E., T. SAUER, AND Y. A. YORKE (EDS.) (1994), *Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems*, Wiley, New York, 418 pp., ISBN 978-0471025566.

PACKARD, N. H., J. P. CRUTCHFIELD, J. D. FARMER, AND R. S. SHAW (1980), Geometry from a Time Series, *Physical Review Letters*, *45*(9), 712-716, doi:10.1103/PhysRevLett.45.712.

PASINI, A., M. LORE, AND F. AMELI (2006), Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system, *Ecological Modelling*, *191*(1), 58-67, doi:10.1016/j.ecolmodel.2005.08.012.

ROHDE, R., R. MULLER, R. JACOBSEN, S. PERIMUTTER, A. ROSENFELD, J. WURTELE, J. CURRY, C. WICKHAM, AND S. MOSHER (2013A), Berkeley Earth Temperature Averaging Process, *Geoinformatics & Geostatistics: An Overview 1*(2), 1-13, doi:http://dx.doi.org/10.4172/2327-4581.1000103.

ROHDE, R., R. A. MULLER, R. JACOBSEN, E. MULLER, S. PERLMUTTER, A. ROSENFELD, J. WURTELE, D. GROOM, AND C. WICKHAM (2013B), A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, *Geoinformatics & Geostatistics: An Overview 1*(1), 1-7, doi:10.4172/2327-4581.1000101.

SATO, M., J. E. HANSEN, M. P. MCCORMICK, AND J. B. POLLACK (1993), STRATOSPHERIC

AEROSOL OPTICAL DEPTHS, 1850-1990, *Journal of Geophysical Research-Atmospheres*, *98*(D12), 22987-22994, doi:10.1029/93jd02553.

SCHMIDT, G. A., ET AL. (2011), Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), *Geoscientific Model Development*, *4*(1), 33-45, doi:10.5194/gmd-4-33-2011.

SCHMIDT, G. A., ET AL. (2012), Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), *Geoscientific Model Development*, *5*(1), 185-191, doi:10.5194/gmd-5-185-2012.

SCHÖNWIESE, C. D., A. WALTER, AND S. BRINCKMANN (2010), Statistical assessments of anthropogenic and natural global climate forcing. An update, *Meteorologische Zeitschrift*, *19*(1), 3-10, doi:10.1127/0941-2948/2010/0421.

SCHREIBER, T., AND A. SCHMITZ (1996), Improved surrogate data for nonlinearity tests, *Physical Review Letters*, *77*(4), 635-638, doi:10.1103/PhysRevLett.77.635.

SCHREIBER, T., AND A. SCHMITZ (2000), Surrogate time series, *Physica D*, *142*(3-4), 346-382, doi:10.1016/s0167-2789(00)00043-9.

SHIOGAMA, H., D. A. STONE, T. NAGASHIMA, T. NOZAWA, AND S. EMORI (2013), On the linear additivity of climate forcing-response relationships at global and continental scales, *International Journal of Climatology*, *33*(11), 2542-2550, doi:10.1002/joc.3607.

SIVAKUMAR, B. (2004), Chaos theory in geophysics: past, present and future, *Chaos Solitons & Fractals*, *19*(2), 441-462, doi:10.1016/s0960-0779(03)00055-9.

SMITH, S. J., J. VAN AARDENNE, Z. KLIMONT, R. J. ANDRES, A. VOLKE, AND S. D. ARIAS (2011), Anthropogenic sulfur dioxide emissions: 1850-2005, *Atmospheric Chemistry and Physics*, *11*(3), 1101-1116, doi:10.5194/acp-11-1101-2011.

SKEIE, R. B., T. K. BERNTSEN, G. MYHRE, K. TANAKA, M. M. KVALEVAG, AND C. R. HOYLE (2011), Anthropogenic radiative forcing time series from pre-industrial times until 2010, *Atmospheric Chemistry and Physics*, *11*(22), 11827-11857, doi:10.5194/acp-11-11827-2011.

SMITH, T. M., R. W. REYNOLDS, T. C. PETERSON, AND J. LAWRIMORE (2008), Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006), *Journal of Climate*, *21*(10), 2283-2296, doi:10.1175/2007jcli2100.1.

ŠTĚPÁNEK, P., P. ZAHRADNÍČEK, AND R. HUTH (2011), Interpolation techniques used for data quality control and calculation of technical series: an example of a Central European daily time series, *Idojaras*, *115*(1-2), 87-98.

STOCKER, T. F., D. QUIN, G.-K. PLATTNER, M. M. B. TIGNOR, S. K. ALLEN, J. BOSCHUNG, A. NAUELS, Y. XIA, V. BEX, AND P. M. MIDGLEY (EDS.) (2013), *IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, 1535 pp., ISBN 978-1-107-05799-1.

THEMEßL, M. J., A. GOBIET, AND G. HEINRICH (2012), Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, *Climatic Change 112*(2), 449-468, doi: 10.1007/s10584-011-0224-4.

TRENBERTH, K. E., J. M. CARON, D. P. STEPANIAK, AND S. WORLEY (2002), Evolution of El Nino-Southern Oscillation and global atmospheric surface temperatures, *Journal of Geophysical Research-Atmospheres*, *107*(D7-8), 4065, doi:10.1029/2000jd000298.

UPPALA, S. M., ET AL. (2005), The ERA-40 re-analysis, *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 2961-3012, doi:10.1256/qj.04.176.

VAN DER A, R. J., M. A. F. ALLAART, AND H. J. ESKES (2010), Multi sensor reanalysis of total ozone, *Atmospheric Chemistry and Physics*, *10*(22), 11277-11294, doi:10.5194/acp-10-11277-2010.

VAN DER WERF, G. R., AND A. J. DOLMAN (2014), Impact of the Atlantic Multidecadal Oscillation (AMO) on deriving anthropogenic warming rates from the instrumental temperature record, *Earth System Dynamics*, *5*(2), 375-382, doi:10.5194/esd-5-375-2014.

WANG, Y. M., J. L. LEAN, AND N. R. SHEELEY (2005), Modeling the Sun's magnetic field and irradiance since 1713, *Astrophysical Journal*, *625*(1), 522-538, doi:10.1086/429689.

WU, S., Z. Y. LIU, R. ZHANG, AND T. L. DELWORTH (2011), On the observed relationship between the Pacific Decadal Oscillation and the Atlantic Multi-decadal Oscillation, *Journal of Oceanography*, *67*(1), 27-35, doi:10.1007/s10872-011-0003-x.

WYATT, M. G., S. KRAVTSOV, AND A. A. TSONIS (2012), Atlantic Multidecadal Oscillation and Northern Hemisphere's climate variability, *Climate Dynamics*, *38*(5-6), 929-949, doi:10. 1007/s00382-011-1071-8.

ZHANG, R., ET AL. (2013), Have Aerosols Caused the Observed Atlantic Multidecadal Variability?, *Journal of the Atmospheric Sciences*, *70*(4), 1135-1144, doi:10.1175/ jas-d-12-0331.1.

ZHANG, Y., J. M. WALLACE, AND D. S. BATTISTI (1997), ENSO-like interdecadal variability: 1900-93, *Journal of Climate*, *10*(5), 1004-1020, doi:10.1175/1520-0442(1997)010 <1004:eliv> 2.0.co;2.

ZHOU, J. S., AND K. K. TUNG (2013), Deducing Multidecadal Anthropogenic Global Warming Trends Using Multiple Regression Analysis, *Journal of the Atmospheric Sciences*, *70*(1), 3-8, doi:10.1175/jas-d-12-0208.1.

ZORITA, E., AND H. VON STORCH (1999), The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, *Journal of Climate*, *12*(8), 2474-2489, doi:10.1175/1520-0442(1999)012<2474:tamaas>2.0 .co;2.

APPENDIX I

MIKŠOVSKÝ, J., AND A. RAIDL (2006), Testing for nonlinearity in European climatic time series by the method of surrogate data, *Theoretical and Applied Climatology*, *83*(1-4), 21-33, doi:10.1007/s00704-005-0130-7.

Theor. Appl. Climatol. 83, 21–33 (2006)
DOI 10.1007/s00704-005-0130-7

Theoretical and Applied Climatology
Printed in Austria

Department of Meteorology and Environment Protection, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Testing for nonlinearity in European climatic time series by the method of surrogate data

J. Mikšovský and **A. Raidl**

With 4 Figures

Received May 12, 2004; revised January 12, 2005; accepted January 16, 2005
Published online May 30, 2005

Summary

Using temperature and pressure records from Czech meteorological stations and NCEP/NCAR reanalysis series, we tested for the presence of detectable nonlinearity in univariate and multivariate climatic time series. The method of surrogate data was utilized for nonlinearity detection – results of nonlinear prediction for the original series were compared to the results for series whose nonlinear structure was randomized. The prediction was done by means of local linear models in the reconstructed phase space. None or very weak nonlinearity was found in the single (univariate) series, and pressure series generally exhibited stronger nonlinearity than series of temperature (daily mean, minimum or maximum). Distinct nonlinearity was found in all tested multivariate systems, especially when both temperatures and pressures were used simultaneously to form the phase space. Nonlinearity tests were carried out for 30-year and 10-year-long datasets and nonlinear behavior was generally more apparent in the longer versions. In addition, the tested systems showed more substantial nonlinearity when the success of short-range prediction was used as the discriminating statistic; with an increase of the prediction time, detectable nonlinearity became weaker and it disappeared completely for long-term prediction.

1. Introduction

Time series analysis is a frequent way of dealing with meteorological signals, both measured and produced by numerical simulations. Although the climate system is essentially nonlinear, linear methods of signal analysis are very common and the linear analysis techniques (such as spectral methods, linear regression or ARMA models) often predominate over the nonlinear ones (artificial neural networks, local or global polynomial fits and others). As the linear approach is usually less complicated, easier to implement, and computationally less demanding, the question arises whether and where nonlinear methods have their place in the study of meteorological signals. The recent development of computers has answered the problem of the high computational demands of the usual nonlinear methods which can now be satisfied by a common personal computer. Therefore, the issue is: Do nonlinear methods offer enough benefits in exchange for their higher demands? Is the inherent nonlinearity strong enough in the climatic series? The intention of this paper is to show the results of nonlinearity tests of a few typical climatic time series and multivariate systems of series and to decide whether nonlinearity is detectable, how extensively it is manifested and, therefore, if the application of nonlinear methods is appropriate.

While the commonly used nonlinear methods, e.g. artificial neural networks or methods based on local models, are generally suitable for dealing with signals of both linear and nonlinear origin, they are usually more complicated than their linear counterparts. This is reflected in the high number of parameters that need to be determined before the method can be used (such as weights in the neural network or coefficients of the local model). Also, due to the higher complexity of nonlinear methods, their application to a strictly linear series (i.e. series without any nonlinear component whatsoever) may even produce worse results than the application of a purely linear method. It is therefore desirable to distinguish between linear and nonlinear tasks, and to match the correct techniques to the nature of the problem to be solved. Identification of a time series as linear or nonlinear is very useful in many scientific disciplines, and so numerous techniques have been proposed for nonlinearity testing (see, e.g. Galka, 2000 or Schreiber and Schmitz, 2000). In this paper, an approach employing so-called surrogate time series (surrogates) is utilized (Theiler et al., 1992). Its principle is quite straightforward. First, the original time series is altered so that its linear structure is preserved while the nonlinear dependencies are randomized. Then results of some nonlinearity-sensitive technique, for instance nonlinear prediction, are compared for the original series and several surrogate series. If the result for the original series differs significantly from those for the surrogates (in the statistical sense), the presence of nonlinearity is positively established. Details on the generation of surrogates and their use are provided in Section 2.3.

The problem with nonlinearity testing of climatic time series has already been addressed by Paluš and Novotná (1994), who compared the redundancy of the original and surrogate time series for series of mean daily temperature and mean daily pressure from Prague-Klementinum, detecting nonlinearity in the latter and finding the former to be linear. Paluš also detected nonlinearity in the bivariate system consisting of temperature and pressure series from Prague-Klementinum (Paluš, 1996). The nonlinear nature of pressure series was confirmed by Casdagli (1997) as well, this time for the Australian records of atmospheric pressure. An example of a meteorology-related analysis can also be found in the paper by Schreiber and Schmitz (2000), where the authors tested for nonlinearity in the series of monthly values of the South Oscillation index, comparing results of nonlinear prediction for the original series and 99 surrogates. Nonlinearity was not detected in this case. Yet another nonlinearity testing procedure was proposed by Tsonis (2001) and applied to a series of average hemispheric available potential energy, which was found to have a significant nonlinear component. It is therefore clear that climatic time series, although originating from an inherently nonlinear system, may exhibit nonlinearity in some cases and appear linear in others. In this paper, we have tried to determine if climate nonlinearity is detectable from the European series of daily temperature and pressure and how extensively it is expressed.

2. Methods

2.1 Phase space reconstruction

Phase space reconstruction is applied as the basis for nonlinearity testing here. Knowledge of the phase space (PS) structure is of great benefit when studying any physical system, since an accurate and unambiguous characterization of its state can be done. For some simple systems, the structure of the phase space is known or can be found without difficulty. In the physics of the climate, the situation is more difficult. First, the essential components of the climate system, atmosphere and ocean, are continuous matter systems, thus having infinite number of degrees of freedom. Several attempts to find low-dimensional dynamics of the climate system were made by a number of authors in the past (see, e.g. Sivakumar (2004) for overview), but no clear and definite proof of the global climate attractor presence has been found. This suggests that such an attractor may not be present, at least not in the well-expressed form typical for some low-dimensional systems. Even if it does exist, its detection may be impossible from the measured time series due to, e.g. insufficient length of the available records. In spite of the fact that the exact description is unattainable, major features of the climate system may still be described by a finite number of variables with sufficient accuracy. Numerical

models, for example, are based on using a few characteristic variables in a finite number of grid points, and they are able to predict weather or model climate with accuracy unrivalled by any other available method. However, even though numerical models' dimensions are finite, their value is still quite high as such models generally have thousands or millions of degrees of freedom. Moreover, numerical models, being very complicated and quite demanding with respect to the formulation of the initial and boundary conditions as well as needed computational power, are too cumbersome and impractical for some tasks. This is where time series analysis methods often come into play. Since time series of many different quantities are plentiful in meteorology, it is reasonable to try to use them to construct some kind of finite-dimensional phase space equivalent. The usability of such a reconstruction may be limited and restricted to certain areas or time periods only. It is, however, possible to capture (some) major features of the climate dynamics with a low number of variables in this way.

In the phase space reconstruction (PSR) process, vector space is constructed from the available time series to be an equivalent of the original phase space, or at least to represent its practically usable substitute. Here,

$$\mathbf{y}_m(t) = (X_1(t), X_2(t), \ldots, X_m(t)) \tag{1}$$

is an m-dimensional vector, describing state of the system in time t, and $X_i(t)$ is its i-th component. The form of $X_i(t)$ depends on the reconstruction technique employed. When just a single time series is used, the most frequently used method of reconstruction is the time delay technique (Packard et al., 1980; Takens, 1981). This uses the time-lagged values of the scalar predictor time series $x_1(t)$ as elements of $\mathbf{y}_m(t)$ as follows:

$$X_i(t) = x_1(t - (i-1)\tau), \quad i = 1, \ldots, m. \tag{2}$$

τ is the time delay. In order to use the time delay method, two scalar values must be determined, characterizing embedding dimension m and time distance τ between the elements of the vector. Although there are many methods of their estimation (see, e.g. Abarbanel, 1996 or Kantz and Schreiber, 1997), the choice can also be done by identifying the combination of m and τ for which nonlinear prediction produces the best results.

Sometimes more than one series is available for the reconstruction. Meteorological measurements are typically done at numerous locations, and several variables are recorded simultaneously, hence multiple series $x_i(t)$ may be utilized (Keppenne and Nicolis, 1989). This allows for PS reconstruction where elements of the vector $\mathbf{y}_m(t)$ are represented by values of m different series at the same time t:

$$X_i(t) = x_i(t), \quad i = 1, \ldots, m. \tag{3}$$

We refer to this reconstruction technique as multivariate PSR. Of course, when the number of the available series S is higher than the embedding dimension, $S > m$, a set of m series has to be picked from those available. This can be done either by means of some dimensionality reduction technique (such as PCA), or by choosing the best fitting subset of the series (e.g. the one which can provide the best prediction, as done in Section 4.1).

Time delay and multivariate reconstruction techniques can also be combined when the embedding dimension is greater than the number of the available series, $1 < S < m$. The general formula, describing time delay, multivariate and combined reconstruction methods, is:

$$X_i(t) = x_{i-((i-1)\operatorname{div} S)S}\,(t - ((i-1)\operatorname{div} S)\tau), \quad i = 1, \ldots, m, \tag{4}$$

where X div Y stands for integer division of X by Y.

2.2 Prediction by local linear models

As soon as the PS is approximated by some of the above described methods, prediction may be applied. The technique of local models (Farmer and Sidorowich, 1987; Casdagli, 1989; Sugihara and May, 1990; Ott et al., 1994) is based on finding the states in the history of the system which correspond to situations similar to that being predicted. The similarity of the two vectors in the PS $\mathbf{y}_m(t)$ and $\mathbf{y}_m(t')$ can easily be quantified by measuring their distance, e.g. using the Euclidean L_2 norm which was applied in the following form here:

$$\|\mathbf{y}_m(t), \mathbf{y}_m(t')\| = \left(\frac{1}{m}\sum_{i=1}^{m} |X_i(t) - X_i(t')|^2\right)^{1/2}. \tag{5}$$

For each predicted state $\mathbf{y}_m(t)$, certain number N of the close (in the sense of minimal distance) states is found in the phase space. These are called the nearest neighbors of $\mathbf{y}_m(t)$ and denoted $\mathbf{y}_m(t,i), i = 1,\ldots,N$; the corresponding times are denoted $t'(t,i), i = 1,\ldots,N$. From the time evolution of the nearest neighbors, a mapping is constructed to approximate the system dynamics in some local neighborhood of $\mathbf{y}_m(t)$. Let the series $x(t)$ be the predictand series. For some time t and prediction done T time steps ahead, predicted value in time $t+T$ is expressed as a linear function of $\mathbf{y}_m(t)$:

$$x^{\mathrm{PRED}}(t+T) \;=\; a_0 + \sum_{i=1}^{m} a_i(t) X_i(t). \tag{6}$$

$a_i(t), i = 0,\ldots,m$ represent coefficients of the local linear models, with $a_0(t)$ being the absolute coefficient. Since an individual mapping is created for each predicted state, coefficients $a_i(t)$ in (6) are not time independent and they must be computed separately for each time t. Their values can be found by solving

$$\begin{pmatrix} x\,(t'(t,1)+T) \\ x\,(t'(t,2)+T) \\ \cdots \\ x\,(t'(t,N)+T) \end{pmatrix} = \begin{pmatrix} 1 & X_1(t,\,1) & \ldots & X_m(t,\,1) \\ 1 & X_1(t,\,2) & \ldots & X_m(t,\,2) \\ \vdots & \vdots & & \vdots \\ 1 & X_1(t,N) & \ldots & X_m(t,N) \end{pmatrix} \cdot \begin{pmatrix} a_0(t) \\ a_1(t) \\ \cdots \\ a_m(t) \end{pmatrix} \tag{7}$$

for $a_i(t)$ ($X_i(t,\,j)$ denotes the i-th component of $\mathbf{y}_m(t,\,j)$). Since mostly $N \gg m$, the problem is overspecified and solvable in the least-squares sense. The quality of the local fits may be reduced if the nearest neighbors, close to $\mathbf{y}_m(t)$ in time (time correlated states), are used. This problem can be avoided by excluding all $\mathbf{y}_m(t^*), |t - t^*| \le t_{\mathrm{DECOR}}$, from the search for the nearest neighbors of $\mathbf{y}_m(t)$ (t_{DECOR} is called decorrelation time).

The success of the prediction can be enumerated by a single scalar value. Here, we employ the common root mean square error (RMSE), defined as:

$$\mathrm{RMSE}\big(x, x^{\mathrm{PRED}}\big) = \left(\frac{1}{K} \sum_{t} \big(x(t) - x^{\mathrm{PRED}}(t)\big)^2 \right)^{1/2}, \tag{8}$$

where $x^{\mathrm{PRED}}(t)$ is the series of the predicted values obtained from (6) and summation is performed over all t for which both original and predicted values exist – number of such cases is denoted K.

2.3 Philosophy of the method of surrogate data

The applied technique of nonlinearity testing is basically the one used by Schreiber and Schmitz (1996, 2000). The null hypothesis to be accepted or rejected by the nonlinearity test was: The series was generated by a Gaussian linear stochastic process of an arbitrary order with constant coefficients, possibly multivariate, and transformed by some nonlinear, but static and invertible measurement function. To test this hypothesis, RMSE of the prediction is computed for the original series (or multivariate system of series) and compared to RMSEs obtained for surrogates. Because the distribution of the surrogates' RMSEs is generally non-Gaussian, the one-sided rank-order test (Theiler et al., 1992) was employed instead of some kind of test using the standard deviation of surrogates' RMSEs (such as t-test or z-test). For testing at the significance level of α, $1/\alpha - 1$ surrogates are generated, i.e. $1/\alpha$ series are used including the original one. If prediction for the original series gives lower RMSE than for all the surrogates, the hypothesis of linear generating process is rejected, with probability α of false rejection. We used a set of 49 surrogates here, thus the probability of false rejection was 2%. Using more surrogate series would result in higher confidence at the cost of slower computation – testing at the 98% confidence level was a compromise between the speed and the reliability of the tests.

Surrogate time series (a.k.a. surrogates) represent a modification of the original data created

so that their linear structure is preserved while nonlinear dependencies are randomized (or, more generally, data created to represent some null hypothesis, such as the one above). As discussed by Schreiber (1998), Schreiber and Schmitz (2000) and Galka (2000), the generation of surrogates can be done in several ways, from a simple phase-randomization of the Fourier spectrum to the application of the simulated annealing method. Although none of the methods can generate perfect surrogates in all cases (i.e. surrogates with exactly the same distribution of values and autocorrelation function as the original data), some of them can provide a sufficient degree of accuracy. IAAFT (iterated amplitude adjusted Fourier transform – Schreiber and Schmitz, 1996) was utilized here, as a method of producing satisfactory surrogates while the computational demands remained moderate. The method is based on alternating steps of power spectrum change and distribution adjustments, performed until a randomized series is obtained with both the distribution of values and power spectrum as close to the original series as possible. In the case of multivariate systems, not just autocorrelations and the distribution of values have to be preserved, but also cross-correlations. The IAAFT procedure can be modified for this task as well – again, see Schreiber and Schmitz (2000). Since the IAAFT surrogates generation procedure is a part of the skilful TISEAN program package (Hegger et al., 1999, downloadable from the web page of the Max Planck Institute for the Physics of Complex Systems in Dresden), we utilized it for surrogate generation here.

One of the potentially treacherous features of the Fourier transform based surrogate generation techniques, IAAFT included, is the problem arising from the possible leap between the first and the last value of the series. The Fourier transform assumes periodicity of the processed series, thus unwanted artifacts may appear in the surrogate series when the leap is significant. This effect can be countered by choosing such subinterval of the studied dataset so that the first-last difference is minimized (Schreiber and Schmitz, 2000). Since the series used for testing here (10935 or 3645-day-long) did not cover the entire 30 or 10 years, we could choose a suitable subset to minimize the end-to-end difference.

As pointed out by Kugiumtzis (1999), the surrogate generation process may unintentionally alter the linear structure of the series, and it is therefore desirable to test whether or not the linear dependencies in the data were noticeably changed. In order to find out how the linear structure of the series was modified in the surrogates and to avoid false rejection due to the change in the linear dependencies, we have also computed results of the strictly linear prediction for original and surrogate series, employing multiple linear regression (MLR). Should the results of the MLR indicate significant change in the linear structure of the surrogates (when RMSE for the original series is out of the min–max range of RMSEs for surrogates), the nonlinearity test could not be considered conclusive, regardless of its outcome. Here, all presented examples passed this test successfully, i.e. the linear structure of the series remained almost unchanged in the surrogates.

3. Data

Time series of daily values from two datasets were used. The first dataset consisted of time series of various variables measured at the Czech meteorological stations: mean daily air temperature (MDT), maximum daily air temperature (MaxDT), minimum daily air temperature (MinDT) and mean daily pressure (MDP), all measured 2 meters above ground level. We have tested data from several stations, but since differences between the results for different stations were rather small, only the results for the series from Prague (station Prague-Ruzyně, 50°6′3″N, 14°15′28″E, altitude 364 m) are presented here. To ascertain the influence of the series length on the results of the nonlinearity tests, two different lengths of the series were used. The 10935-day-long series covered years 1971 to 2000 (1969 to 1998 in the case of the pressure series – only measurements prior to 1999 were available), the 3645-day-long series were for the years 1991–2000 (1989–1998 for pressure series). The numbers of days, 10935 and 3645, do not represent exactly 30 or 10 years. We used these somewhat shortened sequences because of the surrogate series generation procedure, where series whose length satisfies the condition $N = 2^{\alpha}3^{\beta}5^{\gamma}$ were required, α, β, γ being arbitrary integer numbers.

This restriction permits the use of the FFT algorithm for the Fourier transform.

The second dataset, the NCEP/NCAR reanalysis (Kistler et al., 2001), does not represent data directly measured at meteorological stations. It was created by processing a large number of existing measurements of various kinds using a time-invariant data assimilation system for the entire processed period (from 1948 on). The reanalysis data are available in a form of values on a regular grid ($2.5^\circ \times 2.5^\circ$). Here, we used time series of mean daily temperature at 1000 hPa (T_{1000}) and series of daily mean sea level pressure (MSLP) from several grid points (see below). Again, two lengths of the series were used – 10935 and 3645 days, representing years 1971 to 2000 and 1991 to 2000, respectively.

4. Results

Nine typical examples of nonlinearity test outcomes are shown here, six of them for time delay PSR from single series and three for multivariate PSR. Each case was tested for both the 30-year and 10-year-long versions of the series. The parameters of the computation were determined individually for each case (see Section 4.1) and they are shown in Table 1, together with nonlinearity test results. The individual cases are identified by roman numerals and lowercase letters indicate the length of the series (a = 10935 days, b = 3645 days). Decorrelation time was 30 days and the prediction one day ahead was used, unless otherwise specified.

4.1 Parameter estimation

Values of three parameters are required before local linear models in the time delay reconstructed phase space can be used as a forecast tool: time delay τ, embedding dimension m and the number of the nearest neighbors utilised N. We used a combination of the parameters which produced the best value of RMSE for the original ('nonsurrogated') series. The simplest situation was for τ, as the best results were generally obtained for $\tau = 1$ day, and this value is therefore used in all phase space reconstructions here. The dependence of RMSE on m and N was more interesting. RMSE usually drops as m increases due to the improved ability of the local models to describe the local dynamics, but it starts to rise again for high values of m. The same behavior was generally observed for N, with a minimum located between a region where there are too few of the nearest neighbors to average out noise and a region where their too high number is resulting in severe averaging and, as a consequence, there is a loss of the ability to distinguish between different states. In the case of time delay reconstruction, the minimum of RMSE is typically very flat and the choice of the proper parameter values is therefore not critical – and, as demonstrated below, the test results are quite robust to perturbations of m or N. An illustration of RMSE dependence on values of m and N can be seen in Fig. 1a for prediction of the 3645-day-long series of daily mean temperature from Prague one day ahead.

In the case of multivariate PSR, the value of the time delay was not needed. Instead, the predictor series $x_i(t)$ $i = 1, \ldots, m$ had to be chosen from the available series. We used series of T_{1000} and MSLP from the grid point 50° N, 15° E as predictands, and 2×35 series of T_{1000} and MSLP as potential predictors, i.e. the series from which the predictors were selected. The potential predictors covered the area between 60° N, 0° E and 40° N, 30° E in the $5^\circ \times 5^\circ$ grid. A relatively straightforward approach to the selection of predictors was used. First, prediction was performed using a one-dimensional PSR, and all potential predictors were tested in the role of $x_1(t)$. The one providing the best prediction was fixed as the first element of $\mathbf{y}_m(t)$. Then the two-dimensional reconstruction was built ($m = 2$), with all remaining (not yet used) series tested in the place of the second element, and, again, the most successful candidate was retained. Series after series were added to the reconstruction in this fashion, until the increase of the embedding dimension ceased to lower the resulting RMSE any further. An example of RMSE dependence on values of m and N in case of multivariate phase space reconstruction is given in Fig. 1b. Unlike in the case of time delay PSR (Fig. 1a), the minimum is deeper and better expressed.

The values of m and N chosen for the individual cases and used in the testing are given in Table 1.

Table 1. Overview of the tested cases, used parameters and nonlinearity tests results. From the left: Identifier of the case (same as in Fig. 2), method of phase space reconstruction, predictand series, predictor series, length of the used series, embedding dimension, number of the nearest neighbors, result of the linearity hypothesis test, arithmetic average of RMSEs for surrogates expressed in percent of RMSE for the original series

Case	PSR method	Predictand	Predictor(s)	Length [day]	m	N	Linear	Δ[%]
I a	Time delay	Prague – MDT	Prague – MDT	*10935*	*9*	*1500*	No	*100.4*
I b	Time delay	Prague – MDT	Prague – MDT	*3645*	*9*	*800*	Not rejected	*100.1*
II a	Time delay	Prague – MinDT	Prague – MinDT	*10935*	*10*	*1500*	No	*100.9*
II b	Time delay	Prague – MinDT	Prague – MinDT	*3645*	*10*	*1200*	Not rejected	*100.5*
III a	Time delay	Prague – MaxDT	Prague – MaxDT	*10935*	*8*	*1200*	No	*100.3*
III b	Time delay	Prague – MaxDT	Prague – MaxDT	*3645*	*10*	*1000*	Not rejected	*100.0*
IV a	Time delay	Prague – MDP	Prague – MDP	*10935*	*3*	*1200*	No	*100.9*
IV b	Time delay	Prague – MDP	Prague – MDP	*3645*	*3*	*800*	Not rejected	*100.4*
V a	Time delay	NCEP/NCAR – T_{1000} (50° N, 15° E)	NCEP/NCAR – T_{1000} (50° N, 15° E)	*10935*	*9*	*1000*	No	*100.4*
V b	Time delay	NCEP/NCAR – T_{1000} (50° N, 15° E)	NCEP/NCAR – T_{1000} (50° N, 15° E)	*3645*	*9*	*1200*	Not rejected	*99.9*
VI a	Time delay	NCEP/NCAR – MSLP (50° N, 15° E)	NCEP/NCAR – MSLP (50° N, 15° E)	*10935*	*4*	*1000*	No	*100.8*
VI b	Time delay	NCEP/NCAR – MSLP (50° N, 15° E)	NCEP/NCAR – MSLP (50° N, 15° E)	*3645*	*4*	*1200*	No	*100.9*
VII a	Multivariate	NCEP/NCAR – T_{1000} (50° N, 15° E)	NCEP/NCAR – T_{1000}s	*10935*	*19*	*250*	No	*106.8*
VII b	Multivariate	NCEP/NCAR – T_{1000} (50° N, 15° E)	NCEP/NCAR – T_{1000}s	*3645*	*13*	*200*	No	*104.8*
VIII a	Multivariate	NCEP/NCAR – MSLP (50° N, 15° E)	NCEP/NCAR – MSLPs	*10935*	*16*	*250*	No	*107.8*
VIII b	Multivariate	NCEP/NCAR – MSLP (50° N, 15° E)	NCEP/NCAR – MSLPs	*3645*	*16*	*200*	No	*108.1*
IX a	Multivariate	NCEP/NCAR – T_{1000} (50° N, 15° E)	NCEP/NCAR – T_{1000}s and MSLPs	*10935*	*19*	*250*	No	*132.1*
IX b	Multivariate	NCEP/NCAR – T_{1000} (50° N, 15° E)	NCEP/NCAR – T_{1000}s and MSLPs	*3645*	*19*	*200*	No	*125.4*

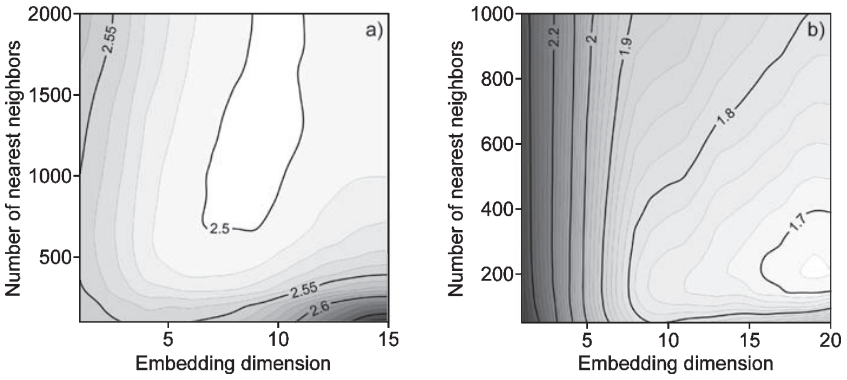


Fig. 1. RMSE [°C] of prediction 1 day ahead as a function of m and N for (**a**) 3645-day-long series of mean daily temperature from Prague (time delay PSR) and (**b**) 3645-day-long NCEP/NCAR T_{1000} series (multivariate PSR from series of T_{1000} and MSLP)

4.2 Nonlinearity tests

To accept or reject the hypothesis of a linear character of the series at the 2% level of significance using a one-sided test, RMSEs of the prediction for the original series and 49 surrogates were computed for each case. The results are presented in Fig. 2 as diagrams which show the value of RMSE for the original series (long bold horizontal line) and values of RMSEs for all 49 surrogates (dots). The arithmetic average of the surrogates' RMSEs is also displayed (short horizontal line), together with their 2σ range. There are two such diagrams for each tested series or multivariate system, one for the 10935-day-long series (denoted a), the other for the 3645-day-long one (denoted b). The identifiers in the upper left corner of the diagrams correspond to the IDs in the first column of Table 1.

The first four cases show the results for the univariate series from Prague, using the time delay method of PS reconstruction. Mean daily temperature (I), minimum daily temperature (II), maximum daily temperature (III) and mean daily pressure (IV) series were examined. All the series exhibit nonlinearity for the longer version of the series only while the shorter series satisfies the hypothesis of a stationary linear generating process. It should also be noted that differences between results from the original series and surrogates are quite small (see the rightmost column of Table 1), even when the linearity hypothesis is rejected by the test. This indicates that even when nonlinearity is detected, it is very weak. Tests of the T_{1000} and MSLP series from NCEP/NCAR grid point 50° N, 15° E (cases V and VI) gave very similar results, with weak nonlinearity detected in the longer version of T_{1000} series and both versions of MSLP series. It appears that the nonlinearity of the climate system is better exhibited in the pressure series, in contrast to the 10-year-long temperature series which seemed completely linear. Such outcomes agree with findings of Paluš and Novotná (1994). Nonlinearity is also more apparent in the longer series, whereas the shorter versions do not usually reveal the nonlinear nature of the atmospheric dynamics. This may partially be a result of the higher discriminating power of the tests for longer series. However, it should be noted that the absolute difference between RMSE for the original series and average RMSE for surrogates was higher for the longer series than for the shorter series in most cases.

While the tests on the univariate series indicated very weak nonlinearity at most, the situation was entirely different for the multivariate systems. We show results for three cases here – prediction of T_{1000} (50° N, 15° E) from multiple T_{1000} series (case VII), prediction of MSLP (50° N, 15° E) from multiple MSLP series (case VIII) and prediction of T_{1000} (50° N, 15° E) from both T_{1000} and MSLP series (case IX). For all three tasks, and for both lengths of the tested series, substantial nonlinearity was detected. The most significant contrast between results from the original series and results from surrogates was observed in case IX.

Since the testing was performed for one specific combination of parameters m and N only, the questions how much do the results depend on the values of the used parameters and if their change could not affect the outcome might be asked. We have investigated this problem and found out that the test results are quite robust with respect to the values of the used parameters.

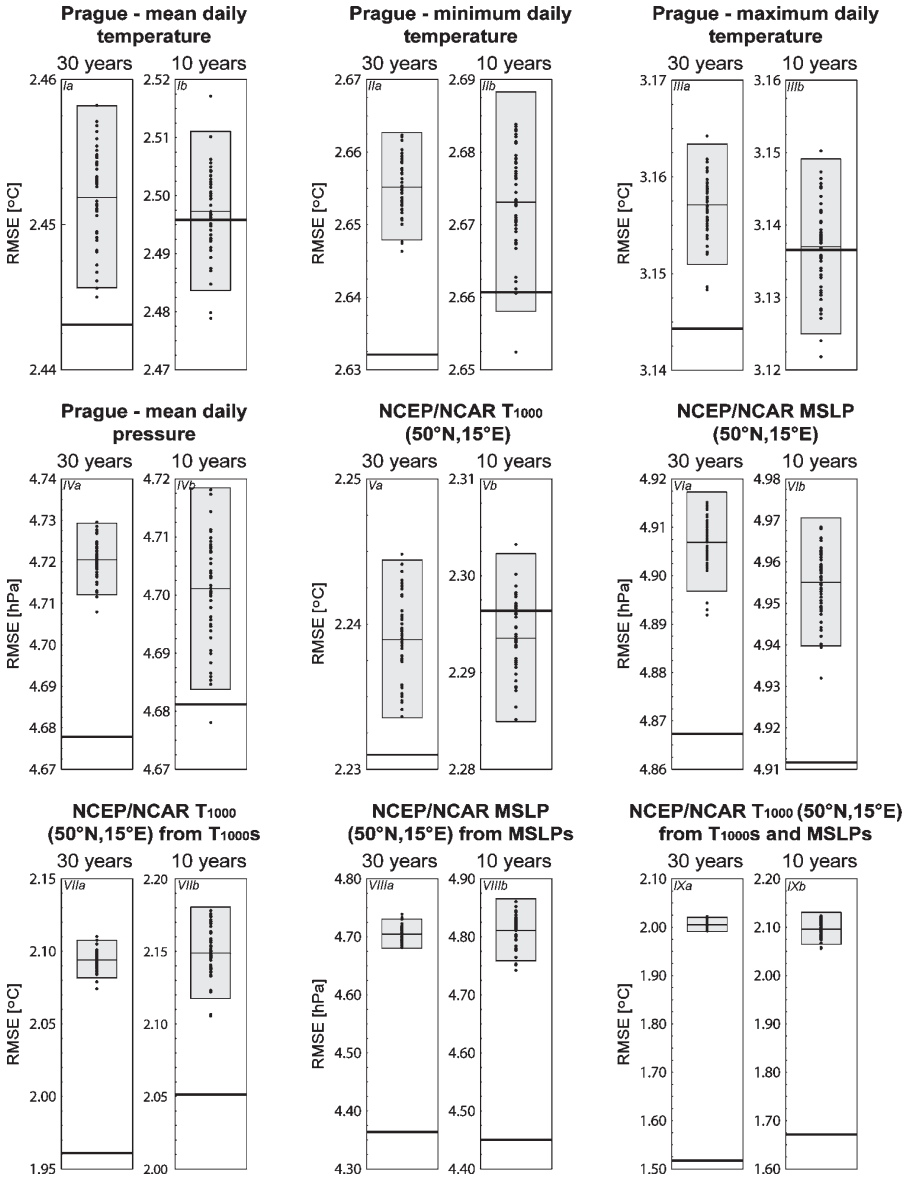


Fig. 2. Results of the nonlinearity tests for nine different cases, each of them in 30-year-long and 10-year-long version of the series. The long horizontal line represents RMSE of the prediction for the original series, dots are RMSEs for the surrogates. The grey rectangle with short horizontal line shows arithmetic average of the surrogates' RMSEs and their 2σ range. Identifiers in the upper left corner of the diagrams correspond to the ones in Table 1

Two examples can be seen in Fig. 3, where the results of the tests from cases I b (Fig. 3a) and IX b (Fig. 3b) are presented for a wider range of embedding dimensions. The linearity hypothesis was accepted in the case I b when $m = 9$ was used, and the same result would have been obtained for any m in the interval 1 to 12. In case IX b, the null hypothesis of a linear generating process was clearly rejected, as it would have been for any m greater than 2. Similar robustness of the results was also verified for values of N.

Since RMSE of time prediction one day ahead was used as the discriminating statistic in all the previous cases, we wondered how the results would change if the prediction time was increased. As the atmospheric system shows typical chaotic behavior at synoptic scales, demonstrated by the divergence of close trajectories

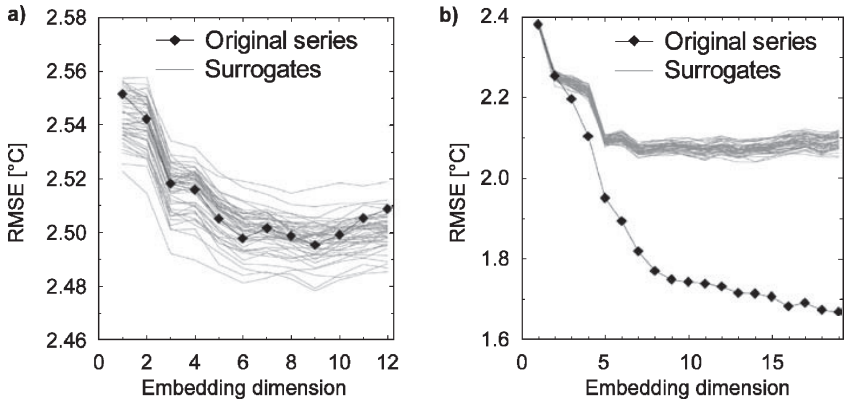


Fig. 3. RMSE of prediction 1 day ahead for the original series and 49 surrogates as a function of embedding dimension for (**a**) 3645-day-long series of mean daily temperature from Prague using time delay PS reconstruction and (**b**) 3645-day-long series of T_{1000} (50° N, 15° E) using multivariate PS reconstruction from T_{1000} and MSLP series

in the PS, successful deterministic prediction is possible for just a few days. Therefore the prediction, linear or not, is degraded to essentially statistical forecast for T beyond some threshold and the eventual advantage of the nonlinear techniques over the linear ones is lost. To test the said effect, we performed the nonlinearity test from case IX b for prediction times T from 1 to 10 days, instead of just one day. Indeed, as seen in Fig. 4, the detectable nonlinearity disappeared for $T \geq 5$ days. Generally, a suitable statistic is required if nonlinearity is to be detected – and the same series or multivariate system of series may exhibit different degrees of nonlinearity for different tasks.

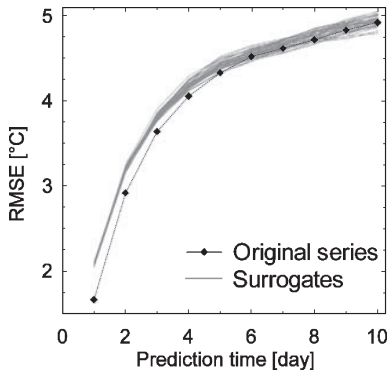


Fig. 4. RMSE of prediction for the original series and 49 surrogates as a function of the prediction time (3645-day-long series of T_{1000} (50° N, 15° E) using multivariate PS reconstruction from T_{1000} and MSLP series – case IX b in Table 1)

5. Discussion

For all the tested univariate series, differences between results from the original series and results from surrogates were very small. Nonetheless, the hypothesis of a linear stochastic generating process was rejected for the 30-year-long versions of the series, in contrast to all but one 10-year-long series. This result could mean that nonlinearity is present, and it can be more easily detected from the longer series. It is necessary to be careful about such an interpretation since the tested hypothesis assumed a time-independent generating process. Because the climate system cannot be considered completely stationary, the question is what was the actual cause of the rejection – was it nonstationarity or nonlinearity? To investigate the possible effect of nonstationarity, we have constructed and tested a synthetic, completely linear series, mimicking the 10935-day-long series of mean daily temperature from Prague (which was tested in case I a, and found to be nonlinear). The synthetic series was created as a sum of an output of an AR(4) process with constant coefficients, periodic signal simulating the annual cycle and linear trend (which is a rough approximation of nonstationarity in the Prague temperature series, with mean daily temperature rising by about 0.3 °C/10 years). Coefficients of the AR model, amplitude of the periodic component and value of the trend were chosen in accordance with the properties of the original measured series. The hypothesis of the stationary linear generating process was not

rejected for the synthetic linear series, therefore nonstationarity in the form of a linear trend is probably not responsible for linearity hypothesis rejection in case I a (or other tested 30-year-long univariate series). Nonetheless, from the practical point of view, the problem of (non)linearity of the tested univariate series is rather academic, due to very small differences between results from the original series and results from surrogates (see the rightmost column of Table 1). Such small differences mean that even if the series are technically nonlinear according to the test, the potential gain from the application of nonlinear methods of time series analysis would be negligible.

The tests performed indicate that nonlinearity is expressed very well in the multivariate systems, although it was slightly less distinct in the 10-year-long versions than in the 30-year-long series. Also, the difference between RMSE for the original series and for the surrogates was quite large, both in absolute numbers and compared to standard deviation of the surrogates' RMSEs. The difference was most significant for the multivariate system consisting of both temperatures and pressures (case IX). In our experience, the more information about the analyzed system introduced into the phase space reconstruction, the more distinct nonlinearity can be expected to be observed. A single time series does usually not provide enough information to enable the nonlinear character of the climate system to be distinctly revealed, while use of multiple series can offer a more complete description of the system, including its nonlinear features.

As an illustration of how various nonlinear methods do actually perform, there is an example of solving the problems V b and IX b (prediction of the 3645-day-long T_{1000} series from the grid point 50° N, 15° E using either time delay or multivariate PSR method) by several forecast techniques in Table 2. In addition to prediction by local linear models, three-layer perceptron neural network (with 8 neurons in the hidden layer, trained by the quasi-Newton method) and radial basis function neural network (150 units in the hidden layer) were employed (for more detailed information on neural networks see, e.g. Haykin, 1999). The results are summarized in Table 2 and compared to outcomes of multiple linear regression (MLR) and persistent prediction (i.e. prediction made simply by taking $x^{\mathrm{PRED}}(t+T) = x(t)$). The results agree with the findings of the nonlinearity tests quite well. While there is no clear difference between the results from the nonlinear methods and MLR when time delay PS reconstruction from a single series is used, the advantage of nonlinear methods is distinct for multivariate PSR.

Table 2. RMSE of prediction of the NCEP/NCAR reanalysis T_{1000} series (grid point 50° N, 15° E) 1 day ahead by various methods, utilizing either time delay (the second column) or multivariate (the third column, series of T_{1000}s and MSLPs used as predictors) phase space reconstruction. 7 years were used as the training set and 3 years as the testing set (RMSE for the testing set is shown). The results for the neural networks were obtained as an average of 10 trainings from random initial weights

Method of prediction	RMSE [°C]	
	Time delay PSR	Multivariate PSR
Persistent	2.43	2.43
Multiple linear regression	2.28	1.99
Local linear models	2.28	1.68
Multilayer perceptron NN	2.29	1.70
Radial basis function NN	2.30	1.63

The probability distribution functions of temperatures and pressures are different, and in the case of temperatures, strongly non-Gaussian. This difference may be viewed as a cause of increased nonlinearity, observed in the case of multivariate phase space reconstruction done from both pressures and temperatures (case IX). However, the null hypothesis allowed the tested series to be modified by a nonlinear measuring function, therefore the difference in shapes of the distribution should not be a cause of the observed nonlinear behavior. To prove this, we repeated the tests from case IX with time series transformed to have zero mean, unit variance and normal (Gaussian) probability distribution. Rank-ordering of a series of Gaussian random numbers (Theiler et al., 1992) was applied for normalization of time series. The relative difference between RMSE for the original series and RMSE for surrogates remained almost unchanged when normalized series were used, showing that the different probability distributions are not a cause of the observed nonlinearity.

Finally, two remarks on the applied nonlinearity testing technique. First, although we used a rank-order test to distinguish between linear and nonlinear cases, the results would be virtually the same if we just checked if RMSE for the original series falls within the $\pm 2\sigma$ range of RMSE for surrogates (except for case IV b – consult Fig. 2 where 2σ ranges are displayed). Still, despite the fact that a Gaussian-based test seems to work quite well, the observed distributions of RMSE values are indeed not completely Gaussian, as seen from diagrams in Fig. 2, and the use of the rank-order test was therefore justified (another kind of nonparametric test, such as the sign test, could have been used to similar effect). Second, we would also like to emphasize that the absolute difference between results from the original series and surrogates is a quite important outcome of the nonlinearity test, as well as the level of significance (whether expressed in multiples of sigma or in some other way). Because the min–max range of results from surrogates can be very small for long series, even very small absolute differences between results for the original and surrogate series can lead to rejection (which may be technically correct, but the practical value of such a result is questionable – this is exactly what happened for all the tested 30-year-long univariate series). Even worse, such small differences can just be a result of some kind of imperfection of the surrogate series (periodicity artifacts, nonstationarity effects, ...), and they can easily cause spurious rejection. Keeping an eye on the values of absolute difference can help to recognize suspicious results, and to identify situations when nonlinearity is really noteworthy.

6. Conclusions

Six different univariate series and three multivariate systems were tested for nonlinearity, each of them in a longer (30 years) and shorter (10 years) version. Success of prediction one day ahead by the method of local linear models in the reconstructed phase space was used as the discriminating statistic. The tested multivariate systems showed substantial nonlinearity for both lengths of the series – it seems that nonlinear dependencies in the multivariate series of temperature and pressure are profound and the application of nonlinear methods of time series analysis is generally suitable and recommended.

The situation was different for the univariate series of daily temperature and pressure. Generally, the null hypothesis of a stationary linear generating process was rejected for the 30-year-long series of temperature (mean, maximum or minimum alike), whilst the 10-year-long series appeared linear. The tests also showed somewhat stronger nonlinearity for the series of atmospheric pressure than in the case of temperature series. Nonetheless, differences between results from the original series and surrogates were very small, regardless of the type of the series or its length. For the analysis of individual univariate series, at least the ones we have examined, the use of nonlinear methods does not seem to be beneficial.

The influence of the length of the prediction time on the nonlinearity tests was investigated as well. The results suggest that nonlinearity is best expressed when the success of short-range prediction is applied as the discriminating statistic. With an increase of the prediction time, detectable signs of nonlinearity weakened, and they disappeared completely for long-term prediction. This implies the dependence of the degree of observed nonlinearity on the performed task, not just the series itself.

Hence, climatic time series reflect the nonlinear nature of the climate system in general, but whether nonlinearity is noticeable depends on the tested system as well as the method applied. This is what we would like to emphasize – nonlinearity as such may be present in the climate system, but sometimes it remains well hidden and insignificant from the point of view of practical time series analysis. Generally, it appears that nonlinear behavior is more distinct in longer series and that multivariate systems show noticeably greater nonlinearity than single series.

Acknowledgements

This work was supported by the Grant Agency of Charles University, grant 227/2002/B-GEO/MFF, Ministry of the environment of the Czech Republic, grant VaV/740/2/03 and Research project 113200004. We would also like to express our gratitude to the Czech hydrometeorological institute for providing the series of measurements from the Czech meteorological stations.

References

Abarbanel HDI (1996) Analysis of observed chaotic data. New York: Springer, 272 pp

Casdagli M (1989) Nonlinear prediction of chaotic time series. Physica D 35: 335–356

Casdagli M (1997) Characterizing nonlinearity in weather and epilepsy data: a personal view. In: Cutler CD, Kaplan DT (eds) Nonlinear dynamics and time series, Building a bridge between the natural and statistical sciences. Rhode Island: Amer Math Soc 252, pp. 201–222

Farmer JD, Sidorowich J (1987) Predicting chaotic time series. Phys Rev Lett 59: 845–848, reprinted in Ott et al. (1994)

Galka A (2000) Topics in nonlinear time series analysis: With implications for EEG analysis. Singapore: World Scientific Publishing, 342 pp

Haykin S (1999) Neural Networks: A comprehensive foundation, 2nd edn. Upper Saddle River, New Jersey: Prentice-Hall, 842 pp

Hegger R, Kantz H, Schreiber T (1999) Practical implementation of nonlinear time series methods: The TISEAN package. CHAOS 9: 413–435 [Available online at http://www.mpipks-dresden.mpg.de/~tisean/TISEAN_2.1/index.html]

Kantz H, Schreiber T (1997) Nonlinear time series analysis. Cambridge: Cambridge University Press, 304 pp

Keppenne CL, Nicolis C (1989) Global properties and local structure of the weather attractor over Western Europe. J Atmos Sci 46: 2356–2370

Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M (2001) The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. Bull Amer Meteor Soc 82: 247–268

Kugiumtzis D (1999) Test your surrogate data before you test for nonlinearity. Phys Rev E 60: 2808–2816

Ott E, Sauer T, Yorke JA (eds) (1994) Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems. New York: John Wiley & Sons, 418 pp

Packard HN, Crutchfield JP, Farmer JD, Shaw RS (1980) Geometry from a time series. Phys Rev Lett 45: 712–716, reprinted in Ott et al. (1994)

Paluš M (1996) Detecting nonlinearity in multivariate time series. Phys Lett A 213: 138–147

Paluš M, Novotná D (1994) Testing for nonlinearity in weather records. Phys Lett A 193: 67–74

Schreiber T (1998) Constrained randomization of time series data. Phys Rev Lett 80: 2105–2108

Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. Phys Rev Lett 77: 635–638

Schreiber T, Schmitz A (2000) Surrogate time series. Physica D 142: 346–382

Sivakumar B (2004) Chaos theory in geophysics: past, present and future. Chaos Soliton Fract 19: 441–462

Sugihara G, May RM (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature 344: 734–741

Takens F (1981) Detecting strange attractors in turbulence. In: Rand DA, Zouny LS (eds) Dynamical systems and turbulence. Lecture Notes in Mathematics 898: 366–381

Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: The method of surrogate data. Physica D 58: 77–94, reprinted in Ott et al. (1994)

Tsonis AA (2001) Probing the linearity and nonlinearity in the transitions of the atmospheric circulation. Nonlinear Proc Geoph 8: 341–345

Authors' address: Jiří Mikšovský (e-mail: jiri.miksovsky@mff.cuni.cz), Aleš Raidl (e-mail: ales.raidl@mff.cuni.cz), Department of Meteorology, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, 180 00 Prague 8, Czech Republic.

APPENDIX II

MIKŠOVSKÝ, J., P. PIŠOFT, AND A. RAIDL (2008), Global Patterns of Nonlinearity in Real and GCM-Simulated Atmospheric Data, in *Nonlinear Time Series Analysis in the Geosciences: Applications in Climatology, Geodynamics and Solar-Terrestrial Physics* (Eds.: Donner, R. V., and S. M. Barbosa), *Lecture Notes in Earth Sciences*, *112*, 17-34, doi:10.1007/978-3-540-78938-3_2.

Global Patterns of Nonlinearity in Real and GCM-Simulated Atmospheric Data

Jiří Mikšovský, Petr Pišoft, and Aleš Raidl

Department of Meteorology and Environment Protection, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic,
jiri.miksovsky@mff.cuni.cz; petr.pisoft@mff.cuni.cz;
ales.raidl@mff.cuni.cz

Abstract. We employed selected methods of time series analysis to investigate the spatial and seasonal variations of nonlinearity in the NCEP/NCAR reanalysis data and in the outputs of the global climate model HadCM3 of the Hadley Center. The applied nonlinearity detection techniques were based on a direct comparison of the results of prediction by multiple linear regression and by the method of local linear models, complemented by tests using surrogate data. Series of daily values of relative topography and geopotential height were analyzed. Although some differences of the detected patterns of nonlinearity were found, their basic features seem to be identical for both the reanalysis and the model outputs. Most prominently, the distinct contrast between weak nonlinearity in the equatorial area and stronger nonlinearity in higher latitudes was well reproduced by the HadCM3 model. Nonlinearity tends to be slightly stronger in the model outputs than in the reanalysis data. Nonlinear behavior was generally stronger in the colder part of the year in the mid-latitudes of both hemispheres, for both analyzed datasets.

Keywords: Nonlinearity, Reanalysis, Global climate model, Surrogates

1 Introduction

The Earth's climate system, as well as its atmospheric component, is an intrinsically nonlinear physical system. This nonlinearity is generally reflected in many series of climatic variables such as atmospheric pressure or temperature, but whether it is detectable and how strong it is depends on the type of the variable [1, 2, 3], geographic area of its origin [2, 4, 5, 6] or length of the signal [3]. The manifestations of nonlinearity in time series can be studied in numerous ways, using different statistics or criteria of the presence of nonlinear behavior. The techniques applied so far to meteorological data involve the calculation of the mutual information or persistence [1, 7, 8], statistics based on the performance of a nonlinear predictive method [3, 4, 9], nonlinear correlations [10] or the examination of the character of the prediction

residuals [2, 4, 5]. Tests using some form of surrogate data are frequently employed [1, 2, 3, 4, 7, 9, 10]. The presence of nonlinearity can also be assessed by comparing the performance of a linear and a nonlinear time series analysis method. In the atmospheric sciences, such studies are frequently associated with the application of statistical methods for prediction [6, 11, 12], or downscaling and postprocessing tasks [6, 13, 14, 15, 16]. Alongside with a wide spectrum of techniques for the detection of nonlinearity, different authors studied diverse types of signals, ranging from various variables related to the local temperature [1, 3, 6, 7, 10, 13, 14, 15, 16] or pressure [1, 2, 3, 4, 5, 7] to characteristics of larger-scale dynamics such as the mean hemispheric available potential energy [8]. Heterogeneity of the methods and datasets applied by different researchers makes it difficult to directly compare the results and use them to create a consistent global picture of the geographic variations of nonlinearity. However, it also seems that there are some systematic regularities in the spatial distribution of nonlinearity or of the related characteristics [2, 5, 6, 10, 17]. Here, we investigate this matter further, using a comparison of the results of linear and nonlinear prediction and tests based on the surrogate data.

A significant portion of the existing studies dealing with the issue of nonlinearity in time series focus on the analysis of individual scalar signals, typically employing time delayed values for the construction of the space of predictors or phase space reconstruction. Due to the complex behavior the atmosphere exhibits, and the relatively small size of the available records, the information content in a single series is limited and often insufficient for an effective application of nonlinear techniques. But meteorological measurements are frequently available for more than one variable, and they are carried out at multiple locations. When a multivariate system is used instead of a single scalar series, more information about the local state of the climate system can be obtained. It also seems that multivariate systems exhibit a generally stronger detectable nonlinear behavior [3]. For these reasons, and because using multiple input variables is common in many tasks of statistical meteorology and climatology, we focused on settings with multivariate predictors in this study. We restricted our attention to just a few of the available variables, defining the temperature and pressure structure of the atmosphere. The two illustrative cases presented here are based on forecasts of daily values of the relative topography 850-500 hPa (which is closely related to the temperature of the lower troposphere) and of the geopotential height of the 850 hPa level (one of the variables characterizing the structure of the field of atmospheric pressure). Along with investigating the character of the series derived from actual measurements (NCEP/NCAR reanalysis), attention was paid to the potential of the global climate model HadCM3 to reproduce the structures detected in the observed data. This should help to assess whether such simulation is able to capture not just the basic characteristics of the Earth's climate, but also the eventual nonlinear features of the respective time series. The utilized datasets are presented in Sect. 2, the techniques applied to quantitatively evaluate

nonlinearity are described in Sect. 3. Section 4 is devoted to the study of the spatial variations of nonlinearity. Section 5 focusses on the influence of the presence of the annual cycle in the series and seasonal changes of the detected patterns. Finally, in Sect. 6, the results are discussed with regard to their possible physical cause and practical implications. Color versions of the presented maps of the geographical distribution of nonlinearity (Figs. 3, 5, 6, 8 and 9) can be accessed at http://www.miksovsky.info/springer2008.htm.

2 Data

Direct atmospheric observations and measurements suffer from a number of potential problems. Their locations are typically unevenly spaced and coverage of some areas of the Earth is limited. Data from different sources are often incompatible and sometimes flawed. This restricts the usability of raw measurements for an analysis such as ours, the goal of which is to derive globally comparable results. To avoid or reduce the aforementioned problems, we used a gridded dataset in this study instead of direct measurements – the NCEP/NCAR reanalysis [18, 19] (hereinafter NCEP/NCAR). The reanalysis is a dataset derived from measurements at weather stations, as well as inputs from rawinsondes, meteorological satellites and other sources. The input observations are processed by a fixed data assimilation system, including a numerical forecast model, and the resulting series are available in a regular horizontal grid of 2.5° by 2.5°. Here, daily values of the geopotential height of the 850 hPa level (hereinafter H850) and 500 hPa level have been employed in a reduced 5° by 5° horizontal resolution, for the period between 1961 and 2000. From the values of the geopotential heights, the relative topography 850-500 hPa (RT850-500) has been computed. This quantity describes the thickness of the layer between the 850 hPa and 500 hPa levels and it is proportional to its mean virtual temperature. According to the classification used by Kalnay et al. [18], geopotential heights fall into the A–category of variables, thus reflecting the character of actual measurements rather than the specific properties of the model applied to create the reanalysis. A typical example of the analyzed series of RT850-500 in the equatorial area and in the mid-latitudes is shown in Fig. 1.

The recently increased interest in climate change instigated an intensive development of the models of the global climate. These simulations, to be reasonably realistic, must describe all key components of the climate system as well as the connections among them. As a result, the models are very complex and demanding with respect to the required computational resources. But despite their sophistication, no model is able to mimic the observed climate with absolute accuracy. A very important task in climate modeling is therefore validating the models, i.e., assessing their ability to reproduce the real climate. The common validation procedures are usually based on the basic statistical characteristics of the model outputs; here, we focus on the ability of a climate

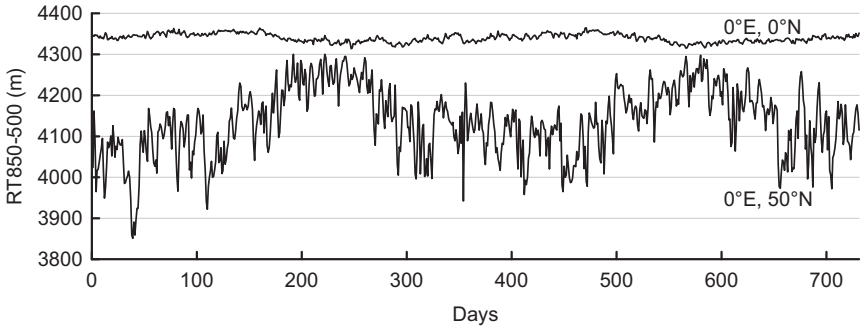


Fig. 1: A section of the analyzed data: Time series of daily values of the relative topography 850-500 hPa in the equatorial area (0°E, 0°N) and in the mid-latitudes (0°E, 50°N), for the years 1991 and 1992.

simulation to produce time series with the same nonlinear qualities as the real climate. For this task, we chose one of the major global climate models, HadCM3 of the Hadley Centre [20, 21]. The model outputs were used in a reduced horizontal resolution of 3.75° (longitude) by 5° (latitude). The model integration employed here was based on the observed concentrations of the greenhouse gasses and estimates of past changes in ozone concentration and sulfur emissions prior to the year 1990, and the emission scenario SRES B2 afterwards [21]. Since we only used the period from 1961 to 2000 for our analysis, and there is just very little difference among the SRES scenarios in the 1990s, the specific scenario choice should not be crucial.

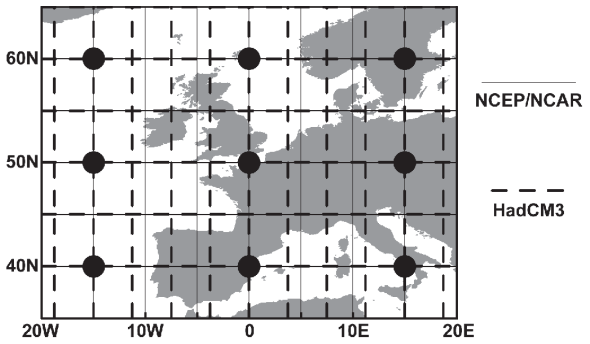


Fig. 2: An example of the structure of the pattern of predictors, displayed for the predictand series located at 0°E, 50°N. Black circles mark the positions of the predictors, the grid illustrates the reduced horizontal resolution of the NCEP/NCAR and HadCM3 data, used in this study.

3 Methods

3.1 General Settings

One of the key issues of the multivariate approach to the construction of the space of predictors is the selection of a suitable set of input variables. Unlike for some simple low-dimensional dynamical systems, a perfect phase-space reconstruction is impossible from climatic time series, due to the complexity of the underlying system. In the case of practical time series analysis tasks, a finite-dimensional local approximation of the phase space may suffice. To predict values of a scalar series in some grid point, we used a pre-set pattern of predictors, centered on the location of the predictand and spanning 30° in longitude and 20° in latitude (Fig. 2). A different configuration of predictors was chosen for each of the two tasks presented: In the case of the RT850-500 forecast, the dimension of the predictor space was $N = 18$, with 9 values of RT850-500 and 9 values of H850 in a configuration shown in Fig. 2. For the forecast of H850, 9 predictors were used, all of which were of the H850 type. Note that, despite the different spatial resolution of the NCEP/NCAR reanalysis and the HadCM3 model, the selected pattern of predictors could be applied for both of them directly, without interpolating the data to a common grid.

All predictors $x_i(t), i = 1, \ldots, N$, were transformed to have zero mean and standard deviation equal to $\sqrt{\cos\varphi}$, using the linear transformation $x_i(t) \to \sqrt{\cos(\varphi)}(x_i(t) - \overline{x}_i)/\sigma_i$ (φ being latitude of the respective grid point, $\overline{x}_i$ mean value of the predictor series and σ_i its standard deviation). Hence, the predictor's variance was proportional to the size of the area characterized by the corresponding grid point. The presented results were derived from the outcomes of prediction one day ahead, carried out for grid points located between 70°N and 70°S (the areas closest to the poles were excluded from the analysis, due to the severe deformation of the applied spatial pattern of predictors in high latitudes).

3.2 Direct Comparison-Based Approach

Our primary technique of quantification of nonlinearity was based on a direct comparison of the root mean square errors (RMSEs) of prediction by a linear reference method, multiple linear regression, and by its nonlinear counterpart, the method of local linear models. In the case of linear regression, the value of the scalar predictand y at time $t+1$ was computed as a linear combination of the values of individual predictors $x_i, i = 1, \ldots, N$, in the previous time step

$$\hat{y}(t+1) = a_0 + \sum_{i=1}^{N} a_i x_i(t)\,, \tag{1}$$

where the coefficients $a_j, j = 0, \ldots, N$, were calculated to minimize the sum of the squared values of the residuals $\hat{y}(t) - y(t)$.

Even a nonlinear system can be described rather well when the linear model is applied locally for smaller portions of the phase space instead of a global linear approximation. This concept has been successfully utilized for the construction of forecast models for many different types of time series. Several related studies are reprinted in [22] and the basic principles of the method of local models are also described, e.g., in [23]. The dynamics in the individual regions of the input space is approximated by linear mappings based on (1), but an individual linear predictive model (or a set of coefficients a_i, respectively) is constructed for each value of t. To create such a local model, only a certain number M of the predictors-predictand pairs, representing the states of the system most similar to the one at time t, is employed to compute the coefficients. The similarity of individual states was quantified by the distance of the respective N-dimensional vectors of predictors $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_N(t))$ here, using the Euclidian norm.

To calculate the out-of-sample root mean square error of the prediction, the analyzed series were divided into two subintervals. The years 1961–1990 were used as a calibration set, i.e., for the computation of the coefficients of the above described models. These were then tested for the years 1991–2000. The values of RMSE we obtained for the prediction by multiple linear regression ($RMSE_{\mathrm{MLR}}$) and local linear models ($RMSE_{\mathrm{LM}}$) were compared by computing

$$SS_{\mathrm{LM}} = 1 - (RMSE_{\mathrm{LM}}/RMSE_{\mathrm{MLR}})^2 \,, \tag{2}$$

which will be referred to as the local models' skill score. Its definition is based on the commonly used concept of a skill score, described, e.g., in [24]. SS_{LM} vanishes when both methods perform equally well in terms of RMSE and it equals to one for a perfect forecast by local models (presuming that $RMSE_{\mathrm{MLR}} \neq 0$). The number M of predictors-predictand pairs used for the computation of the coefficients of the local models is one of the adjustable parameters of the method of local models. Depending on the specific structure of the local climate system, different values of M may be suitable to minimize RMSE. Here, local models constructed with $M = 250$, 500 and 1000 were tested for each grid point; the variant giving the lowest RMSE was then used in the subsequent analysis.

3.3 Surrogate Data-Based Approach

The above described approach yields results which are interesting from a practical perspective, but, strictly speaking, it only refers to a relation of two particular techniques, both of which may have their specifics. Another method, which does not rely on comparing different mappings, exists. It uses modified series (so-called surrogate series or surrogates), which preserve selected properties of the original signal, but are consistent with some general null hypothesis. Here, the hypothesis is that the data originates from a linear

Gaussian process, the output of which may have been modified by a static monotonic nonlinear filter. The values of a nonlinearity-sensitive statistic are then compared for the original series and multiple surrogates, and if a statistically significant difference is detected, the null hypothesis is rejected. It should be noted that the formal rejection does not necessarily prove the presence of nonlinearity in the signal, as it can be caused by other reasons, such as nonstationarity of the series or imperfection of the surrogate-generating procedure. For details see, e.g., [9], where the principles of the surrogate data-based tests are presented in depth, or [25], where the usability of several methods of generating surrogates is discussed for various geophysical data.

For each grid point, 10 surrogates were created from the respective multivariate system of time series. Prediction by the method of local linear models was carried out for each of the surrogates and an arithmetic average $RMSE_{\mathrm{SURR}}$ of the resulting RMSEs was computed. A skill score-based variable, analogous to (2), was then calculated using RMSE for the original series $RMSE_{\mathrm{LM}}$ and $RMSE_{\mathrm{SURR}}$:

$$SS_{\mathrm{SURR}} = 1 - (RMSE_{\mathrm{LM}}/RMSE_{\mathrm{SURR}})^2. \tag{3}$$

In order to keep the computational demands at a reasonable level, the surrogate data-based analysis was performed just for $M = 250$. Also, the years 1991–2000 were used for both calibration and testing of the mappings. The surrogate series were generated by the iterative amplitude adjusted Fourier transform [26] in its multivariate form [9]; the program package TISEAN by Hegger et al. [27] was applied for this task.

4 Spatial Patterns of Nonlinearity

Figure 3a shows the geographical distribution of the local models' skill score SS_{LM}, obtained for the NCEP/NCAR RT850-500 forecast. The most prominent feature of the detected pattern is the strong latitudinal variance of nonlinearity. Near the equator, just very small and mostly statistically insignificant difference between the performance of purely linear regression and local linear models was found. Nonlinear behavior becomes visibly stronger in the midlatitudes, and it is more pronounced on average in the northern hemisphere, where major nonlinearity was detected for all grid points north of circa 25°N (Fig. 4). In the southern hemisphere, the strongest nonlinearity was located in a band approximately between 25°S and 50°S. This structure seems to be well reproduced by the HadCM3 model (Fig. 3b), although the nonlinear behavior is slightly stronger in the model data in the northern hemisphere – see Table 1, columns 1 and 2. The spatial correlation of the SS_{LM} fields for the NCEP/NCAR and HadCM3 data was evaluated by computing the Pearson correlation coefficient, after linear interpolation of the HadCM3 data-based values of SS_{LM} to the 5° by 5° grid of NCEP/NCAR. For the entire area

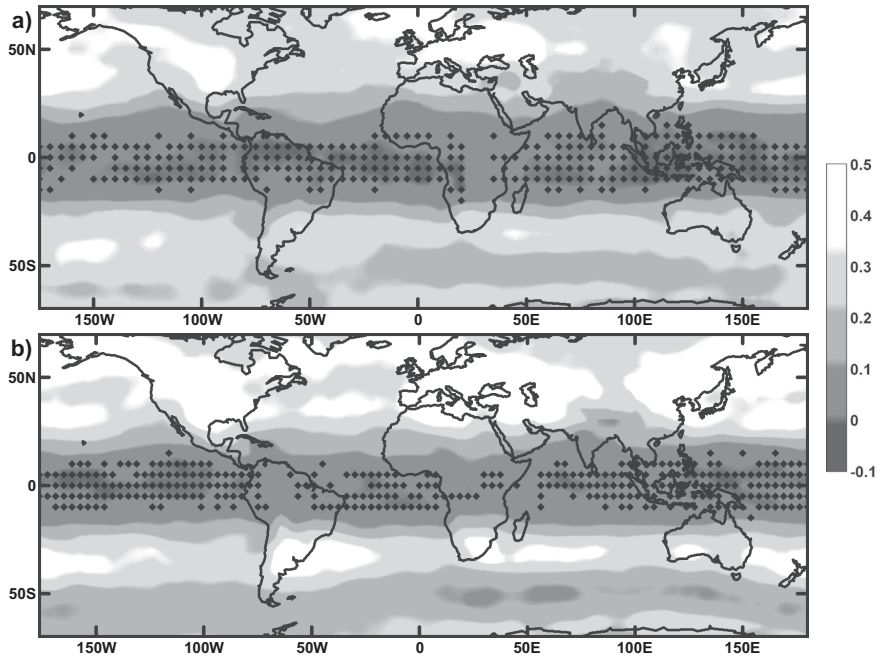


Fig. 3: Geographical distribution of the local models' skill score SS_{LM}, obtained for the RT850-500 prediction, using the NCEP/NCAR (**a**) and HadCM3 (**b**) data. Diamonds mark the positions of the grid points where daily errors of prediction by the method of local models were not statistically significantly lower than for linear regression at the 95% confidence level, according to the one-sided paired sign test.

between 70°N and 70°S, the correlation was 0.91. When just extratropical areas were taken into account, the resemblance of the SS_{LM} patterns was stronger in the northern hemisphere than in the southern one (Table 1, column 3). Similar values of correlation were also obtained when the Spearman rank-order correlation coefficient was used instead of the Pearson one.

Aside from the dominant latitudinal dependence, the detected nonlinearity patterns also exhibited a distinct finer structure. As can be seen in Fig. 3a for the NCEP/NCAR reanalysis data, local maxima of nonlinearity were found over Europe, North America, East Asia and the northern part of the Pacific Ocean, and east of the landmasses of the southern hemisphere. The HadCM3 data yielded a very similar pattern (Fig. 3b). After the average latitudinal structure was filtered out by subtracting the respective latitudinal averages from the values of SS_{LM} in every grid point, the spatial correlation of the NCEP/NCAR and HadCM3 SS_{LM} patterns was still rather high, though the resemblance of both fields was clearly stronger in the northern hemisphere (Table 1, column 4).

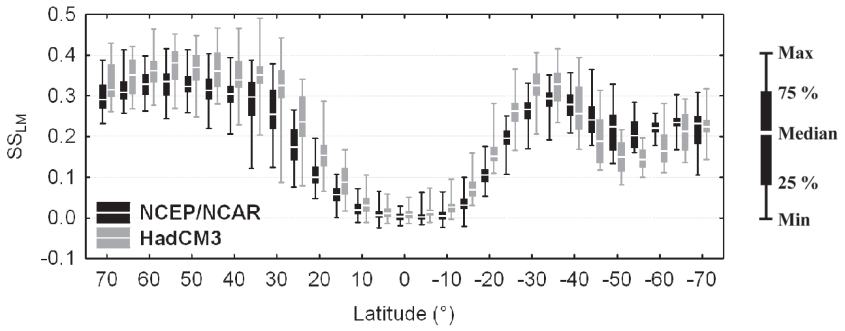


Fig. 4: Distribution of SS_{LM} in different latitudes, obtained for the RT850-500 prediction (latitude values are positive north of the equator).

Table 1: Regional averages of SS_{LM}, obtained for the RT850-500 prediction in the case of the NCEP/NCAR (column 1) and HadCM3 (column 2) data and spatial correlations of the NCEP/NCAR and HadCM3 SS_{LM} patterns for the original values of SS_{LM} (column 3) and after the average latitudinal dependence has been filtered out (column 4).

	SS_{LM}		Correlation	
Region	NCEP/NCAR	HadCM3	Original	Filtered
25°N–70°N	0.29	0.34	0.75	0.60
20°S–20°N	0.04	0.06	0.89	0.45
70°S–25°S	0.24	0.23	0.55	0.48

When the results of the H850 prediction were applied as a basis for a nonlinearity detection, a somewhat different pattern emerged (Fig. 5). The basic latitudinal structure with very weak nonlinearity in the equatorial area was still present, but other details of the detected structure differed from the ones found for the RT850-500 prediction. In the northern hemisphere, maximum values of SS_{LM} were located over the northwestern part of the Atlantic Ocean and the adjacent part of North America, as well as over the northern part of the Pacific Ocean. Both these maxima were rather well expressed, while the rest of the northern hemisphere exhibited weaker nonlinearity. In the southern hemisphere, the maxima of SS_{LM} were less localized. The overall degree of nonlinearity was lower than for the RT850-500 prediction (Table 2, columns 1 and 2). The similarity of the patterns obtained from the NCEP/NCAR and HadCM3 data was again very strong, with a value of global spatial correlation of 0.9. The nonlinearity was stronger on average in the HadCM3 outputs than in the NCEP/NCAR reanalysis. As for the match of the patterns of SS_{LM} with filtered-out latitudinal dependence, there was still a high positive correlation, stronger in the northern hemisphere (Table 2, column 4).

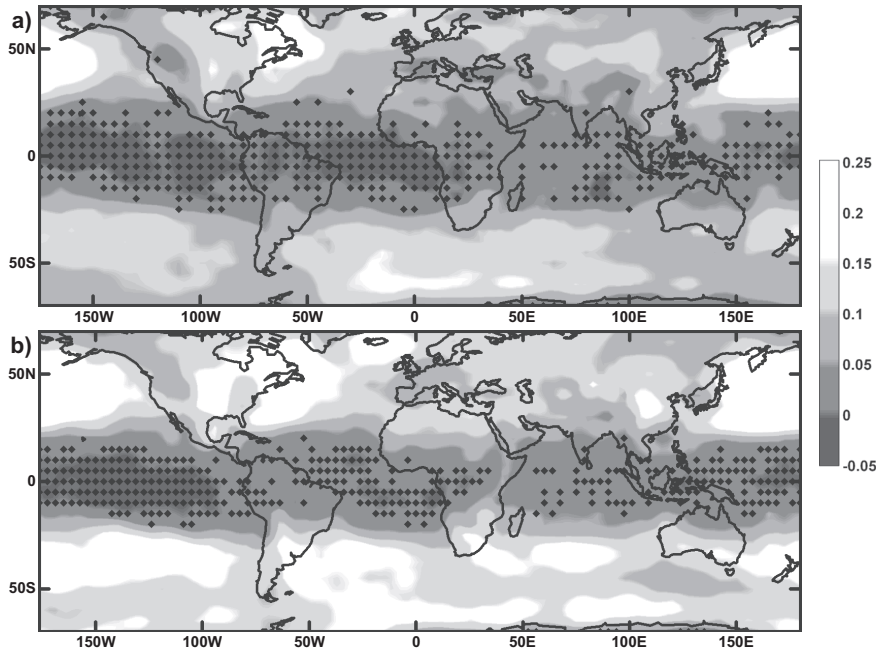


Fig. 5: Same as Fig. 3, for the prediction of H850 instead of RT850-500.

Table 2: Same as Table 1, for the prediction of H850 instead of RT850-500.

	SS_{LM}		Correlation	
Region	NCEP/NCAR	HadCM3	Original	Filtered
25°N–70°N	0.10	0.13	0.84	0.85
20°S–20°N	0.02	0.03	0.71	0.52
70°S–25°S	0.09	0.14	0.75	0.71

As can be seen from Fig. 6, the pattern of nonlinearity obtained for the RT850-500 prediction by means of surrogate data and expressed through SS_{SURR} is very similar to the one presented above for the direct comparison technique (Fig. 3a). To illustrate the distribution of RMSE in the ensemble of surrogates, a more detailed example of the outcomes is shown in Fig. 7 for the grid points along the 0° meridian. The results for the HadCM3 data are not shown, but they also confirm the outcomes of the direct comparison of multiple linear regression and local linear models. Similarly, surrogate data-based verification of the results derived from the H850 prediction showed no major differences either.

It should be mentioned that when an identical setting is used for direct comparison-based and surrogate data-based tests, including an equal size of the calibration set, SS_{LM} is systematically smaller than SS_{SURR}. The reason

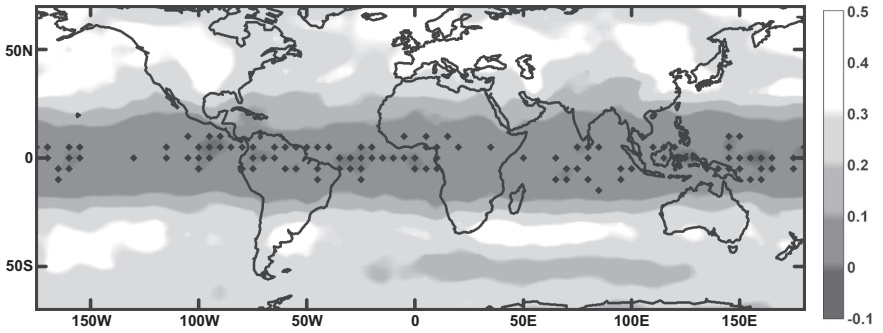


Fig. 6: Geographical distribution of SS_{SURR}, obtained for the RT850-500 prediction, using the NCEP/NCAR data. Diamonds mark positions of the grid points, where the value of RMSE for the original series was not smaller than for all 10 surrogates. This is equivalent to the non-rejection of the hypothesis of a linear Gaussian generating process at the confidence level of about 91%, according to the usually applied one-sided rank-order test, described, e.g., in [9]. Testing at a higher confidence level would require more surrogates, but even then, the results would be almost identical, as additional tests have shown for selected individual grid points.

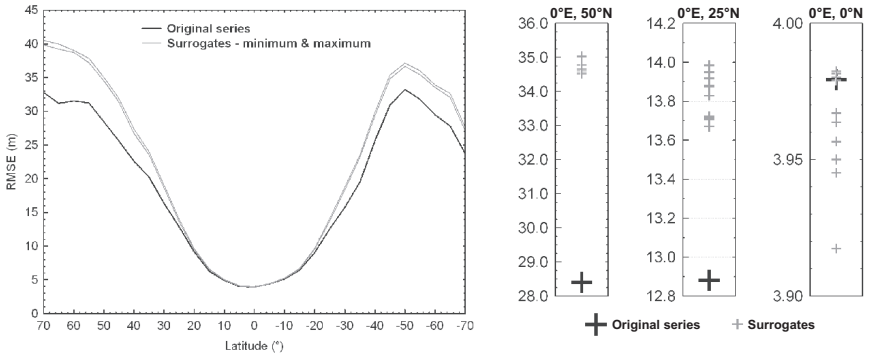


Fig. 7: *Left panel*: RMSE of the RT850-500 prediction by the local linear models method and its range for the respective surrogate series, for grid points at 0°E (NCEP/NCAR data). *Right panels*: Values of RMSE (m) obtained for the original series and individual surrogates in the three selected grid points.

for this difference is related to the behavior of the method of local models for purely linear series. When the processed signal contains no deterministic nonlinear component (like surrogates do) and M is smaller than the size of the calibration set, the method of local models performs slightly worse than linear regression. Our choice of a shorter calibration set for the surrogate data-based tests (Sect. 3.3) has actually partly compensated for this shift, because the magnitude of detected nonlinearity generally decreases with the reduction of the size of the calibration set.

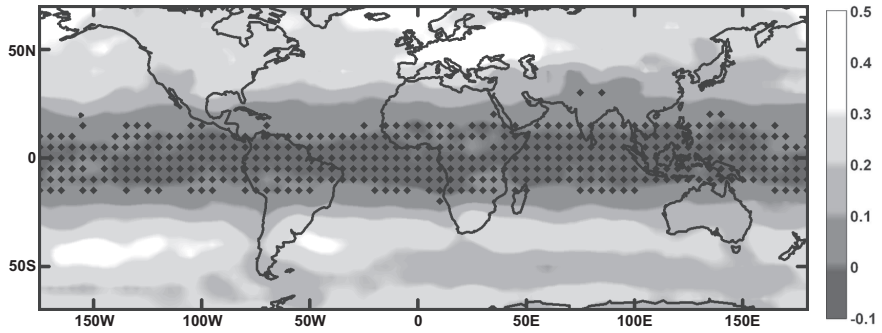


Fig. 8: Same as Fig. 3a, for series with removed annual cycle.

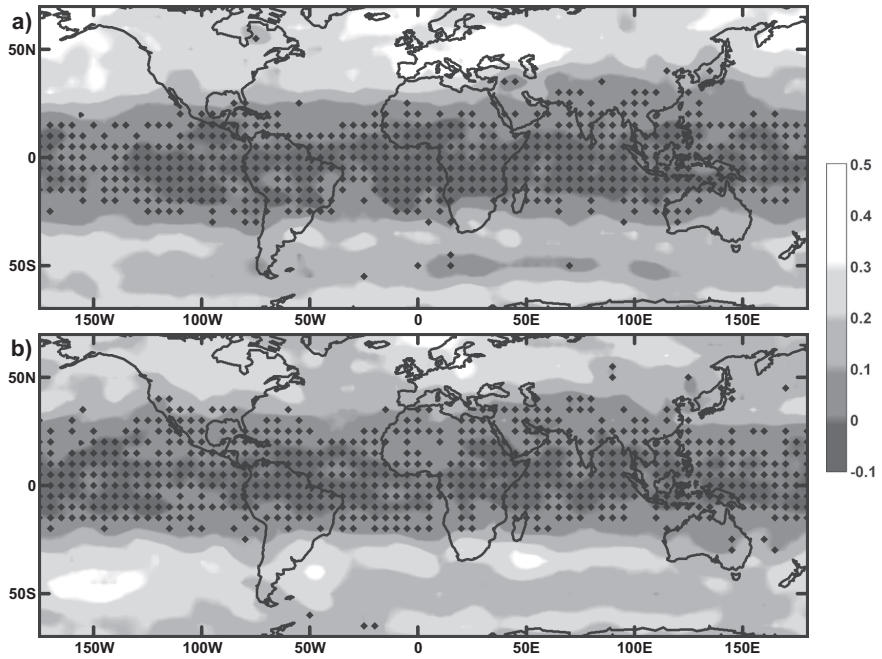


Fig. 9: Same as Fig. 3a, for the DJF (**a**) and JJA (**b**) seasons (winter and summer in the northern hemisphere) instead of the entire year.

5 Seasonal Variations of Nonlinearity

The annual cycle is among the strongest oscillations in the climate system. It dominates series of many climatic variables, especially in higher latitudes (see example in Fig. 1). This also means that the geographical areas with a well-defined annual cycle coincide to some degree with the regions, where strong nonlinearity was detected. To assess the possible relationship, we repeated some of the tests for the series with removed annual cycle. The removal was carried out by subtracting the mean climatological annual cycle of the respective variable, computed for the years 1961–2000 and smoothed by an 11–day moving average. An example of the results is shown in Fig. 8, for the RT850-500 prediction. As a comparison to Fig. 3a reveals, the values of SS_{LM} generally decreased after the annual cycle removal. Although this change was relatively small on average (e.g., the average value of SS_{LM} decreased from 0.29 to 0.23 in the area north of 25°N, and from 0.24 to 0.21 south of 25°S), it was profound in the regions with the highest amplitude of the annual cycle of RT850-500. For instance, the maximum of SS_{LM}, originally detected over East Asia and the adjacent part of the Pacific Ocean, disappeared almost completely. In the southern hemisphere, the changes associated with the annual cycle removal were generally smaller. In the case of the H850 prediction, the shape of the pattern of SS_{LM} remained practically identical for the annual cycle-free series, though the average degree of nonlinearity also slightly decreased.

In many situations, the annual cycle cannot be treated as simply an oscillation superposed to the variations at other time scales. Different seasons are associated with different atmospheric dynamics in many regions, and properties of the analyzed time series, including their eventual nonlinearity, may thus periodically vary throughout the year. Because of this, the analysis of climatic data is often performed separately for different parts of the year, typically seasons or months. We used this approach to investigate the seasonal variations of SS_{LM}. The results below are shown for the parts of the year corresponding to climatological winter (December, January and February – DJF) and summer (June, July and August – JJA) of the northern hemisphere. When the analysis was carried out for separate seasons, the RMSE of the prediction by linear regression decreased for most grid points in the annual average. The performance of the method of local models usually became worse, primarily due to the reduction of the amount of data available for the calibration of the mappings. As a result, the average magnitude of nonlinearity decreased somewhat, compared to the situation when the series were analyzed as the whole. Despite this change, the basic features of the patterns of SS_{LM} were still the same, as can be seen from an example of the results based on the RT850-500 forecast (Fig. 9). In the equatorial area, the nonlinearity remained very weak or undetectable in all seasons. In higher latitudes, the patterns retained some of the basic shape, detected for the year as the whole, but their magnitude visibly varied with the season. The overall nonlinearity was stronger in the

DJF season than in JJA in the northern hemisphere, while in higher latitudes of the southern hemisphere, this variation was reversed and JJA exhibited stronger nonlinearity than DJF on average (Fig. 10, Table 3, columns 1 and 2). The seasonal changes were stronger expressed in the northern hemisphere. The seasonal variation was well simulated by the HadCM3 model (Table 3, columns 3 and 4) and it was also detectable in the results based on the forecast of H850, for both the NCEP/NCAR and HadCM3 data (not shown).

6 Discussion

All performed analyses revealed a common basic latitudinal structure with just negligible nonlinearity in the equatorial regions, but generally stronger nonlinear behavior in the mid-latitudes of both hemispheres. A detailed analysis of the factors behind the observed patterns might be problematic, because they do not seem to be a result of a single driving force, but rather their complex combination. There are, however, some possible links worth mentioning. In the case of the results based on the RT850-500 prediction, there may be a connection between more pronounced nonlinearity in the mid-latitudes and the activity of the polar front. The strongest nonlinear behavior over Europe and North America seems to coincide with the position of the zones where air masses of different origin often interact. In the southern hemisphere, where the landmasses are less extensive, areas of the strongest nonlinearity are typically located rather east of the continents, possibly because of the interaction of the landmass with the prevailing westerlies. Between approximately 50°S and 60°S, where the amount of land is very small, nonlinearity is weaker on average. A removal of the annual cycle from the series slightly decreases the magnitude of detected nonlinearity, but except for the regions where the annual variation is very strong (East Asia), the effect of the annual cycle presence does not dominate the results. For the H850 forecast, there appears to be a certain connection of the areas with strong nonlinearity to the zones of high horizontal gradient of H850. In the northern hemisphere, such areas are typically associated with deep stationary cyclones, which are usually present over the North Atlantic and North Pacific during winter. The match is not perfect though, and there may be some other factors involved. Altogether, it seems that nonlinearity tends to be stronger in the regions with more complex dynamics, where strong driving or perturbing factors are in effect. This hypothesis is supported by the fact that nonlinearity is generally more pronounced during the colder season in the mid-latitudes of both hemispheres, i.e., in situations when the temperature gradient between the equatorial area and the polar region is strongest. The fact that the seasonal variations are more distinct in the northern hemisphere is probably an effect of the uneven distribution of the continents, resulting in a larger influence of the continental climate in the northern mid-latitudes.

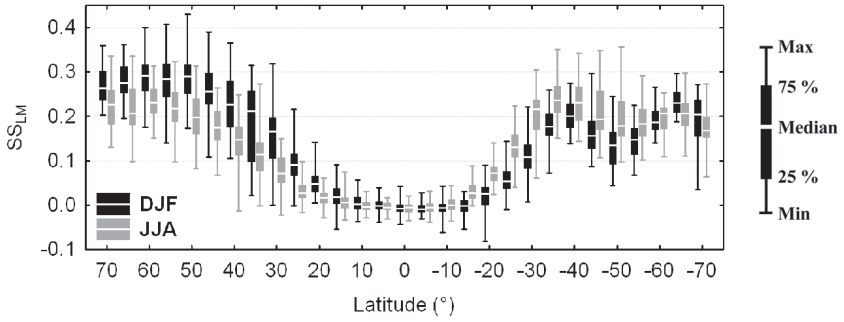


Fig. 10: Distribution of SS_{LM}, obtained for the prediction of RT850-500 for the DJF and JJA seasons in different latitudes (NCEP/NCAR data).

Table 3: Seasonal variations of nonlinearity (expressed by SS_{LM}) in the NCEP/NCAR and HadCM3 data, for the RT850-500 prediction.

	NCEP/NCAR		HadCM3	
Area	DJF	JJA	DJF	JJA
25°N–70°N	0.23	0.16	0.27	0.20
20°S–20°N	0.01	0.01	0.02	0.02
70°S–25°S	0.16	0.20	0.16	0.18

The two presented cases, based on the prediction of geopotential height and relative topography one day ahead, represent just a fraction of possible settings. From additional tests, carried out for different predictand-predictors combinations, it seems that the basic structure with weak nonlinearity in the equatorial area is typical for most situations. On the other hand, the finer details of the detected patterns vary, especially with the type of the predictand. The exact number and geographical configuration of predictors seem to be less important, as long as they sufficiently characterize the local state of the atmosphere. Beside the type of the studied variables, we also paid attention to the sensitivity of the results to the specific details of the tests. It appears that the results are rather robust to the changes of the size of the source area of predictors, although a use of a too big or too small area leads to a general increase of the prediction error and a weakening of detected nonlinearity. The outcomes remain very similar when the input data are pre-processed by principal component analysis, instead of using the point-wise predictors directly. The method of eventual normalization of the predictors also does not appear to be of major importance. The observed patterns of nonlinearity seem to be rather stable in time, i.e., the specific choice of the analyzed period does not have any major effect on the outcomes of the tests. The relatively most distinct changes compared to the presented results were detected in the

NCEP/NCAR data when the 1960s were chosen as the testing set instead of the years 1991–2000, especially in the southern hemisphere. This difference can probably be contributed to the variations in the amount of observational data, entering the reanalysis, as discussed below. The applied tests were all based on prediction one day ahead – with an increase of the lead time, nonlinearity quickly weakened and it became undetectable for predictions more than approximately five days ahead, even in the regions where the nonlinear behavior was originally strongest. This is in good agreement with the fact that a deterministic weather prediction is impossible for too long lead times, regardless of the method.

Most of the patterns of nonlinearity identified in the NCEP/NCAR reanalysis data were also found in the outputs of the HadCM3 model. From the perspective of applied nonlinear time series analysis tasks (such as statistical downscaling carried out by nonlinear methods), the fact that a climate model is able to reproduce the character of the observed data is encouraging. Still, from the results obtained for a single representative of global climate models, it is not possible to infer whether all existing climate simulations do behave in a similar fashion. It is interesting that the correspondence of the structures found in the NCEP/NCAR and HadCM3 data tends to be better in the northern hemisphere. Although this fact can at least partially be a consequence of the specifics of the model's physics, it might also be contributed to the character of the reanalysis data. To assess the possible influence of the specific properties of the NCEP/NCAR reanalysis, we repeated some of the tests for another commonly used gridded dataset based on observations, the ERA-40 reanalysis [28]. Although some differences were found, the resemblance of the results from the NCEP/NCAR and ERA-40 data was generally strong in the northern hemisphere, but somewhat weaker in the southern one. This implies that caution is needed in interpretation of the model-reanalysis differences, particularly in the southern latitudes, as they may be a result of a limited amount of observational data used by the reanalysis (and possibly some other specifics of the NCEP/NCAR dataset), not just imperfections of the climate model. This especially applies to the period preceding the era of meteorological satellites – e.g., the amount of data entering the NCEP/NCAR reanalysis is very low before the year 1979 south of approximately 40°S [19].

We have shown that the direct comparison of prediction by linear regression and by local linear models yields nonlinearity patterns very similar to the approach based on the application of local linear models for surrogate data. A practical advantage of the direct comparison lies in its speed, as there is no need for multiple realizations of a nonlinear model. This is especially convenient in the case of an analysis like ours, carried out for thousands of grid points and repeated for numerous settings. Another benefit of the direct comparison is that it provides specific information about the potential gain from employing a nonlinear method; its fundamental drawback is that such information may only be valid for the combination of the methods applied.

7 Conclusions

By analyzing the series of selected atmospheric variables, we were able to confirm the presence of systematic geographical and seasonal variations of nonlinearity. Simple and unequivocal physical explanation of the results beyond the basic tropics/mid-latitudes and summer/winter contrast may be problematic, because the finer details of the detected patterns are probably a product of multiple influences and they are subject to the type of the predictand variable and some other factors. To find out whether any other general regularities exist would require a systematic analysis performed for a large number of variables and pressure levels. Regardless of the exact cause of the detected structures, their character was simulated fairly well by the HadCM3 model. From the practical perspective, this finding is rather promising, as it confirms that data produced by the current generation of global climate models can be utilized for the study of nonlinear properties of the climate system.

Acknowledgements. This study was supported by the Czech Science Foundation (grant 205/06/P181) and by the Ministry of Education of the Czech Republic (research plan MSM0021620860). The presented work would not be possible without the utilized datasets: NCEP/NCAR reanalysis (obtained from NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov), HadCM3 model outputs (provided by the Met Office Hadley Centre) and ERA-40 reanalysis (obtained from the Data Server of the European Centre for Medium-Range Weather Forecasts). The authors would also like to express their gratitude to the two anonymous reviewers of the manuscript and to R. Donner for their valuable comments and suggestions.

References

1. M. Paluš, D. Novotná: Testing for nonlinearity in weather records, Phys. Lett. A **193**, 67 (1994)
2. D. A. S. Patil, B. R. Hunt, J. A. Carton: Identifying low-dimensional nonlinear behavior in atmospheric data, Mon. Weather Rev. **129**, 2116 (2001)
3. J. Miksovsky, A. Raidl: Testing for nonlinearity in European climatic time series by the method of surrogate data, Theor. Appl. Climatol. **83**, 21 (2006)
4. M. C. Casdagli: Characterizing Nonlinearity in Weather and Epilepsy Data: A Personal View. In:*Nonlinear Dynamics and Time Series*, ed by C. D. Cutler, D. T. Kaplan (American Mathematical Society, Providence, Rhode Island 1997) pp 201–222
5. G. Sugihara, M. Casdagli, E. Habjan, et al.: Residual delay maps unveil global patterns of atmospheric nonlinearity and produce improved local forecasts, P. Natl. Acad. Sci. USA **96**, 14210 (1999)
6. J. Miksovsky, A. Raidl: Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures, Nonlinear Proc. Geoph. **12**, 979 (2005)
7. M. Paluš: Detecting nonlinearity in multivariate time series, Phys. Lett. A **213**, 138 (1996)

8. A. A. Tsonis: Probing the linearity and nonlinearity in the transitions of the atmospheric circulation, Nonlinear Proc. Geoph. **8**, 341 (2001)
9. T. Schreiber, A. Schmitz: Surrogate time series, Physica D **142**, 346 (2000)
10. I. Bartos, I. M. Jánosi: Nonlinear correlations of daily temperature records over land, Nonlinear Proc. Geoph. **13**, 571 (2006)
11. V. Pérez-Muñuzuri, I. R. Gelpi: Application of nonlinear forecasting techniques for meteorological modeling, Ann. Geophysicae **18**, 1349 (2000)
12. B. Tang, W. W. Hsieh, A. H. Monahan, F. T. Tangang: Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures, J. Climate **13**, 287 (2000)
13. A. Weichert, G. Bürger: Linear versus nonlinear techniques in downscaling, Climate Res. **10**, 83 (1998)
14. J. T. Schoof, S. C. Pryor: Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, Int. J. Climatol. **21**, 773 (2001)
15. M. Casaioli, R. Mantovani, F. P. Scorzoni, et al.: Linear and nonlinear post-processing of numerically forecasted surface temperature, Nonlinear Proc. Geoph. **10**, 373 (2003)
16. E. Eccel, L. Ghielmi, P. Granitto, et al.: Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models, Nonlinear Proc. Geoph. **14**, 211 (2007)
17. W. von Bloh, M. C. Romano, M. Thiel: Long-term predictability of mean daily temperature data, Nonlinear Proc. Geoph. **12**, 471 (2005)
18. E. Kalnay, M. Kanamitsu, R. Kistler, et al.: The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc. **77**, 437 (1996)
19. R. Kistler, E. Kalnay, W. Collins, et al.: The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation, Bull. Amer. Meteor. Soc. **82**, 247 (2001)
20. C. Gordon, C. Cooper, C. A. Senior, et al.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, Clim. Dynam. **16**, 147 (2000)
21. T. C. Johns, J. M. Gregory, W. J. Ingram, et al.: Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios, Clim. Dynam. **20**, 583 (2003)
22. E. Ott, T. Sauer, J. A. Yorke (eds.): *Coping with Chaos: Analysis of Chaotic Data and The Exploitation of Chaotic Systems* (John Wiley & Sons, New York 1994)
23. H. Kantz, T. Schreiber: *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge 1997)
24. D. S. Wilks: *Statistical Methods in the Atmospheric Sciences*, 2nd edn (Elsevier, Amsterdam 2006)
25. V. Venema, S. Bachner, H.W. Rust, C. Simmer: Statistical characteristics of surrogate data based on geophysical measurements, Nonlinear Proc. Geoph. **13**, 449 (2006)
26. T. Schreiber, A. Schmitz: Improved surrogate data for nonlinearity tests, Phys. Rev. Lett. **77**, 635 (1996)
27. R. Hegger, H. Kantz, T. Schreiber: Practical implementation of nonlinear time series methods: The TISEAN package, CHAOS **9**, 413 (1999)
28. S. M. Uppala, P. W. Kållberg, A. J. Simmons, et al.: The ERA-40 re-analysis, Quart. J. R. Meteorol. Soc. **131**, 2961 (2005)

APPENDIX III

MIKŠOVSKÝ, J., AND A. RAIDL (2005), Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures, *Nonlinear Processes in Geophysics*, *12*(6), 979-991.

Nonlinear Processes in Geophysics, 12, 979–991, 2005
SRef-ID: 1607-7946/npg/2005-12-979
European Geosciences Union

Nonlinear Processes in Geophysics

Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures

J. Miksovsky and A. Raidl

Dept. of Meteorology, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Received: 16 June 2005 – Revised: 15 September 2005 – Accepted: 15 September 2005 – Published: 9 November 2005

Abstract. We investigated the usability of the method of local linear models (LLM), multilayer perceptron neural network (MLP NN) and radial basis function neural network (RBF NN) for the construction of temporal and spatial transfer functions between different meteorological quantities, and compared the obtained results both mutually and to the results of multiple linear regression (MLR). The tested methods were applied for the short-term prediction of daily mean temperatures and for the downscaling of NCEP/NCAR reanalysis data, using series of daily mean, minimum and maximum temperatures from 25 European stations as predictands. None of the tested nonlinear methods was recognized to be distinctly superior to the others, but all nonlinear techniques proved to be better than linear regression in the majority of the cases. It is also discussed that the most frequently used nonlinear method, the MLP neural network, may not be the best choice for processing the climatic time series – LLM method or RBF NNs can offer a comparable or slightly better performance and they do not suffer from some of the practical disadvantages of MLPs.

Aside from comparing the performance of different methods, we paid attention to geographical and seasonal variations of the results. The forecasting results showed that the nonlinear character of relations between climate variables is well apparent over most of Europe, in contrast to rather weak nonlinearity in the Mediterranean and North Africa. No clear large-scale geographical structure of nonlinearity was identified in the case of downscaling. Nonlinearity also seems to be noticeably stronger in winter than in summer in most locations, for both forecasting and downscaling.

Correspondence to: J. Miksovsky
(jiri.miksovsky@mff.cuni.cz)

1 Introduction

Within the last two decades, numerous new methods of time series analysis have been developed for dealing with nonlinear data (see, e.g. Abarbanel, 1996; Kantz and Schreiber, 1997; Haykin, 1999; Galka, 2000 for an overview), and a lot of them have found their place in the study of meteorological signals (see examples in Sect. 4). But it has also been shown that application of nonlinear methods does not automatically grant better results than use of their linear counterparts (e.g. Tang et al., 2000), despite the fact that the meteorological series originate from an inherently nonlinear system. Application of nonlinearity tests reveals some climatic data sets to appear linear (Palus and Novotna, 1994; Schreiber and Schmitz, 2000; Miksovsky and Raidl, 2005), while others may exhibit nonlinear characteristics (Palus and Novotna, 1994; Palus, 1996; Tsonis, 2001; Miksovsky and Raidl, 2005). In this paper, our intention was to address the problem of nonlinearity of the atmospheric time series from a rather practical point of view and to ascertain the performance of several nonlinear methods of time series analysis for two typical meteorological problems: construction of temporal (prediction) and spatial (downscaling) mappings at synoptic time scales. The examined methods included a nonlinear technique which has already become common in the atmospheric sciences (multilayer perceptron neural network – MLP NN), as well as methods which are less common, at least to date – local linear models in the reconstructed phase space (LLM) and radial basis function neural networks (RBF NN). Performance of nonlinear methods was compared both to each other and to the results of multiple linear regression (MLR). Aside from presenting examples of the obtained results, including their spatial and seasonal variances, we have tried to draw conclusions about the tested methods' disadvantages and strong points, as well as the pros and cons associated with their implementation.

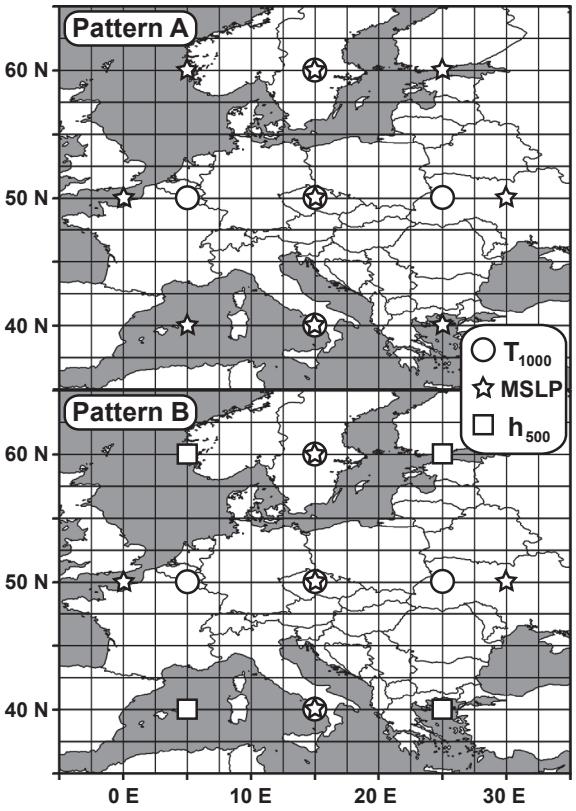


Fig. 1. The pattern of predictors used in the case of prediction (Pattern A) and downscaling (Pattern B), displayed for the predictand series located at 50° N, 15° E. The grid of horizontal and vertical lines illustrates the full resolution of the NCEP/NCAR reanalysis data set.

2 Methods

2.1 Choice of predictors

The first problem which needs to be addressed before time series analysis methods can be applied is the issue of the structure of the predictor space. The climate system, as well as its local subsystems, has an infinite number of degrees of freedom, meaning that an infinite number of variables would be required for its state exact description. But there are methods by which the state can be characterized, at least approximately and locally, in a relatively low number of quantities. Techniques referred to as phase space (PS) reconstruction represent a way of achieving such a description. The most classical method of phase space reconstruction, time delay reconstruction from a scalar series (Packard et al., 1980; Takens, 1981), has been applied for analysis of climatic time series at a number of occasions, from early attempts to discover some kind of climate attractor (among many e.g. Fraedrich, 1986; Keppenne and Nicolis, 1989), to its practical implementations for the forecast of meteorological or hydrological variables (e.g. Abarbanel, 1996; Pérez-Muñuzuri and Gelpi, 2000; Jayawardena and Gurung, 2000). Many more examples from various climate-related disciplines can be found in the paper by Sivakumar (2004). But it also turns out that the information content in a single time series is not always sufficient for the climate system's state characterization. This is especially true when the nonlinear component of the analyzed signal is to be studied (Miksovsky and Raidl, 2005). Fortunately, meteorological measurements (or numerical model outputs, reanalyses and similar data sets) are typically available for several variables and in numerous locations, which allows for the use of multivariate phase space reconstruction (Keppenne and Nicolis, 1989). The vector in the reconstructed phase space, characterizing the system's state in time t (and representing the vector of predictors), is denoted $\boldsymbol{y}(t)$ here,

$$\boldsymbol{y}\,(t) = (X_1\,(t)\,, X_2\,(t)\,,\ \ldots\,,\ X_m\,(t))\,, t = 1, \ldots, L\,, \qquad (1)$$

where m characterizes the dimension of the reconstructed PS and is usually referred to as the embedding dimension, L is the length of the series and $X_i(t)$, i=1, …, m, are elements of $\boldsymbol{y}(t)$. For multivariate reconstruction from m scalar series $x_i(t)$, elements of $\boldsymbol{y}(t)$ take the following simple form:

$$X_i\,(t) = x_i\,(t)\,, i = 1, \ldots, m\,. \qquad (2)$$

An important issue is the choice of suitable predictor series, i.e. specifying which series, and how many, will be used in the role of $x_i(t)$. We used NCEP/NCAR reanalysis series as potential predictors here, meaning that thousands of series of different quantities from numerous grid points and pressure levels were available, whereas only a few predictors were needed. Choice of predictors is a nontrivial problem and it can be done in several ways such as using step-wise regression or reducing the dimensionality of the predictor space by means of principal component analysis, but no approach can generally be considered to be the absolutely best one. Also, use of different sets of predictors may sometimes result in quite different outcomes, as shown by Huth (2004) for temperature downscaling. We have tested several methods for predictors selection and decided to utilize a pre-set pattern of input variables, consisting of values of T_{1000}, MSLP and h_{500} from different grid points (see Sect. 3 for data sets description). Use of the pre-set pattern is fast and easy to implement; it does not favor the MLR method like the use of the step-wise linear regression could and it also gave better results (in terms of root mean squared error – RMSE) than use of the principal components as predictors, similarly to the findings of Huth (2002). Moreover, using the same type and number of predictors for different tested methods, locations and seasons makes intercomparison of the results easier, because the composition of the predictor space need not be taken into account as one of the variables. The patterns used for prediction (Sect. 4.1) and downscaling (Sect. 4.2) both had dimension m=14; their structure is shown in Fig. 1.

2.2 Methods used

Mapping from the predictor vector $\boldsymbol{y}(t)$ to predictand $x^{PRED}(t)$ (transfer function) can be generally expressed as

$$x^{PRED}(t) = \Phi(\boldsymbol{y}(t)) = \Phi(X_1(t), X_2(t), ..., X_m(t)) . \quad (3)$$

The exact form of Φ depends on the mapping construction technique employed. We tested four methods here, one linear and three nonlinear. Their brief descriptions are in Sects. 2.2.1 (multiple linear regression), 2.2.2 (local linear models), 2.2.4 (MLP neural networks) and 2.2.5 (RBF neural networks). The respective forms of Φ are represented by Eqs. (4), (5), (6) and (8).

2.2.1 Multiple linear regression

Linear methods still represent the most frequently used tool of time series analysis. They are usually less complicated than their nonlinear counterparts, with lower demands regarding computational power, and, unlike nonlinear methods, without many parameters to be determined prior to their application. We used multiple linear regression (MLR) here as a typical representative of linear techniques. The mapping had the form

$$x^{PRED}(t) = v_0 + \sum_{i=1}^{m} v_i X_i(t) , \quad (4)$$

where v_i, i=0, ..., m, are the regression coefficients.

2.2.2 Method of local linear models

Origin of the method of local models dates back to the second half of the eighties and it is associated with research focused on the problems of chaotic dynamics, strange attractors and phase space reconstruction. The method was shown to be suitable for the prediction of low-dimensional chaotic systems (Farmer and Sidorowich, 1987), as well as simple physical (e.g. Sauer, 1993) or biological (e.g. Sugihara and May, 1990) systems. In meteorology, too, its applications have been demonstrated, for example, for cloud coverage forecast (Pérez-Muñuzuri and Gelpi, 2000) or rainfall prediction (Sivakumar et al., 2000). A detailed description of the method and more examples of its application can be found in the monographs by Abarbanel (1996) or Kantz and Schreiber (1997).

Since successful phase space reconstruction enables characterization of the system's state by an m-dimensional vector like Eq. (1), it is also possible to quantify the similarity of different states, typically by computing the Euclidian distance of their respective vectors $\boldsymbol{y}$. In order to construct a mapping approximating the time evolution from time t or spatial relation between different variables in time t, a certain number N of states $\boldsymbol{y}(t, j)$, j=1, ..., N, is found in the history of the system as states with the smallest distance to $\boldsymbol{y}(t)$. From the relation between such states and the corresponding values $x(t, j)$ of the predictand, a mapping can be constructed which approximates some local neighbourhood of $\boldsymbol{y}(t)$ in the phase space by a linear model:

$$x^{PRED}(t) = v_0(t) + \sum_{i=1}^{m} v_i(t) X_i(t) . \quad (5)$$

Note that, unlike in Eq. (4), coefficients $v_i(t)$ are not time-invariant and they need to be computed separately for each time t. Computation of m+1 coefficients $v_i(t)$ from N pairs of $\boldsymbol{y}(t, j)$ and $x(t, j)$ is a linear regression problem, solvable in the least-squares sense. It should be mentioned that using a fixed value of the number of nearest neighbors N is not the only possible way of defining local neighborhood in the phase space. It is also possible to work with the directly specified size of the neighborhood (e.g. Hegger et al., 1999), or to pick an individual value of N for every t, as done by Jayawardena et al. (2002). Here, however, all local models were realized using a constant N for the entire analyzed series.

2.2.3 Artificial neural networks

Neural networks (NNs) have become very popular in various scientific areas as a convenient tool for many practical tasks, including time series analysis and data processing. Typical artificial neural network is a complex structure, consisting of some number of interconnected, simple signal processing units – artificial neurons (or, shortly, neurons). Neurons typically have several inputs and a single output; the information received by inputs is processed by the neuron and the outcome is then transmitted to its neighbors. For data processing tasks, so-called feedforward NNs are the ones most frequently applied. A typical feedforward NN consists of several layers of neurons – one, called the input layer, which receives inputs, then one or more hidden layers of signal-processing neurons, and finally, an output layer in which the results are computed to their final form. The output layer can comprise of one or more neurons (producing one or more output values simultaneously), but we only used single-output networks here. Two different types of feedforward networks were studied in this paper: Multi-Layer Perceptrons (MLPs) and Radial Basis Function (RBF) networks. For more information on NNs, see, e.g. monographs by Haykin (1999) or Principe et al. (2000).

2.2.4 Multilayer perceptron neural networks

MultiLayer Perceptrons (MLPs) are by far the best known and most frequently used design of neural networks – to such an extent that they are sometimes (incorrectly) viewed as being equivalent to NNs generally. The operation performed by neurons in the hidden layer of MLP is a weighted summation. For MLP with one hidden layer and a single neuron in the output layer, the transfer function can be expressed as

$$x^{PRED}(t) = v_0 + \sum_{i=1}^{M} v_i f\left(w_{0i} + \sum_{j=1}^{m} w_{ji} X_j(t)\right), \quad (6)$$

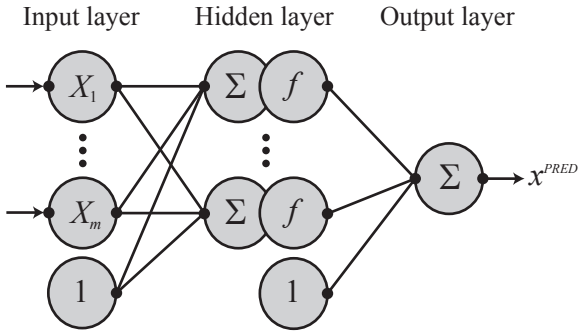


Fig. 2. The schematic structure of multilayer perceptron neural network.

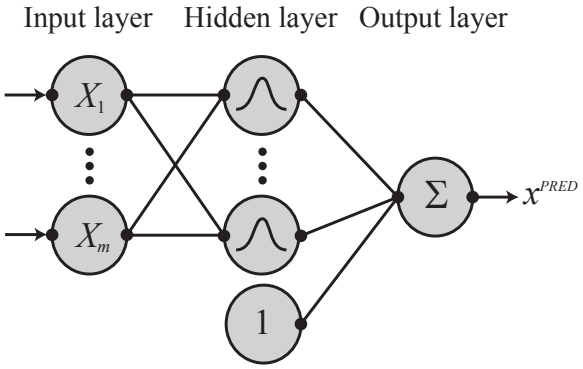


Fig. 3. The schematic structure of radial basis function neural network.

where w_{ji} is the weight of the connection between the i-th neuron in the hidden layer and the j-th neuron in the input layer, v_i is the weight of the connection between the i-th neuron in the hidden layer and the neuron in the output layer, and M is the number of neurons in the hidden layer (structure of a MLP network is schematically shown in Fig. 2). Function $f(x)$, called the activation function, is generally nonlinear. We used the logistic function,

$$f(x) = \frac{1}{1 + \exp(-x)} . \tag{7}$$

Apparently, the character of the mapping in Eq. (6) is given by the values of the weights v_i and w_{ji}, which have to be determined before the network can be used. Values of weights are computed using examples of input-output pairs from some training set (supervised learning). The basic learning technique, backpropagation of errors, is an iterative procedure using values of errors for the training cases to calculate the weights adjustments. The backpropagation technique exists in several variants, differing in their implementation complexity, needed learning time, and tendency to be trapped in a local minimum of the error function instead of reaching the global minimum. We have used the quasi-Newton method for MLPs training (e.g. Haykin, 1999), with weights initialized to uniformly distributed random values.

2.2.5 Radial basis function neural networks

In many aspects similar to MLPs, radial basis function neural networks (RBF NNs – Fig. 3) are feedforward networks with one hidden layer and one or more neurons in the output layer. There are two principal differences between MLPs and RBF networks. First, instead of the weighted summation of the inputs, performed by the neurons in the hidden layer of MLPs, RBF networks employ radial basis functions, usually the Gaussian ones. The transfer function can be expressed as

$$x^{PRED}(t) = v_0 + \sum_{i=1}^{M} v_i \exp\left(-\frac{\| \boldsymbol{y}(t) - \boldsymbol{c}_i \|^2}{2\sigma_i^2}\right), \tag{8}$$

where $\boldsymbol{c}_i$ is the position of the centre of the i-th radial basis function (assigned to the i-th neuron in the hidden layer, i=1, …, M), σ_i characterizes the width of the i-th RBF, and the meaning of v_i and M is the same as in Eq. (6).

The second major difference between RBF and MLP neural networks is in the learning algorithm. While weights in MLPs are determined in some sort of iterative learning procedure, training of RBF NNs can be done in two separate phases. First, positions and shapes of the radial basis functions are specified (i.e. $\boldsymbol{c}_i$ and σ_i, i=1, …, M). The simplest way to do this is to take M randomly chosen times $t_1, t_2, \ldots, t_M$ from the training set and use the corresponding vectors in the input space as centres of RBFs, $\boldsymbol{c}_i = \boldsymbol{y}(t_i)$, i=1, …, M. A more sophisticated way of setting $\boldsymbol{c}_i$, k-means clustering (MacQueen, 1967; Haykin, 1999), did not notably improve the results in our case. Values of σ_i can be set individually for each neuron, but in praxis a single value is often used, $\sigma_i = \sigma$, i=1, …, M, which we followed. As soon as the positions of RBF centres and sigma are set, finding the values of weights v_i, i=0, …, M, is a linear problem which can be easily solved using the least-squares approach.

2.3 Computation settings

When testing a time series analysis method, it is important to construct the mapping from one part of the data set, and then to test its performance on an independent interval. For neural networks, the former set is usually referred to as a training set while the latter is called a testing set. We use the same convention for LLM and MLR methods, too.

All utilized nonlinear methods here need one or more parameters to be determined before they can be applied (number of nearest neighbors N for LLM method, number of hidden neurons M for both types of neural networks, width of radial functions σ for RBF networks). The parameters were chosen to give the lowest RMSE for the testing set (a range of parameter values was tested and the best performing one was then used for the actual computation). RMSE dependence on the above mentioned parameters typically had a broad, flat minimum and so it was relatively easy to pick

their optimal values. An exception to this rule was observed for the number of hidden neurons in MLPs, when the dependence of RMSE on M exhibited notable fluctuations, due to the sensitivity of the results to the initial values of the network's weights. This was compensated for by repeated training (see below), but even then the error curve was visibly less smooth than for RMSE as a function of N or RBF NNs' M.

Optimization of the parameters was done for the entire year as a whole, not separately for individual seasons. For different tasks, the number of nearest neighbors for the LLM method ranged between 200 and 700. As for the MLP neural networks, the best results were typically obtained with the number of hidden neurons 8 to 12, but sometimes as high as 20. In the case of RBF NNs, the optimal number of hidden units was much higher, usually between 200 and 400. The width of radial functions was set to σ=3.2 (very little sensitivity of the results to its value was observed, so we kept it fixed for all tasks). Training of neural networks, both MLP and RBF, was performed 5 times from random initial weights (or RBF centres' positions, respectively) and the presented results are an average of this five-member ensemble. In retrospect, repeated learning was probably not necessary for the RBF networks, since variance of the results within the ensemble was quite low. In the case of MLPs, a single realization could profoundly misrepresent the performance of the method, so using an entire ensemble was important to reduce the effect of sensitivity to the initial values of weights. Input values for all neural networks were normalized to range [0,1] by subtracting the minimum of the series and then dividing the value with the series' max-min range. MLP networks were trained for 1500 epochs by the quasi-Newton method.

An important part of the input data is information about the season of the year. In order to introduce this information to the computations, application of the MLR and LLM methods was done separately for each season. For NNs, we did perform the training for the entire training set as a whole, and information about the season was introduced by four extra neurons in the input layer, each of which was assigned to one season (the effective dimension of the input space was therefore m=18 for all neural networks). These neurons' activations were equal to 1 for the season they controlled, and 0 otherwise (i.e. just one of these four neurons was active for a given time – the one assigned to the corresponding season). Another possibility of introducing information about the season would be using sine and cosine of the Julian day (as done by Trigo and Palutikof, 1999). The usual climatological definition of the seasons was utilized:

$$\begin{aligned}
\text{Winter} &= \text{December} + \text{January} + \text{February} &&= \text{DJF},\\
\text{Spring} &= \text{March} + \text{April} + \text{May} &&= \text{MAM},\\
\text{Summer} &= \text{June} + \text{July} + \text{August} &&= \text{JJA and}\\
\text{Autumn} &= \text{September} + \text{October} + \text{November} &&= \text{SON}.
\end{aligned}$$

3 Data

The first data set utilized was the NCEP/NCAR reanalysis by Kistler et al. (2001), obtained from the page of NOAA-CIRES CDC at http://www.cdc.noaa.gov. It is available for the years from 1948 on, and it covers an entire world with reanalysis of several meteorological variables in the regular 2.5°×2.5° grid, including data for various pressure levels. Series of mean daily temperature at the level 1000 hPa (T_{1000}) were used here, as well as series of mean sea level pressure (MSLP) and series of geopotential height of the 500 hPa level (h_{500}).

As for directly measured meteorological series, perhaps the largest publicly available data set of European temperature, precipitation and pressure measurements was collected by Klein Tank et al. (2002) and it is obtainable from the Internet page of the European Climate Assessment and Dataset (ECA&D – http://eca.knmi.nl/). Measurements from many different sources were assembled by the authors, so series of various quality and length are part of ECA&D, and many of them contain missing values. We used series of daily mean, minimum and maximum temperature from 25 European stations – see Table 1 for their list. Many more series can be obtained from ECA&D – these 3×25 were selected in order to cover most of Europe with series available for the years 1961 to 2000, or ending not too long before the year 2000, and containing as little missing values as possible.

4 Results

4.1 Prediction

Forecast of future weather is what most people view as a fundamental purpose of meteorology's existence. And although today's weather forecasts are made almost exclusively by means of numerical models, there are many supplementary tasks for which time series analysis can be more suitable, for example, because the available data do not allow for use of a generally demanding NWF model, because of unsuitable spatial or temporal scale or due to a lack of available computing capacity. It was shown that nonlinear methods can be applied for tasks like prediction of road temperatures (Shao, 1998), precipitation forecasting (Waelbroeck et al., 1994; Hall et al., 1999; Sivakumar et al., 2000), or forecast of cloud cover (Pérez-Muñuzuri and Gelpi, 2000), although in some cases nonlinear methods do not seem to be more suitable than the linear ones, as demonstrated, for example, by Tang et al. (2000) for Central Pacific SST forecast.

The gridded NCEP/NCAR reanalysis data set represents a suitable basis for study of spatial distribution of the predictive potential of nonlinear methods. First, we performed the prediction of the T_{1000} series one day ahead by the LLM and MLR methods for every grid point in the area between 65° N, 25° W and 25° N, 45° E, i.e. for 493 points in total. Years 1961–1990 were used as the training set, while the testing set consisted of the years 1991–2000. The structure of the predictor space is demonstrated in Fig. 1a for the grid point 50° N, 15° E (the pattern was moved to be centered on the location of the predicted series). Unlike for all the other computations in this paper, a fixed number of nearest

Table 1. A list of stations used for downscaling tests. Asterisk (*) marks the stations for which at least one of the series of daily mean, minimum or maximum temperature did not cover the entire period 1961–2000.

Station	Country	Latitude	Longitude	Alt. (m)	NCEP/NCAR nearest grid point
Beograd*	Yugoslavia	44°48′ N	20°28′ E	132	45°00′ N, 20°00′ E
Berlin	Germany	52°27′ N	13°18′ E	55	52°30′ N, 12°30′ E
Bordeaux	France	44°50′ N	00°41′ W	47	45°00′ N, 00°00′
Bologna	Italy	44°29′ N	11°15′ E	60	45°00′ N, 10°00′ E
Bremen	Germany	53°03′ N	08°47′ E	4	52°30′ N, 10°00′ E
Brindisi*	Italy	40°38′ N	17°56′ E	10	40°00′ N, 17°30′ E
Eskdalemuir	UK	55°19′ N	03°12′ W	242	55°00′ N, 02°30′ W
Helsinki	Finland	60°10′ N	24°57′ E	4	60°00′ N, 25°00′ E
Heraklion*	Greece	35°20′ N	25°11′ E	39	35°00′ N, 25°00′ E
Hurbanovo	Slovakia	47°53′ N	18°12′ E	115	47°30′ N, 17°30′ E
Karlsruhe	Germany	49°01′ N	08°23′ E	114	50°00′ N, 07°30′ E
Kremsmuenster*	Austria	48°03′ N	14°08′ E	383	47°30′ N, 15°00′ E
Larissa*	Greece	39°39′ N	22°27′ E	74	40°00′ N, 22°30′ E
St. Petersburg*	Russia	59°58′ N	30°18′ E	6	60°00′ N, 30°00′ E
Linkoeping	Sweden	58°24′ N	15°32′ E	93	57°30′ N, 15°00′ E
Malaga	Spain	36°40′ N	04°29′ W	7	37°30′ N, 05°00′ W
Marseille	France	43°18′ N	05°24′ E	75	42°30′ N, 05°00′ E
Moskou*	Russia	55°50′ N	37°37′ E	156	55°00′ N, 37°30′ E
Oslo Blindern	Norway	59°57′ N	10°43′ E	94	60°00′ N, 10°00′ E
Oxford	UK	51°46′ N	01°16′ W	63	52°30′ N, 02°30′ W
Paris	France	48°49′ N	02°20′ E	75	50°00′ N, 02°30′ E
Praha	Czech Rep.	50°05′ N	14°25′ E	191	50°00′ N, 15°00′ E
Tortosa	Spain	40°49′ N	00°29′ E	48	40°00′ N, 00°00′
Utsira Fyr	Norway	59°18′ N	04°53′ E	55	60°00′ N, 05°00′ E
Valentia Observatory*	Ireland	51°56′ N	10°13′ W	9	52°30′ N, 10°00′ W

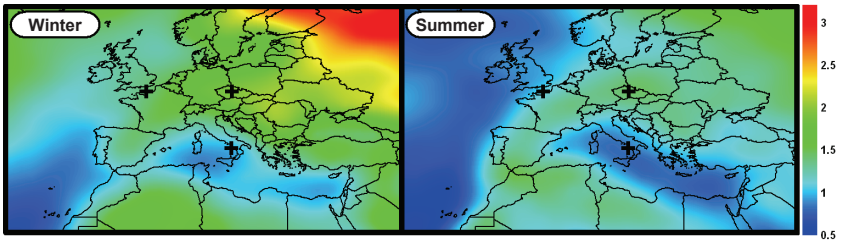


Fig. 4. RMSE (°C) of 1-day ahead prediction of temperature at the 1000 hPa level (T_{1000}) by LLM method. Three black crosses mark the positions of the grid points from Table 2.

neighbors was used for all grid points, N=250. The results are shown in Fig. 4 (RMSE for LLM method) and 5 (RMSE for LLM method, divided with RMSE for MLR, which we will call effective nonlinearity, since it quantifies an improvement which can be gained from replacing linear MLR with the nonlinear method of local models).

The lowest values of RMSE were detected over the Mediterranean Sea and Atlantic Ocean, and, in summer, in the Near East. Errors in summer were generally lower than in winter. The absolute values of the errors coincide well with values of average interdiurnal temperature change, i.e. errors were higher in the areas with higher temperature variability. When the spatial structure of effective nonlinearity was analyzed, the observed pattern was more complex. In winter, most of continental Europe appears to be a region of increased nonlinearity, with the highest differences between RMSE for the LLM and MLR methods in Western Europe and Russia. Nonlinearity seems to be rather weak in the maritime areas, as well as in Northern Africa and the Near East. In summer, nonlinearity is weaker than in winter in most areas, and there is a clear contrast between the Mediterranean area (very weak nonlinearity, except for Tunisia and Northern Libya) and the rest of Europe (rather stronger nonlinearity, especially in Northern France, Belgium and Netherlands). This observed pattern bears an interesting resemblance to the positions of the climatic zones, as nonlinearity

Table 2. RMSE (°C) of 1-day ahead prediction of NCEP/NCAR T_{1000} series for three different grid points. The values in bold indicate that the nonlinear method gave better results than multiple linear regression, according to the paired Wilcoxon test at the 95% confidence level. The values in the first row show RMSE of persistent prediction.

	50° N, 0° E				50° N, 15° E				40° N, 15° E			
	DJF	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**	**SON**	**DJF**	**MAM**	**JJA**	**SON**
Pers.	2.12	1.8	1.73	1.74	2.73	2.41	2.31	2.31	1.42	1.15	0.99	1.23
MLR	1.75	1.48	1.3	1.5	2.3	1.9	1.65	1.86	1.11	0.98	0.83	0.99
LLM	**1.41**	**1.28**	**1.07**	**1.21**	**1.93**	**1.57**	**1.44**	**1.54**	**0.96**	**0.88**	0.83	**0.88**
MLP	**1.52**	**1.35**	1.18	**1.24**	**1.9**	**1.54**	**1.46**	**1.57**	**0.98**	0.92	0.81	**0.89**
RBF	**1.39**	**1.26**	**1.07**	**1.15**	**1.87**	**1.52**	**1.38**	**1.54**	**0.95**	**0.87**	**0.79**	**0.88**

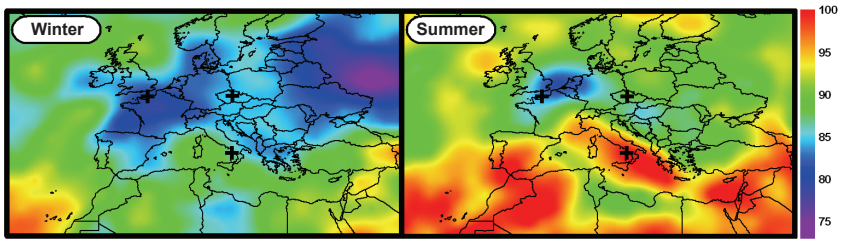


Fig. 5. Effective nonlinearity for 1-day ahead prediction of temperature at the 1000 hPa level (T_{1000}), i.e. RMSE for LLM method expressed in % of RMSE for multiple linear regression (lower values indicate stronger nonlinearity). Three black crosses mark the positions of the grid points from Table 2.

seems to be stronger in the temperate zone and weaker in the subtropical and tropical areas (various climate classifications can be found in Essenwanger, 2001).

There is a comparison of 1-day ahead prediction results in Table 2 for all tested methods and all seasons of the year for three selected NCEP/NCAR grid points: 50° N, 0° E (a grid point which exhibited strong effective nonlinearity in both summer and winter), 50° N, 15° E (medium nonlinearity in both seasons) and 40° N, 15° E (medium nonlinearity in winter, very weak in summer). Positions of the respective grid points are indicated by crosses in Figs. 4 and 5. Aside from comparing RMSEs, differences between individual methods were ascertained using the paired Wilcoxon test (e.g. Wilks, 1995), applied at the absolute values of daily errors. The significance of the differences between errors from the MLR and nonlinear methods is indicated in Table 2 – values in bold signal that the nonlinear method was better than MLR at the 95% level of confidence. RBF NN typically gave the best results of all three nonlinear methods in terms of RMSE, but the difference of error medians was not significant, according to the Wilcoxon test, in most of the tested cases. On the other hand, nonlinear methods, especially LLM and RBF NNs, typically gave better results than MLR, both with respect to RMSE and to the significance of the difference of daily errors.

Visually, there was just little difference between the series of predictions from all four tested methods. Superiority of nonlinear techniques in the geographic areas with strong effective nonlinearity was, however, clearly visible in the distribution of errors, with residuals close to zero being relatively more frequent for nonlinear methods than for MLR – see example in Fig. 6. What all methods had in common was a certain tendency to underestimate high values of temperature and overestimate the low ones, thus actually

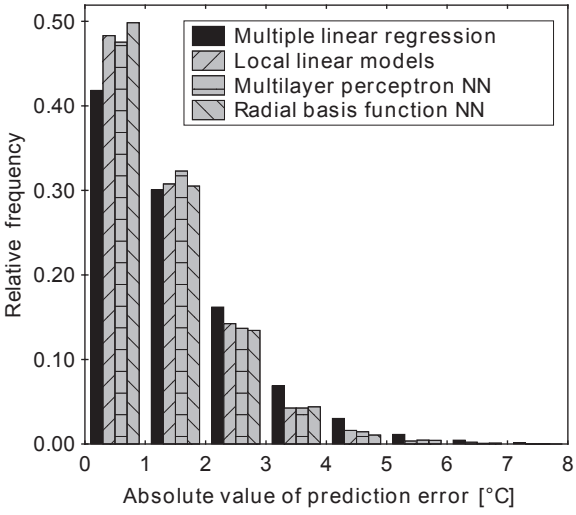


Fig. 6. Histogram of the absolute values of the prediction errors for 1-day ahead prediction of NCEP/NCAR T_{1000} series at 50° N, 15° E (whole year).

Table 3. RMSE (°C) of NCEP/NCAR reanalysis downscaling for four European stations. The values in bold indicate that the nonlinear method gave better results than multiple linear regression, according to the Wilcoxon test at the 95% confidence level. Underline indicates that the method performed better than all the other three, according to the Wilcoxon test. The values in the first row for each station show RMSE obtained by means of mean climatology.

		Mean temperature				Min. temperature				Max. temperature			
		DJF	MAM	JJA	SON	DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Praha	**Clim.**	4.68	3.96	3.44	3.53	4.75	3.45	2.71	3.50	4.81	4.71	4.31	4.07
	MLR	1.88	1.29	1.19	1.40	2.27	1.76	1.45	1.85	1.85	1.66	1.52	1.77
	LLM	**1.73**	**1.26**	1.17	**1.28**	2.23	1.72	1.44	**1.74**	**1.67**	1.66	1.52	1.69
	MLP	**1.77**	1.24	1.20	1.34	2.16	1.70	1.41	**1.75**	**1.72**	**1.63**	**1.52**	1.68
	RBF	**1.75**	**1.22**	1.21	**1.29**	**2.18**	1.70	1.42	**1.72**	**1.73**	1.63	**1.52**	1.68
Oslo	**Clim.**	4.67	3.46	2.84	3.38	5.01	3.35	2.71	3.83	4.44	4.24	3.62	3.52
	MLR	2.44	1.64	1.33	1.58	2.96	2.00	1.71	2.19	2.54	2.38	1.92	1.97
	LLM	**2.12**	1.65	**1.30**	**1.49**	**2.69**	1.96	1.72	**2.09**	**2.27**	**2.34**	**1.88**	**1.86**
	MLP	**2.13**	1.57	1.33	1.50	**2.69**	**1.83**	1.67	**2.06**	**2.28**	2.31	**1.89**	**1.77**
	RBF	**2.10**	**1.59**	1.32	**1.45**	**2.67**	**1.84**	1.67	**2.09**	**2.26**	**2.29**	**1.87**	**1.74**
Oxford	**Clim.**	3.46	2.86	2.46	2.83	3.78	3.14	2.45	3.44	3.66	3.37	3.40	2.93
	MLR	1.54	1.09	1.16	1.23	2.29	1.94	1.61	2.20	1.68	1.61	1.83	1.37
	LLM	**1.44**	**1.07**	1.15	**1.21**	**2.19**	**1.81**	1.61	**2.11**	**1.55**	**1.56**	**1.81**	1.33
	MLP	**1.44**	1.06	**1.15**	**1.24**	**2.18**	**1.79**	**1.57**	2.13	1.60	**1.59**	1.79	1.35
	RBF	**1.43**	1.08	1.15	1.22	**2.16**	**1.78**	**1.58**	**2.13**	**1.53**	**1.55**	**1.79**	1.31
Bordeaux	**Clim.**	3.75	3.13	2.97	3.25	4.25	3.16	2.72	3.63	3.98	4.25	4.15	3.86
	MLR	1.75	1.47	1.45	1.50	2.50	2.38	2.18	2.37	2.26	2.05	1.87	1.87
	LLM	**1.69**	**1.42**	1.45	**1.44**	**2.40**	**2.22**	**2.11**	**2.21**	**2.18**	**2.00**	1.88	1.85
	MLP	1.72	1.44	1.40	1.45	**2.39**	**2.25**	**2.07**	**2.19**	2.19	**1.97**	1.85	1.80
	RBF	**1.66**	**1.43**	1.40	1.45	**2.38**	**2.26**	**2.07**	**2.22**	**2.16**	**2.01**	1.86	1.81

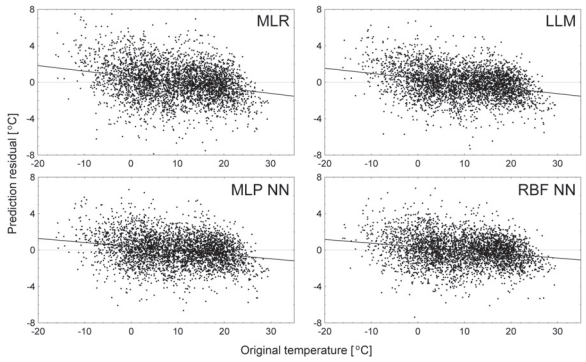


Fig. 7. Plot of the prediction residuals (the predicted temperature minus the original observed temperature) against the original temperature for 1-day ahead prediction of NCEP/NCAR T_{1000} series at 50° N, 15° E (whole year). Bold line represents the linear fit of the residuals.

decreasing the variance of the series of predictions compared to the original one. This kind of behavior was observed for all seasons, as well as the year as a whole, and it is demonstrated for the grid point 50° N, 15° E in Fig. 7. The problem of a deformed distribution of values is commonly encountered in the context of the application of empirical models in climate research, particularly in statistical downscaling, and various strategies have been proposed for handling it, such as variance inflation or partial randomization (e.g. von Storch, 1999). All results presented here are direct outputs of the transfer functions, without being subject to any form of additional postprocessing.

4.2 Downscaling

Statistical downscaling of large-scale data is another common task of meteorological time series analysis, and one in which nonlinear methods are sometimes used. Several studies have been published devoted to downscaling or postprocessing of temperatures by nonlinear methods, mostly MLPs (Trigo and Palutikof, 1999; Schoof and Pryor, 2001; Marzban, 2003; Casaioli et al., 2003) or neural networks based on RBF functions (Weichert and Bürger, 1998). Here, downscaling of the gridded large-scale data was done for the predictand series of daily mean, minimum and maximum temperatures. NCEP/NCAR reanalysis series were used as predictors. The pattern of predictors (Fig. 1b) was centered on the NCEP/NCAR grid point closest to the respective station (see the last column of Table 1). Years 1961–1990 were used as the training set, and years 1991–2000 as the testing set. In some cases, the testing set was shorter than 10 years

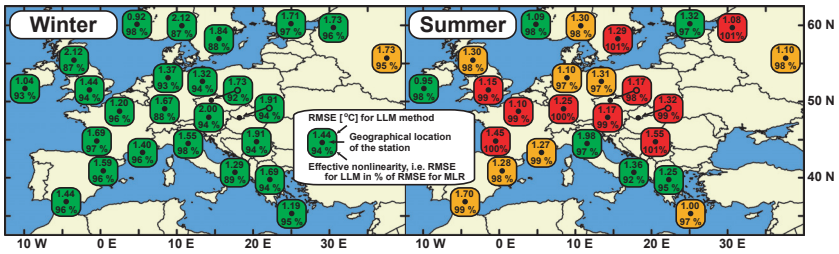


Fig. 8. Results of NCEP/NCAR reanalysis downscaling by LLM and MLR methods for 25 European stations – daily mean temperature. Color of the station's background shows the result of the test of the hypothesis that the absolute values of daily errors were the same for both methods, against the alternative that errors were smaller for LLM – red means that the hypothesis was not rejected at the level of confidence of 95%, yellow that it was rejected at 95%, but not at 99%, and green means that the difference or errors was significant at the 99% level of confidence (one-sided paired Wilcoxon test was used).

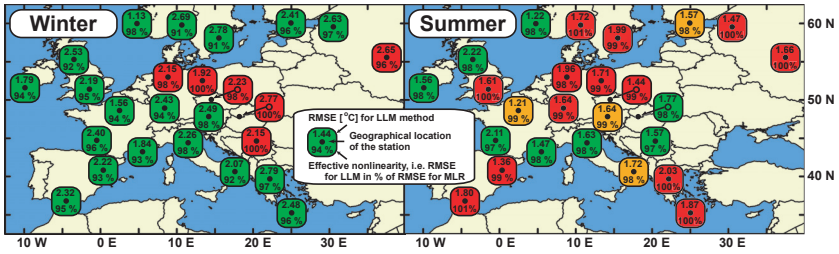


Fig. 9. Same as Fig. 8, for daily minimum temperature.

because the predictand temperature series was not available for the whole interval (such stations are marked with an asterisk in Table 1). Moreover, some of the series contained missing values – in those cases, the respective days were excluded from the computation. The number of missing values was, however, always very small in comparison with the total size of the data set, so their presence should not have caused any major shift in the results.

For all 25 stations from Table 1, results of downscaling by MLR and LLM methods for daily mean (Fig. 8), minimum (Fig. 9) and maximum (Fig. 10) temperature in winter and in summer are presented. Aside from computing the relative differences between RMSE for the LLM and MLR methods, we also tested the statistical significance of the difference of the medians of the absolute values of daily errors for the LLM and MLR methods. The testing was done by the one-sided paired Wilcoxon test, and its outcomes are represented by different colors of the respective station's background in Figs. 8 to 10. Full results for the four stations are shown in Table 3. Situations when nonlinear method gave better results than MLR at the 95% level of confidence are indicated by bold print. To compare the results to outcomes of a low-skill method, the table also contains RMSE obtained by means of mean climatology (i.e. when monthly mean values of the predictand were used in the role of $x^{PRED}(t)$ from Eq. (3)).

Nonlinear methods generally performed better than MLR, but there were distinct differences between the seasons – most profoundly nonlinear behavior was typical for winter, while in summer, nonlinear techniques seemed to grant just a very small, if any, improvement in the comparison with MLR at most stations. This can be easily seen from Fig. 11, where histograms of the absolute values of downscaling errors are shown for mean daily temperature from Oslo (a station, where the contrast between winter and summer was particularly clear). A tendency was observed in all methods to underestimate high temperatures and overestimate the low ones, as in the case of the forecasts in Sect. 4.1.

Situations when one of the methods was superior to all the others, according to the Wilcoxon test, were rather rare, and they are marked with an underline in Table 3. None of the nonlinear methods could be identified as the best one in all cases, or the majority of the cases. However, when one of the methods was tested as the best performing one, it was usually either the method of local models or the RBF neural network.

Unlike the forecasts studied in Sect. 4.1, results of downscaling were influenced by local conditions of individual stations and by eventual problems with series quality (shorter length of some series, as well as the fact that the series in ECA&D were collected from many different sources, and they may not be mutually as easily comparable as NCEP/NCAR reanalysis data). Thus, we have not tried to draw any detailed conclusions about the geographical structure of errors or nonlinearity based on the behavior of individual stations. A few points can be made, nonetheless:

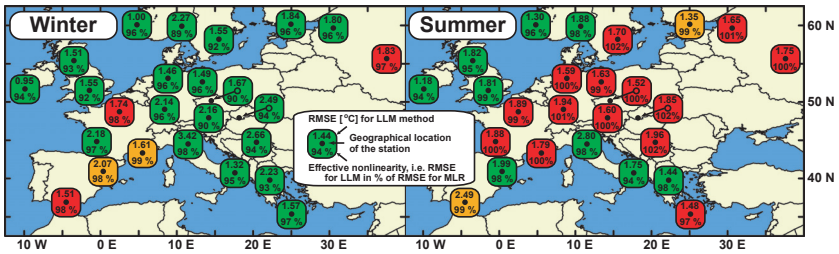


Fig. 10. Same as Fig. 8, for daily maximum temperature.

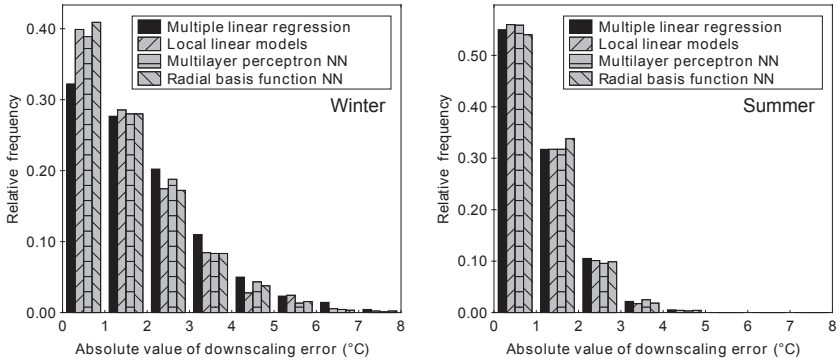


Fig. 11. Histogram of the absolute values of downscaling errors for mean daily temperature from Oslo, Norway, in winter (left) and summer (right).

– Values of RMSE were higher in winter than in summer at most stations, for mean and minimum temperatures; for maximum temperature, almost half of the stations showed higher values of error in summer. The effective nonlinearity was typically stronger in winter than in summer for all three types of temperature. In summer, nonlinear methods even performed a little worse than MLR sometimes, and for many stations, the difference in the medians of errors was not conclusive at the confidence level of 95%.

– Errors were lower for daily mean temperatures than for maximum or minimum ones in most locations. This is not surprising as mean values are already averaged in time, and therefore smoother. The extreme temperatures represent values in a single moment during the day, hence they are more prone to be affected with a noise component of the atmospheric dynamics, and more difficult to determine.

– In Sect. 4.1, we pointed out that nonlinearity in the Mediterranean area seems to be significantly weaker than for the rest of Europe. Here, a similar rule does not apply and there is no clear connection between the geographic position of the station and the degree of exhibited nonlinearity, as even geographically close stations exhibited quite a different behavior in a number of cases.

5 Discussion

The presented analyses, carried out for different tasks and various series, do not identify any of the tested nonlinear methods to be definitely superior to the others. It is, however, obvious that in most cases the LLM, MLP NN and RBF NN methods outperformed multiple linear regression. Application of nonlinear methods seems, therefore, to be useful for meteorological time series analysis at synoptic time scales, but the issue remains as to which one of them should be used preferably. Aside from the performance itself, attention should also be paid to the application properties of the contestants. Advantages and drawbacks of individual methods may be summarized as follows:

– **Method of local linear models'** advantage is that its use need not be preceded by the potentially time-consuming training of weights. On the other hand, the execution itself takes longer than for the already trained NNs, due to the search for nearest neighbors, which needs to be carried out for every processed state. This method is faster than both types of NNs when applied just once for a given setting, as we did in order to obtain the results for Figs. 4 and 5. Its speed also makes estimation of suitable parameters (specifically, N) by a trial-and-error technique easier.

– **MLP**s are the most frequently used nonlinear technique of the tested ones, which is an advantage by itself, due

to support in many existing software applications. Their features are well documented and thoroughly discussed by a number of authors (among others Haykin, 1999; Principe et al., 2000). Perhaps the most severe drawback of MLPs is the possibility that the learning procedure will end in a local minimum of the error function instead of reaching the global one. The seriousness of this problem can be reduced by means of repeated training or by application of a global minimum search technique (such as the one used by Casaioli et al., 2003), but usually at the cost of increased training time. There is also the possibility of overtraining, when the learning procedure runs for too long a time, and the network becomes overoptimized for the description of the cases from the learning set, thus losing its ability to generalize. None of these problems is unsolvable, but they cause the application of MLPs to be potentially tricky, with many nontrivial decisions to be made and the learning process which needs to be supervised.

– **RBF NN**s do not suffer from the "local minimum trap" problem (at least not their form that we used here), because the supervised optimization of the weights to the output layer can be treated as a linear problem. Certain variations of the results can arise from the random selection of the RBF centers, but these were very small in our case. There is also no need to worry about the network becoming overtrained. Nonetheless, it is still possible to overfit the network in the sense of too many free parameters (when too many hidden neurons are used). Training of RBF NNs is usually faster than that of MLPs (especially considering the need for repeated training of MLPs).

Therefore, the RBF NNs and LLM method can offer better application properties than MLPs in many respects. Considering their relatively easy implementation, we believe that these methods can be recommended as a worthy alternative to MLPs.

The error and nonlinearity maps in Figs. 4, 5 and 8 to 10 offer some insight into the geographic structure of spatiotemporal relations between the atmospheric variables, but they should be interpreted carefully. For prediction (Figs. 4 and 5), the fact that the reanalysis and not measured data was used may be partly responsible for the character of the observed structures. There is also an issue of selection of the input variables – the presented results were obtained for one specific pattern of predictors, and it can be argued that the maps could change if a different one was used. We repeated the tests for various sets of predictors, differing in both type and number of variables used, and although the details of the maps changed, the basic geographic structure of nonlinearity seemed quite robust and unaffected by the modification of the computation settings. There still was an area of distinct nonlinearity over continental Europe, more apparent in winter than in summer, while over the Mediterranean and Northern Africa, nonlinearity was weaker. These results suggest that the nonlinear character of the climate system is strongly reflected in the atmospheric time series in the midlatitudes, and that the relations between them cannot be fully described by linear mappings. In the tropical areas south of Europe, a purely linear description seems more sufficient.

Interpretation of the downscaling tests needs to be done with respect to their possible dependence on the tested stations' local conditions. The observed profound differences between even close stations suggest that conclusions about the skill and suitability of time series analysis methods, obtained for a limited area or a single station, cannot be automatically generalized and applied to different locations. Noteworthy is the seasonal variance of the results, which implies the presence of a strong detectable nonlinear component in the relations between predictors and predictand in winter, and a rather linear character of these relations in summer.

It is also interesting that the effective nonlinearity is generally weaker for downscaling than for forecasting tasks (compare results in Figs. 5 and 8). This does not necessarily mean that the relations between predictors and predictand are intrinsically more linear in the case of downscaling. The reason for this difference may lie in the fact that while the forecasting tests were done solely on reanalysis data, which are relatively smooth by construction, downscaling predictands are local measurements. They reflect the situation in just a very small area, so it is possible for them to contain a stronger high-dimensional component, which may be nonlinear in its nature, but has a character of noise and is too complex to be described by any time series analysis method, even a nonlinear one.

6 Conclusions

Our primary intention was to make a comparison of three nonlinear methods and to identify the one most suitable for the analysis of climatic data at synoptic time scales. It turned out that the most commonly applied nonlinear method, multilayer perceptron neural network, may not necessarily be the best possible choice. The other two tested nonlinear techniques, method of local linear models and radial basis function neural network, performed equally well, even better in many cases, they do not suffer from MLPs' drawbacks, such as the problem of local minima of the error function or the danger of overtraining, and their implementation is relatively easy. We can therefore recommend them as a worthy alternative to MLPs.

The constructed maps of the geographical structure of errors and effective nonlinearity revealed profound spatial variations of the results. In case of prediction, a distinct contrast between the temperate zone (stronger nonlinearity) and the subtropical and tropical zones (weaker nonlinearity) was found, especially in summer. As for downscaling tests, we were not able to identify any apparent large-scale geographical distribution of nonlinearity, probably because the results for individual stations were influenced by the local conditions at least as much as by the large-scale climate dynamics.

For both forecasting and downscaling, profound seasonal dependence of the results was observed, with errors being larger and nonlinearity being stronger in winter than in summer in most locations.

Acknowledgements. This work was supported by the Grant Agency of Charles University, grant 227/2002/B-GEO/MFF, Ministry of the Environment of the Czech Republic, grant VaV/740/2/03 and Research project 113200004. We would also like to express our gratitude to B. Sivakumar for his valuable comments on the manuscript, and to thank the two anonymous reviewers of the previous version of the manuscript.

Edited by: A. Tsonis
Reviewed by: B. Sivakumar

References

Abarbanel, H. D. I.: Analysis of Observed Chaotic Data. Springer, New York, 272 pp., 1996.

Casaioli, M., Mantovani, R., Scorzoni, F. P., Puca, S., Speranza, A., and Tirozzi, B.: Linear and nonlinear post-processing of numerically forecasted surface temperature, Nonlin. Processes Geophys., 10, 373–383, 2003,
SRef-ID: 1607-7946/npg/2003-10-373.

Essenwanger, O. M.: General Climatology: Classification of Climates (World Survey of Climatology, Vol. 1 Part C), Elsevier, Amsterdam, 113 pp., 2001.

Farmer, J. D. and Sidorowich, J. J.: Predicting Chaotic Time Series, Phys. Rev. Lett., 59, 845–847, 1987, reprinted in Ott et al. (1994).

Fraedrich, K.: Estimating the Dimensions of Weather and Climate Attractors, J. Atmos. Sci., 43, 419–432, 1986.

Galka, A.: Topics in Nonlinear Time Series Analysis – With Implications for EEG Analysis, World Scientific Publishing, Singapore, 342 pp., 2000.

Hall, T., Brooks, H. E., and Doswell, C. A.: Precipitation Forecasting Using a Neural Network, Weather Forecast., 14, 338–345, 1999.

Haykin, S.: Neural networks: A comprehensive foundation (2nd edition), Prentice Hall, Upper Saddle River, New Jersey, 842 pp., 1999.

Hegger, R., Kantz, H., and Schreiber, T.: Practical implementation of nonlinear time series methods: The TISEAN package, CHAOS, 9, 413–435, 1999.

Huth, R.: Statistical Downscaling of Daily Temperature in Central Europe, J. Climate, 15, 1731–1742, 2002.

Huth, R.: Sensitivity of Local Daily Temperature Change Estimates to the Selection of Downscaling Models and Predictors, J. Climate, 17, 640–652, 2004.

Jayawardena, A. W. and Gurung, A. B.: Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach, J. Hydrol., 228, 242–264, 2000.

Jayawardena, A. W., Li, W. K., and Xu, P.: Neighbourhood selection for local modelling and prediction of hydrological time series, J. Hydrol., 258, 40–57, 2002.

Kantz, H. and Schreiber, T.: Nonlinear time series analysis. Cambridge University Press, Cambridge, 304 pp., 1997.

Keppenne, C. L. and Nicolis, C.: Global Properties and Local Structure of the Weather Attractor over Western Europe, J. Atmos. Sci., 46, 2356–2370, 1989.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP–NCAR 50–Year Reanalysis: Monthly Means CD-ROM and Documentation, B. Am. Meteorol. Soc., 82, 247–268, 2001.

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., et al.: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, Int. J. Climatol., 22, 1441–1453, 2002.

MacQueen, J.: Some methods for classification and analysis of multivariate observations, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, 281–297, 1967.

Marzban, C.: A Neural Network for Post-processing Model Output: ARPS, Mon. Weather Rev., 131, 1103–1111, 2003.

Miksovsky, J. and Raidl, A.: Testing for nonlinearity in European climatic time series by the method of surrogate data, Theor. Appl. Climatol., in press, 2005.

Ott, E., Sauer, T., and Yorke, J. A. (eds.): Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems, John Wiley & Sons, New York, 418 pp., 1994.

Packard, H. N., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S.: Geometry from a time series, Phys. Rev. Lett., 45, 712–716, 1980, reprinted in Ott et al. (1994).

Palus, M.: Detecting nonlinearity in multivariate time series, Phys. Lett. A, 213, 138–147, 1996.

Palus, M. and Novotna, D.: Testing for nonlinearity in weather records, Phys. Lett. A, 193, 67–74, 1994.

Pérez-Muñuzuri, V. and Gelpi, I. R.: Application of nonlinear forecasting techniques for meteorological modeling, Ann. Geophys., 18, 1349–1359, 2000,
SRef-ID: 1432-0576/ag/2000-18-1349.

Principe, J. C., Euliano, N. R., and Lefebvre, W. C.: Neural and Adaptive systems: Fundamentals through Simulations, John Wiley & Sons, New York, 656 pp., 2000.

Sauer, T.: Time Series Prediction by Using Delay Coordinate Embedding., in: Time Series Prediction: Forecasting the Future and Understanding the Past, Santa Fe Institute Studies in the Science of Complexity XV, edited by: Weigend, A. S. and Gershenfeld, N. A., Addison-Wesley, Reading, 643 pp., 1993, reprinted in Ott et al. (1994).

Schoof, J. T. and Pryor, S. C.: Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, Int. J. Climatol., 21, 773–790, 2001.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D, 142, 346–382, 2000.

Shao, J.: Improving Nowcasts of Road Surface Temperature by a Backpropagation Neural Network, Weather Forecast., 13, 164–171, 1998.

Sivakumar, B.: Chaos theory in geophysics: past, present and future, Chaos Soliton. Fract., 19, 441–462, 2004.

Sivakumar, B., Berndtsson, R., Olsson, J., Jinno, K., and Kawamura, A.: Dynamics of monthly rainfall-runoff process at the Göta basin: A search for chaos, Hydrol. Earth Syst. Sci., 4, 407–417, 2000,
SRef-ID: 1607-7938/hess/2000-4-407.

Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature, 344, 734–740, 1990, reprinted in Ott et al. (1994).

Takens, F.: Detecting Strange Attractors in Turbulence, in: Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Vol. 898, edited by: Rand, D. A and Zouny, L. S., Springer, New

York, 366–381, 1981.

Tang, B., Hsieh, W. W., Monahan, A. H., and Tangang, F. T.: Skill Comparisons between Neural Networks and Canonical Correlation Analysis in Predicting the Equatorial Pacific Sea Surface Temperatures, J. Climate, 13, 287–293, 2000.

Trigo, R. M. and Palutikof, J. P.: Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach, Climate Res., 13, 45–59, 1999.

Tsonis, A. A.: Probing the linearity and nonlinearity in the transitions of the atmospheric circulation, Nonlin. Processes Geophys., 8, 341–345, 2001,
SRef-ID: 1607-7946/npg/2001-8-341

von Storch, H.: On the Use of "Inflation" in Statistical Downscaling, J. Climate, 12, 3505–3506, 1999.

Waelbroeck, H., López-Peña, R., Morales, T., and Zertuche, F.: Prediction of Tropical Rainfall by Local Phase Space Reconstruction, J. Atmos. Sci., 51, 3360–3364, 1994.

Weichert, A. and Bürger, G.: Linear versus nonlinear techniques in downscaling, Climate Res., 10, 83–93, 1998.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic Press, San Diego, 467 pp., 1995.

APPENDIX IV

HUTH, R., J. MIKŠOVSKÝ, P. ŠTĚPÁNEK, M. BELDA, A. FARDA, Z. CHLÁDOVÁ, AND P. PIŠOFT (2015), Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature, *Theoretical and Applied Climatology*, *120*(3-4), 533-553, doi:10.1007/s00704-014-1190-3.

Theor Appl Climatol (2015) 120:533–553
DOI 10.1007/s00704-014-1190-3

ORIGINAL PAPER

Comparative validation of statistical and dynamical downscaling models on a dense grid in central Europe: temperature

Radan Huth · Jiří Mikšovský · Petr Štěpánek · Michal Belda · Aleš Farda · Zuzana Chládová · Petr Pišoft

Received: 11 October 2013 / Accepted: 29 May 2014 / Published online: 13 June 2014

Abstract Minimum and maximum temperature in two regional climate models and five statistical downscaling models are validated according to a unified set of criteria that have a potential relevance for impact assessments: persistence (temporal autocorrelations), spatial autocorrelations, extreme quantiles, skewness, kurtosis, and the degree of fit to observed data on both short and long times scales. The validation is conducted on two dense grids in central Europe as follows: (1) a station network and (2) a grid with a resolution of 10 km. The gridded dataset is not contaminated by artifacts of the interpolation procedure; therefore, we claim that using a gridded dataset as a validation base is a valid approach. The fit to observations in short time scales is equally good for the statistical downscaling (SDS) models and regional climate models (RCMs) in winter, while it is much better for the SDS models in summer. The reproduction of variability on long time scales, expressed as linear trends, is similarly successful by both SDS models and RCMs. Results for other criteria suggest that there is no justification for preferring dynamical models at the expense of statistical models—and vice versa. The non-linear SDS models do not outperform the linear ones.

R. Huth
Department of Physical Geography and Geoecology, Faculty of Science, Charles University, Albertov 6, 128 43 Praha 2, Prague, Czech Republic

R. Huth (✉) · Z. Chládová
Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
e-mail: huth@ufa.cas.cz

R. Huth · P. Štěpánek · A. Farda
Czech Globe–Global Change Research Centre, Academy of Sciences of the Czech Republic, Brno, Czech Republic

J. Mikšovský · M. Belda · P. Pišoft
Department of Meteorology and Environment Protection, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

P. Štěpánek
Czech Hydrometeorological Institute, Regional Office, Brno, Czech Republic

A. Farda
Czech Hydrometeorological Institute, Prague, Czech Republic

1 Introduction

Global climate models (GCMs) are designed to describe climatic features at large spatial scales and are usually run in effective resolutions exceeding a hundred of kilometres. For these reasons, they are not capable of directly providing regional or even local information. However, it is the regional and local scales that are needed in climate change impact studies. The scale mismatch between what GCMs are able to provide and what is required in the impact studies can be bridged in several ways. Two of the approaches, which have been used most commonly, have undergone a rapid development during the last decade: regional climate models (RCMs), also referred to as dynamical downscaling, and statistical downscaling (SDS).

We use the term 'statistical downscaling' consistently with Benestad et al. (2008). This means that SDS in this paper refers to the 'Perfect Prog' approach (i.e. the approach based on relationships between observed predictors and observed predictands) according to the nomenclature used in Maraun et al. (2010) and does not include weather-generator approaches and bias-correction methods.

An extremely important task is a comparison of a climate simulated by a model with reality, that is, validation. The necessary, but not sufficient, condition for having confidence in simulated future climates is an evidence of the models'

ability to simulate present climate conditions. Both the RCMs and SDS have been validated extensively in terms of basic characteristics (mean, variability) of two most relevant climatic variables, temperature and precipitation. However, there are relatively few studies that have compared relative merits of RCMs and SDS, as already noted in a review paper by Fowler et al. (2007). The majority of the comparison studies (Kidson and Thompson 1998; Murphy 1999; Wilby et al. 2000; Hellström et al. 2001; Busuioc et al. 2006; Haylock et al. 2006; Schmidli et al. 2007; Flaounas et al. 2013; Khalili et al. 2013) suggest that the performance of RCMs and SDS in reproducing daily and monthly temperature is comparable. In addition to it, validation studies of RCMs have, to a large extent, been driven by demands of climate modellers; therefore, validation studies frequently concentrate on variables relevant for a further development of RCMs themselves but not directly for climate impacts. On the other hand, compared with RCMs, SDS models have been subjected to a much more detailed scrutiny as to their ability to simulate more advanced characteristics, which are directly relevant for society and/or important for climate impact models, such as higher-order statistical moments, temporal and spatial autocorrelations, prolonged extreme events and periods, etc. (Kalvová and Nemešová 1998; Easterling 1999; Huth et al. 2000, 2001, 2003, 2008; Busuioc and von Storch 2003; Kettle and Thompson 2004; Bachner et al. 2008; Rauscher et al. 2010).

There are two broad families of SDS methods (Benestad et al. 2008) as follows: linear (represented by multiple linear regression and canonical correlation analysis) and non-linear (represented by artificial neural networks and also analog-based and classification methods). Several studies comparing their performance (e.g., Trigo and Palutikof 1999, 2001; Schoof and Pryor 2001; Mpelasoka et al. 2001; Coulibaly et al. 2005; Mikšovský and Raidl 2005; Eccel et al. 2007; Huth et al. 2008) do not provide a clear answer as to which family is superior; hence, the relative performance of linear and non-linear methods is another unresolved issue in the context of the present study.

SDS models have typically been evaluated according to their ability to approximate the observed time series. The ability to reproduce short-term day-to-day time variations has, however, little to do with the ability to reproduce variations on long-term decadal time scales, on which the climate change signal proceeds. In spite of it, only few SDS models and RCMs have been subjected to a validation of trends or reproduction of contrasting recent climate states (von Storch et al. 1993; Busuioc and von Storch 2003; Giorgi et al. 2004; Kettle and Thompson 2004; Benestad et al. 2007; Lorenz and Jacob 2010; Bukovsky 2012; Ceppi et al. 2012).

The brief review of recent literature and gaps in our knowledge implies that a validation study comparing RCMs with both linear and non-linear SDS models, going beyond means and standard deviations, that is, concentrating on advanced statistical characteristics and examining long-term trends, would constitute a step forward in understanding the downscaling methods. This is what the current study takes on as its goal: to compare the ability to simulate the advanced characteristics of climate variables, potentially relevant in various climate change impact studies, between the dynamical and statistical downscaling models, that is, to subject them to the same validation procedure. Fowler et al. (2007) pose several recommendations for regional climate change research; we follow at least two of them: a call for a 'coordinated intercomparison and diagnostics of models (…) (RCMs, SDS)' and a wider use of 'higher-order statistics … to measure downscaling skill.'

Our analysis concentrates on advanced validation criteria. We intentionally do not present validation of the first two statistical moments, mean and variance, for two main reasons. First, the first two moments are in the focus of the vast majority of validation studies, so their treatment here would not be innovative. And second, statistical post-processing approaches to correct the model outputs towards observed mean and variance have recently become available, such as quantile mapping (Déqué 2007; Themeßl et al. 2012). Moreover, there are approaches to correct the mean and variance, which are directly embedded in the statistical downscaling framework, such as variance inflation and noise (both white and temporally or spatially autocorrelated) addition (e.g., von Storch 1999; Huth 2002). The criteria we employ include temporal and spatial autocorrelations, third and fourth statistical moments (skewness and kurtosis), and extreme quantiles (5th and 95th) of statistical distributions. We take advantage of the fact that both the dynamical and statistical models are driven by reanalysis data in this study, which means that time series they produce can be directly compared with observations. Therefore, we include into the set of validation characteristics also correlations with the observed time series as a measure of short-term fit to observations, and linear trends as a measure of changes on longer time scales on which climate change proceeds. The analysis is carried out on a daily basis and covers maximum and minimum temperature.

In validation, the outputs of a model are compared with observations. Whereas the observations are located on an irregular network of stations, the dynamical model (GCM, RCM) outputs are provided on a regular grid. (Indeed, SDS procedures can yield output either at stations or grid points, depending on what they were trained on.) Dynamical models thus provide area-aggregated rather than point-specific data, which makes a direct comparison between station data and gridded model output less straightforward, especially for variables with a short correlation distance, such as precipitation (e.g., Skelly and Henderson-Sellers 1996). Therefore, the validation can be made potentially fairer to dynamical models if the observations are transformed from stations to a grid. The

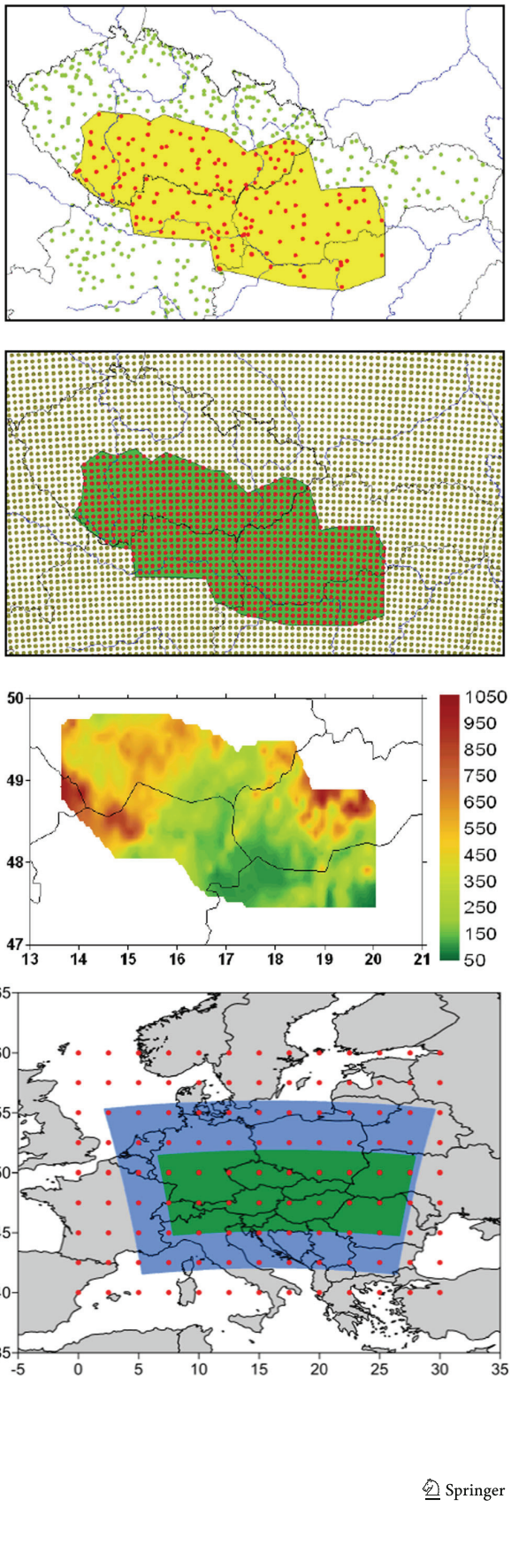


Fig. 1 *Top*: analysis domain shown in yellow; *dots* indicate the available climatological stations (*green*/*red* for stations inside/outside the domain). *Second* from *top*: 10-km grid; *red*/*green dots* indicate gridpoints inside/outside the domain (denoted as a *green area*). *Third* from *top*: elevation of terrain within the analysis domain (m above sea level). *Bottom*: integration domains of RegCM (*blue*) and ALADIN (*green*), and the source area of potential predictors for the SDS techniques with gridpoints is indicated by *red dots*

transformation procedure must of course take into account characteristics specific for each variable; see e.g. Osborn and Hulme (1998) for precipitation.

This is the approach we take in the current study. We use the same grid for analyzing observed data, outputs from two RCMs, and running and evaluation of five SDS models. This makes a comparison of validation results among individual models as fair as possible. To get more faith that gridded observations are not contaminated by artifacts from the interpolation procedure, we analyze the station data and SDS models developed on the station data as well and compare results based on the two datasets.

2 Datasets and methods

The region of interest covers a part of central Europe along the joint borders of the Czech Republic, Austria, Slovakia, and Hungary (Fig. 1). This particular region was selected mainly because of joint efforts to conduct coordinated climate change impact studies by the all four countries within its area. The variables examined are daily maximum and minimum temperature. We intentionally do not deal with daily mean temperature since the formulas for its calculation differ between countries, and in some countries, they even changed during the study period. All the temperature data were interpolated onto the regular grid on which the ALADIN-Climate/CZ RCM was integrated, which has a 10-km resolution. Altogether, 832 grid points cover the domain (Fig. 1).

Our task is to find out how well the downscaling models reproduce fine-scale structures from large-scale information. Therefore, the downscaling models, both dynamical and statistical, were driven by ERA-40 reanalysis (Uppala et al. 2005). That is, RCMs were nested into ERA-40 data in both cases using a double nesting in order to overcome problems of a resolution jump, with the grid step of the intermediate model being 25 km. The SDS models were developed and trained on ERA-40 data. The RCMs were integrated for the period 1961–2000; the SDS models were trained on the 30-year period 1961–1990, while the period 1991–2000 was used as an independent sample. All the validation is conducted on the latter period. The downscaling models are described in more detail below in Secs. 2.2 and 2.3 and are summarized in Table 1.

Table 1 List of downscaling models

Abbreviation	Type	Brief description	Location in text
ALA	RCM	ALADIN-Climate	Sec. 2.2.1
REG	RCM	RegCM3	Sec. 2.2.2
MLR	SDS	Multiple linear regression	Sec. 2.3.1
LLM	SDS	Locally linear model	Sec. 2.3.2
LCM	SDS	Locally constant model (analog model)	Sec. 2.3.2
RBF	SDS	Radial basis function-based neural network	Sec. 2.3.3
MLP	SDS	Multilayer perceptron neural network	Sec. 2.3.3

2.1 Observed data

The original observed data were taken from 165 stations in all the four participating countries. The station data were quality checked thoroughly and then subjected to the interpolation procedure. Here, only a brief description of the interpolation procedure is provided; for more details, refer to Štěpánek et al. (2011).

Temperature series at a particular grid point were calculated from up to six neighbouring (nearest) stations within the distance of 300 km at maximum, with the allowed maximum difference in altitude of 500 m. Before applying inverse distance weighting, data at the neighbouring stations were standardized relative to the altitude of the base grid point. The altitudes applied in the calculation of grid point series were taken from the model of terrain with a resolution of 1 km. The standardization was carried out by means of linear regression, taking into account the dependence of values of a particular meteorological element on altitude, individually for each station and each day. Each standardized value was checked if it did not differ excessively from the original value. For the weighted average (using inverse distances as weights), the power of weights equal to 1 was applied. In the case of temperature, standardized neighbour values outside the 20 to 80 % percentile range were not considered in the calculation of final values (i.e. a trimmed mean was applied). A similar procedure was applied to complete and adjust the station series.

2.2 Regional climate models

2.2.1 ALADIN-climate/CZ

The ALADIN-climate/CZ RCM was developed from numerical weather prediction model ALADIN, which has been used at the Czech Hydrometeorological Institute, in the version run operationally in 2003 and 2004. Compared to its older versions, an improved radiation parameterization, developed on the basis of the original by Ritter and Geleyn (1992), was incorporated. Surface processes are controlled by the ISBA scheme (Noilhan and Mahfouf 1996), and deep convection follows the formalism described by Bougeault (1985). Other improvements included improved semi-Lagrangian horizontal diffusion scheme and enhancements in vertical diffusion/turbulence processes. For a more detailed description, refer to Farda et al. (2010). The integration area consists of almost 11,000 points (148 in each of 74 latitude rows, the projection centre being approximately at 48.25° N, 17° E, which almost exactly corresponds to the centre of the analysis domain) organized in a grid with a constant horizontal resolution of 10 km. For the integration domain, see Fig. 1d. The model in this setup uses 43 atmospheric levels in the vertical.

2.2.2 RegCM3

RegCM is one of the RCMs commonly used for climate simulations in the region of central Europe; for its description, see Halenka et al. (2006) and the references therein. RegCM3 was implemented with a horizontal resolution of 10 km, on the grid consisting of 184×164 grid points, with a central point located at 49.0° N, 15.8° E, and with 23 vertical levels. Its integration domain is shown in Fig. 1d. For this application, the RegCM3 simulations were performed using the following physical parameterizations: SUBEX non-convective precipitation scheme (Elguindi et al. 2007), Grell convective scheme (Grell 1993), NCAR CCM3 radiation scheme (Kiehl et al. 1996) and BATS surface model (Dickinson et al. 1993). Since the position of RegCM's grid was different from that of ALADIN, the RegCM outputs were interpolated onto the grid of the ALADIN model during post-processing, using inverse distance weighting, in order for all the data to be available on a common grid.

2.3 Statistical downscaling models

Five variants of statistical downscaling models were employed to represent transfer functions between the ERA-40 predictors and local minimum or maximum temperature measurements at the grid points or stations: multiple linear regression (MLR), two types of local models in the phase space (locally linear models, LLM, and locally constant models, LCM) and two architectures of artificial neural

networks (radial basis function-based networks, RBF, and multilayer perceptron, MLP).

2.3.1 Multiple linear regression

To estimate temperature T at the target site by multiple linear regression (MLR), a weighted average of the predictors $P_i, i=1,\ldots,M$ was computed as follows:

$$\widehat{T} = \alpha_0 + \sum_{i=1}^{M} \alpha_i P_i \tag{1}$$

Coefficients α_i were determined to minimize the sum of the squared errors $\widehat{T}-T$ over the training period of data. MLR is one of the most commonly used downscaling techniques, offering easy implementation and high computation speed. On the other hand, in its basic form, this method lacks skill to describe even simple non-linear relations between predictors and predictand.

2.3.2 Local models

It has been shown (e.g. contributions in Ott et al. 1994) that linear mappings can be used to realistically approximate dynamics of even strongly non-linear systems, providing that linear models are applied locally for just small sections of the phase space. Here, a separate linear mapping in the form of Eq. 1 was created for each day in the validation part of data. The respective coefficients α_i were computed from a set of N_L days from the calibration period when the state of the regional weather system most resembled the one for the processed day. The similarity of individual states was measured by the Euclidean distance of the respective vectors of predictors, $\mathbf{x}=(P_1,P_2,\ldots,P_M)$. This method of locally linear models (LLM) represents a relatively straightforward generalization of linear regression, able to take potential non-linear components of the predictors-predictand relation into account. N_L was set to 500 days for all computations here (i.e. roughly 5 % of the size of the training period).

The second employed variant of the local model technique, the method of locally constant models (LCM), was based on an even simpler principle: the closest synoptic situation in the calibration segment of data was identified (again, using Euclidian distance of the vectors of predictors), and the corresponding value of the predictand (here, maximum or minimum temperature) was used as the estimate of downscaled temperature. In the atmospheric sciences, this approach is usually recognized as the method of analogues (e.g. Zorita and von Storch 1999).

2.3.3 Neural networks

Two architectures of artificial neural networks (NNs) were employed (their detailed description can be found, e.g., in Haykin 1999). The first one, the radial basis function (RBF) neural network was built around neurons whose activation decreases with the distance of the input vector of predictors x from reference input vectors $\mathbf{x}_n, n=1,\ldots,N_R$. The respective transfer function can be expressed as follows:

$$\widehat{T} = w_0 + \sum_{n=1}^{N_R} \left(w_n \exp\left(-\|\mathbf{x}-\mathbf{x}_n\|^2 / \left(2\sigma^2\right) \right) \right) \tag{2}$$

Weights w_n were optimized to minimize the error for the training part of the data by the least squares method. Vectors x_n were randomly sub-sampled from the training set. N_R was set to 200, widths of the radial functions were kept fixed at $\sigma^2=50$.

The second type of neural network, multilayer perceptron (MLP), was implemented in a form with a single hidden layer,

$$\widehat{T} = w_0 + \sum_{n=1}^{N_M} \left(w_n \tanh\left(v_{0n} + \sum_{m=1}^{M} v_{mn} P_m \right) \right) \tag{3}$$

where v_{mn} and w_n represent weights of connections between neurons in the input and hidden layer and in the hidden and output layer, respectively, and N_M denotes the number of neurons in the hidden layer. The training, initiated from random values of w_n and v_{mn}, was carried out by the error back propagation procedure run for a pre-determined number of 1,000 epochs. The optimum combination of the network's complexity (determined by the number of connections between neurons) and the number of iterations of the learning procedure was identified by tests carried out for a few trial settings. The networks in the applied configuration suffered from just a mild over learning or under learning, the sensitivity to the initialization of weights was small, and the presented results are therefore based on just a single realization of the MLP NN model.

2.3.4 Settings of SDS models

A crucial issue related to the application of all above-described forms of downscaling models is the selection of input variables. The number of potential predictors is large in the ERA-40 data, but the high correlation between series from nearby locations makes the actual information content substantially smaller. Only a small fraction of the available predictors was thus needed to enter the regression: typically about 10 to 20 predictors were enough to lower the out-of-sample RMSE close to its absolute minimum. Due to this fact, and to provide a better mutual comparability of results, all SDS

models were built with the same number of predictors, 15. A step-wise procedure based on the multiple linear regression, adding variables to minimize RMSE in each step, was applied to identify the suitable structure of input data. The predictors were drawn from ERA-40 reanalysis series of temperature at the 850-hPa level, geopotential height of the 500-hPa level, and relative humidity at the 700-hPa level, in the area bounded by 40° N, 60° N, 0° E, and 30° E, with a 2.5° step (Fig. 1d). The selection was performed individually for each combination of predictand (maximum or minimum temperature) and target location (i.e. grid point or weather station). While it would be ideal to carry out the selection of predictors separately for individual methods, the numerical demands of nonlinear mappings are too high for repeated computations, and the same predictor sets were therefore employed in all the five SDS methods. The mappings were constructed for the year as a whole, without separating individual seasons or months. All predictors were used in a normalized form.

2.4 Validation criteria

We subject the downscaling models to the evaluation according to the following three groups of criteria: their spatio-temporal structure, characteristics of their distributions, and the correspondence of their time series to observations.

The temporal structure is quantified by autocorrelations lagged by 1 day (we refer to it as 'persistence' for simplicity). The spatial structure is quantified by spatial correlations (hereafter referred to as 'autocorrelations'). In both cases, the calculations are conducted for anomalies from the mean annual cycle, and the Pearson correlation is used.

Statistical distributions of downscaled temperature variables are characterized by their skewness, defined as a standardized third moment, and kurtosis, defined as a normalized fourth moment. Both these characteristics are equal to zero for a Gaussian distribution. Skewness provides a measure of asymmetry, its negative (positive) value indicating a heavier left (right) tail of a distribution. Kurtosis is a measure of the peakedness of a distribution, its negative (positive) values suggesting a flatter (more peaked) maximum with light (heavy) tails. We also validate the 5th and 95th percentiles, corresponding to moderate extremes, which are potentially relevant in various impact sectors.

Furthermore, we employ two measures of correspondence between downscaled and observed temperatures. Correlations measure the fit of models to the reality, focusing on short time scales; Pearson correlation coefficient is used. On the other hand, linear trends were calculated in observed and downscaled values by regressing seasonal mean values against time. Their comparison provides information on how well the models describe a longer-term behaviour of temperature.

All criteria are evaluated for both maximum and minimum temperatures, for the year as a whole and for winter (DJF) and summer (JJA) separately and both against the gridded and station-observed dataset. The RCM outputs are defined on the gridded dataset only; we decided to avoid their interpolation onto the station network; therefore, both ALADIN and RegCM are evaluated against the gridded observations only.

3 Spatio-temporal structure

3.1 Persistence

Persistence of maximum temperature for the whole year is mapped for all downscaling models at both networks (gridded and station-based) in Fig. 2.

First, let us compare the two networks. Obviously, the gridded data yield more spatial details thanks to its higher spatial resolution, especially in the areas with complex terrain, such as Slovakian mountains and foothills in the northeast. Apart from the finest scale, there is a very good agreement between the gridded and station-observed dataset: lower values in the west, with several spots of persistence values below 0.72; highest values at the northeast and southeast corners, with an area of decreased persistence near Bratislava where borders of Slovakia, Austria, and Hungary connect. A similarly satisfying agreement between the gridded and station datasets can be seen for all the SDS models.

The MLR model overestimates persistence over almost all the domain. There is a tendency to a weaker overestimation by the LLM, RBF and MLP models; their patterns of persistence are very similar to each other, with a lack of west–east gradient seen in the observations. The LCM model is unable to capture the correct level of persistence, severely underestimating it. The ALADIN RCM generally underestimates persistence, with the lowest values located in the southwest, whereas RegCM seems to be closer to the observed values, with an underestimation mainly in Austria and central Slovakia and an overestimation in the northwestern part of the domain. In order to better document the degree of agreement between the models and observations, differences in persistence between the models and observations were calculated at all gridpoints, and their histograms were plotted (Fig. 3). Clearly, the MLR model overestimates persistence at almost all gridpoints, while the overestimation by the LLM model (and, very similarly, by the RBF and MLP models, not shown) is less ubiquitous, although still notable. The general underestimation of persistence by the ALADIN RCM and an ambiguous signal of the RegCM with a prevailing slight overestimation, accompanied by a much less frequent, though stronger underestimation, can also be seen. The comparison conducted for station datasets yields very similar results and is not shown here.

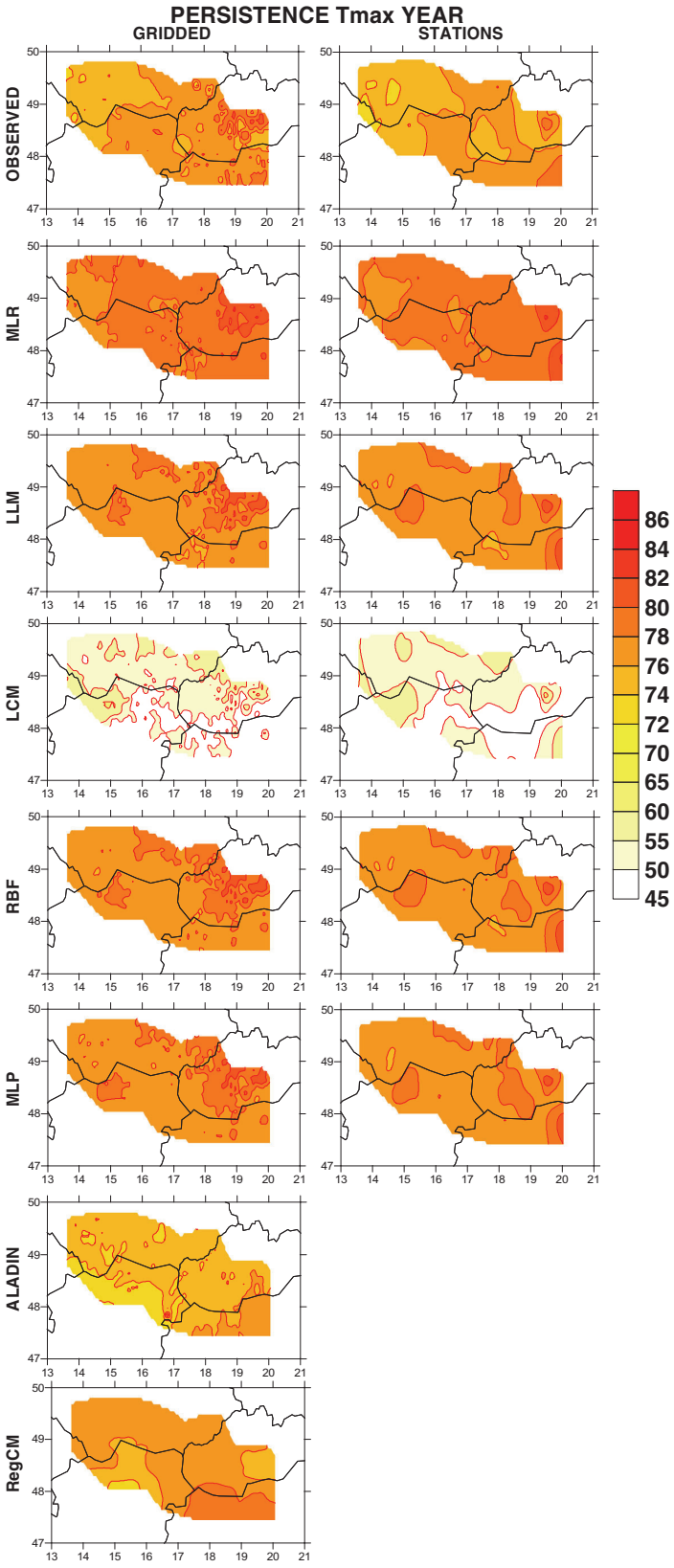


Fig. 2 Persistence (lag-1-day autocorrelations × 100) of maximum temperature for the gridded (*left*) and station (*right*) datasets, for observations (*top*) and individual downscaling models, all for the whole year

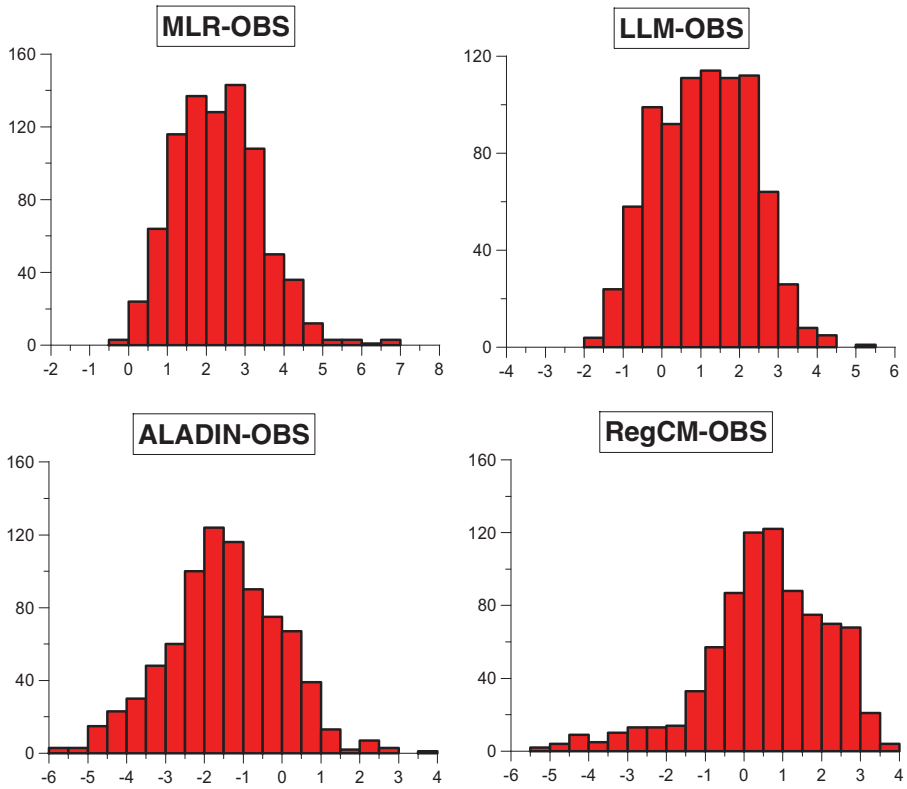


Fig. 3 Histograms (in absolute frequencies) of differences in persistence (×100) of maximum temperature, whole year, between the selected downscaling models and observations, for gridded datasets

Persistence values for all the methods are summarized in box-and-whisker plots in Fig. 4. We discuss results for the whole year first (Fig. 4 top). For observations, persistence is slightly lower in the station data; nevertheless, results for the observed and gridded data agree well with each other, with persistence for the station data tending to be slightly lower for all SDS methods. This can be seen as a supporting argument for considering both the datasets as equivalent, both representing the real climatic features. The LCM model clearly leaves the line: it severely underestimates the persistence. Other SDS models overestimate it, the MLR method exhibiting the highest persistence of all downscaling methods for maximum temperature, but lowest (together with MLP) for minimum temperature. Minimum temperature has lower persistence than maximum temperature; this feature is reproduced by all the models except for RegCM.

The seasonal analysis (Fig. 4 bottom) suggests that the general overestimation of persistence is mainly a reflection of summer conditions when all the models overestimate the persistence rather strongly, and more so for minimum temperature. The considerably lower persistence of minimum than maximum temperature in summer is reproduced by the SDS models, though with much smaller difference (except for the outlying LCM model), whereas both RCMs incorrectly simulate summer persistence of minimum temperature higher than that of maximum temperature. In winter, the persistence of maximum temperature is well simulated by MLR whereas being underestimated by all other models; on the other hand, the persistence of minimum temperature is underestimated by the ALADIN and MLR models while being well simulated by other models (aside from LCM again).

3.2 Spatial autocorrelations

Spatial autocorrelations were calculated for all pairs of gridpoints and all pairs of stations. Figure 5 brings an example of autocorrelations of maximum temperature with the northwestern-most point (49.78° N, 14.00° E) and station (Neumětely, 49.85° N, 14.04° E) for the whole year. The features displayed here are typical of autocorrelations with other base points or stations. Similarly to persistence, we can see more details in gridded data in the areas with complex terrain, except for RegCM. Apart from these fine structures, differences between the gridded and station datasets are only very minor again.

In observations, the autocorrelation drops gradually from the reference grid point/station to values of slightly over 0.70 at the opposite southeastern edge of the domain. The LCM model manifests a drastic underestimation; all other SDS models overestimate the autocorrelations, that is, the decline of autocorrelations with distance is slower than in

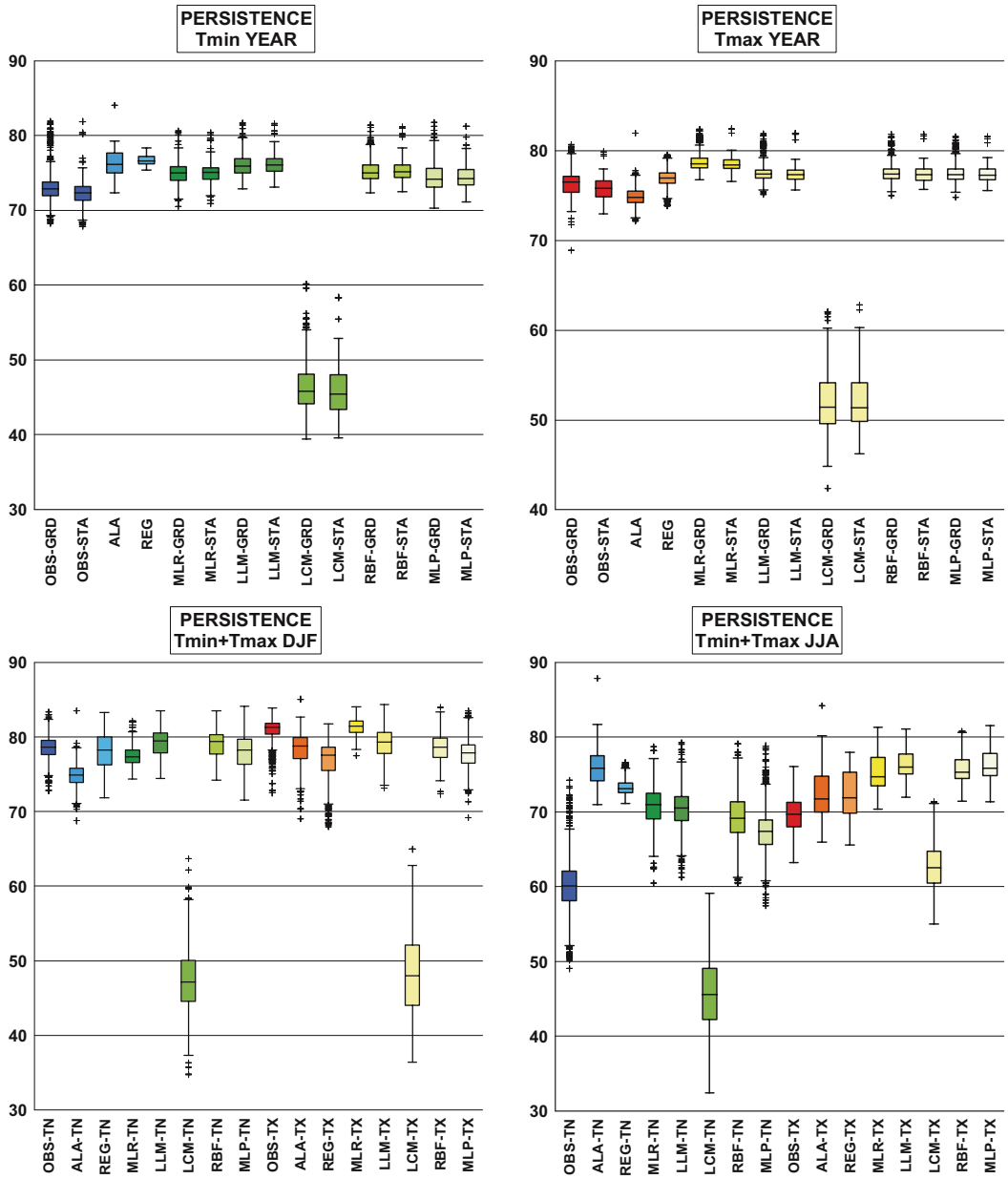


Fig. 4 *Box-and-whisker plots* for persistence (×100). *Top left*: whole year, minimum temperature; *top right*: whole year, maximum temperature; *bottom left*: DJF, both temperatures; *bottom right*: JJA, both temperatures. The *box ranges* from the lower to the upper quartile, the inner line being the median. The *whiskers* point to the minimum and maximum values, respectively, unless they are more than 1.5 times the inter quartile range lower (*higher*) than the lower (*upper*) quartile. In that case, the data points lying beyond these limits are denoted by small crosses. Gridded and station dataset is denoted as GRD and STA, respectively; minimum and maximum temperature is denoted as TX and TN, respectively

reality, the error being largest for MLR. Concerning the RCMs, RegCM exhibits a fairly realistic behaviour, while in ALADIN, the decrease of autocorrelation with distance is too fast.

In the following, we summarize autocorrelations for all pairs of gridpoints or stations by plotting the autocorrelation values against the distance between the sites. Such plots typically display diffuse clouds of points; therefore, in order to better visualize the dependence on distance, exponential fits to the data are added to the plots. First, the observed data are compared between the gridded and station dataset for minimum temperature and whole year (Fig. 6). Both fitted exponentials are fairly close to each other, with a difference of about 0.02, which is almost constant across the whole range of

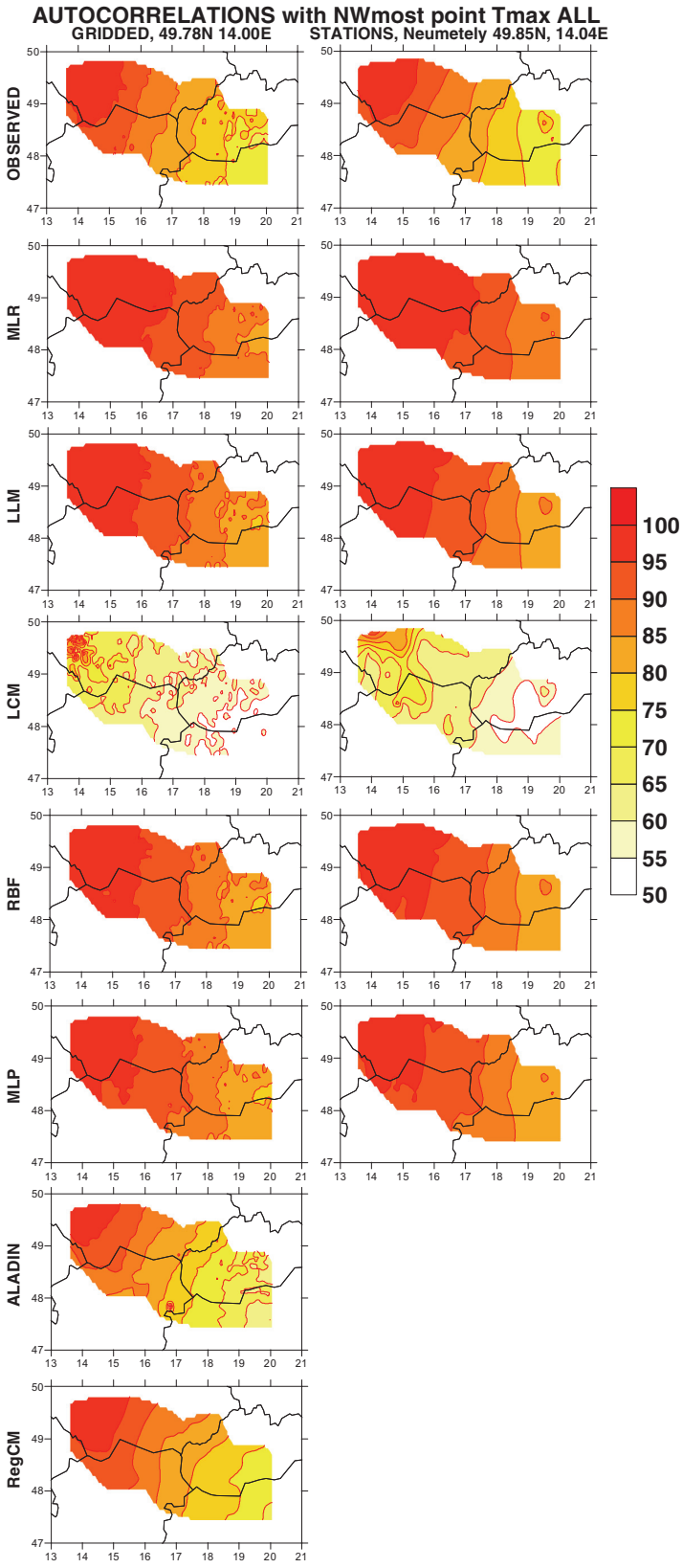


Fig. 5 Spatial autocorrelations (×100) with the grid point at 49.78° N, 14.00° E (*left column*) and the Neumětely station (49.85° N, 14.40° E, right column). Otherwise, as in Fig. 2

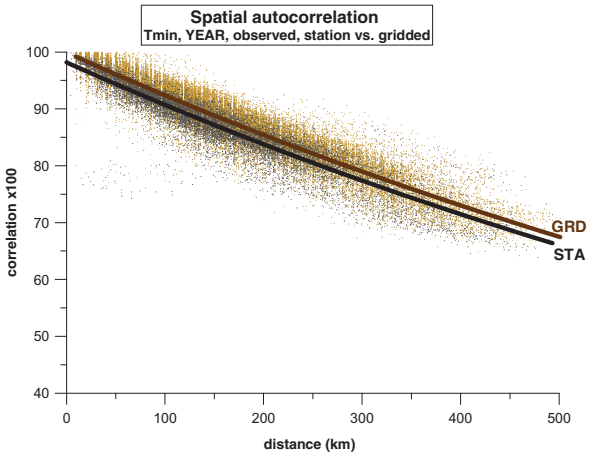


Fig. 6 Dependence of autocorrelation on distance, for observed data, minimum temperature and whole year, for gridded (*brown*) and station (*black*) datasets. *Dots* correspond to individual grid point/station pairs, *bold lines* are exponential fits

distances. The same plot for maximum temperature shows an even closer agreement, the two exponential fits being almost identical for short distances and differing by a maximum of 0.01 for the longest distances (not shown). Analogically to the persistence, the high similarity in the autocorrelation plots between the two datasets allows us to take them as equivalent. Therefore, we only present results for the gridded dataset here without any loss of generality.

Figure 7 displays autocorrelation plots for both temperatures, for the whole year (top panels) and the two seasons (bottom panels). A comparison of the top panels reveals that in reality, the autocorrelations are slightly lower for minimum temperature than for maximum temperature. The LCM model underestimates the autocorrelations severely, other SDS models overestimate them, the LLM, RBF and MLP models being quite close one to another especially for maximum temperature. The overestimation is larger for MLR. Of the two RCMs, RegCM seems to perform slightly better, overestimating the autocorrelation for minimum temperature (with a smaller error than the SDS models especially for long distances) and being almost correct for maximum temperature. ALADIN underestimates autocorrelations for both temperature variables.

The bottom panels of Fig. 7 provide an insight into the seasonality of the performance of the models in terms of autocorrelations. The LCM model is again an outlier in all cases; similarly to persistence, its autocorrelations for maximum temperature in JJA are considerably higher, but still remain, except for the longest distances below any other model. In general, RCMs seem to do a better job than SDS models in simulating autocorrelations. All SDS models overestimate the autocorrelations (except for the shortest distances and except for outlying LCM), MLR being the model for which the overestimation is strongest in all cases. The performance of the LLM, RBF and MLP SDS models is again fairly similar. In winter, a systematic error is apparent in both RCMs: ALADIN underestimates, while RegCM overestimates the autocorrelations. In summer, the RCMs are unable to reproduce the diurnal cycle of autocorrelations, while in reality and in SDS models, the autocorrelations are higher for maximum temperature; ALADIN makes no difference between them (the two blue lines in Fig.7d almost coincide) and RegCM even simulates minimum temperature to autocorrelate more strongly.

4 Statistical distributions

In this section, we present the validation of several characteristics of statistical distributions of both temperature variables. Specifically, we discuss skewness and kurtosis, and also the 5th and 95th percentiles as descriptors of extremes. We display results for the gridded dataset only because results for the two datasets are very similar, analogously to the temporal and spatial autocorrelations, and their presentation would thus be redundant.

Figure 8 contains maps for low temperatures (5th percentile of minimum temperature) in winter and high temperatures (95th percentile of maximum temperature) in summer. The observed values clearly reflect terrain features and the elevation in particular. In winter, extremely low temperatures are warmest in the lowlands of Lower Austria (easternmost part of Austria in the analysis domain), while coldest low temperatures are observed in the Slovakian mountains (northeastern edge of the domain), Czech-Moravian highlands (northwest of the domain) and in particular in the Bohemian Forest along the border of Germany. In summer, extremely high temperatures are hottest in the Hungarian and Slovakian lowlands, while they are least hot in the hilly and mountain areas of central Slovakia, Czech Republic and also along the borders of the Czech Republic with Austria and Germany.

All the models reproduce the major features of the extreme temperature patterns. However, their majority simulate the extremely cold events too warm and extremely warm events too cool. The only exception is the LCM SDS model for winter low temperatures and the ALADIN RCM for summer high temperatures. However, ALADIN exhibits another deficiency in the spatial variability of extreme temperatures: the low minima in winter are too little spatially variable while the gradient in the high maxima in summer is considerably exaggerated. All these effects can also be seen in box-and-whisker plots in Fig. 9. The overestimation of the 5th temperature percentile (that is, the left tail being simulated too short or light) in winter is much less pronounced in maxima (for which

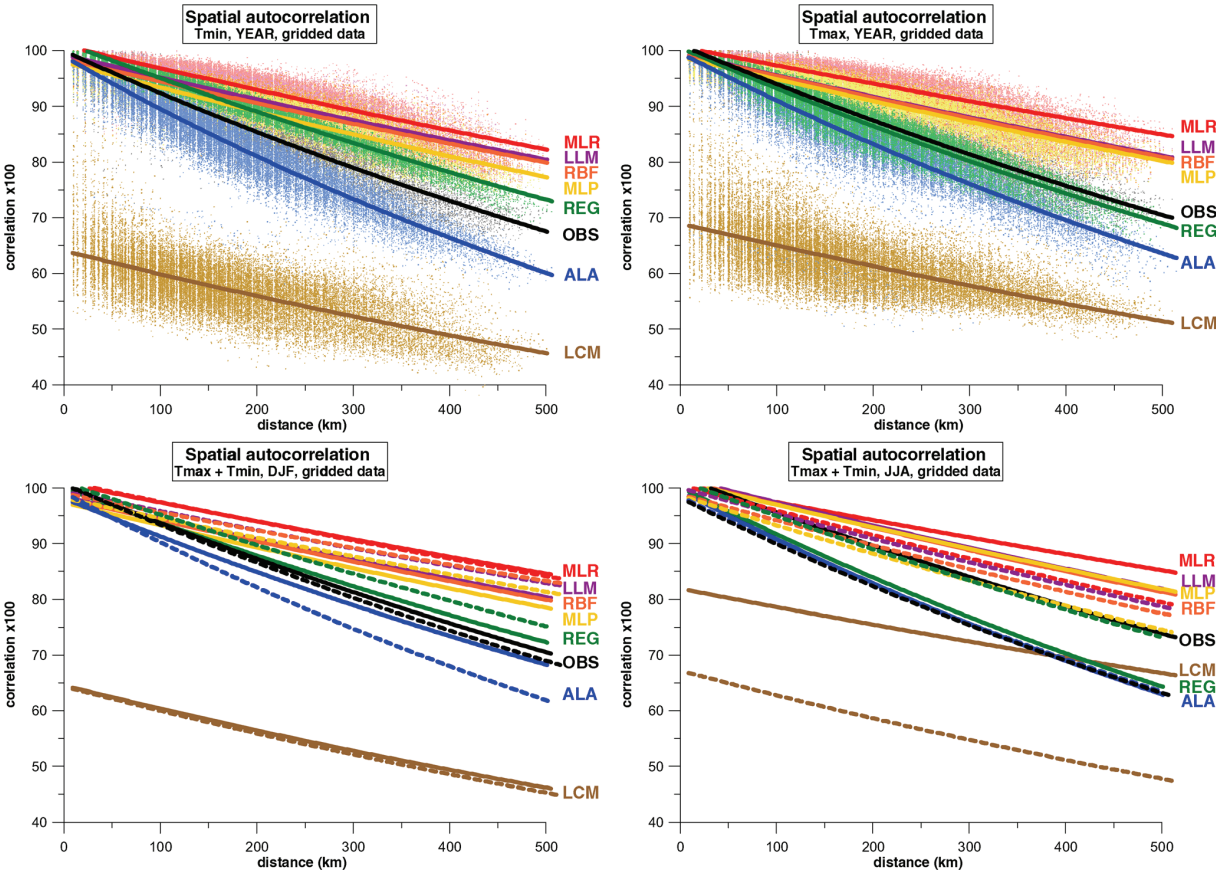


Fig. 7 Dependence of autocorrelation on distance for the gridded dataset. *Dots* correspond to individual grid point/station pairs, *bold lines* are exponential fits: observed (*black*), ALADIN (*blue*), RegCM (*green*), MLR (*red*), LLM (*violet*), LCM (*brown*), RBF (*orange*) and MLP (*yellow*). *Top left*: whole year, minimum temperature; *top right*: whole year, maximum temperature; *bottom left*: DJF, both temperatures; *bottom right*: JJA, both temperatures. In the *bottom* panels, *dots* are omitted, and *solid* (*dashed*) *lines* correspond to maximum (minimum) temperature

the errors generally remain below 1.5 °C) than in minima (for which the errors are between 1 and 3 °C except LCM). In summer, SDS models underestimate high summer temperature quantiles, the error being stronger for maximum temperature; ALADIN exaggerates regional variations of the 95th percentile, while RegCM underestimates the 95th percentile of maximum temperature.

Skewness and kurtosis do not manifest any clear geographical structure over the domain, their spatial patterns being more or less chaotic; therefore, displaying maps would be redundant. We present results of their validation in terms of box-and-whisker plots only.

Observed minimum temperature is negatively skewed (Fig. 10), more in winter and less in summer, indicating a heavier left tail of its distribution. This feature is reproduced by all the models, RegCM underestimating it (i.e. producing overly symmetric statistical distributions) in both seasons. This underestimation is present in minimum temperature outputs from ALADIN and MLR in DJF; both models reproduce minimum temperature skewness correctly in JJA. The underestimation of the negative skewness by both RCMs in winter appears to be the cause of too warm 5th minimum temperature percentiles reported above. LLM, RBF and MLP models are successful in a correct reproduction of skewness in both seasons. Observed maximum temperature is skewed slightly positively in both seasons. No model reproduces this, most of the models producing skewness near zero. RegCM outputs exhibit an unrealistically large range of skewness values in summer.

Kurtosis (Fig. 11) in summer is negative (i.e. the distribution is less peaked than Gaussian, with thinner tails) for both minimum and maximum temperatures. This is more or less reproduced by all the models; only maximum temperature in RegCM possesses kurtosis of the opposite sign. In winter, kurtosis is mostly positive for minimum temperature, which is again reproduced by most models except ALADIN and MLR. Near-zero kurtosis of maximum temperature with a tendency towards negative values is correctly simulated by all the models except LCM and, in summer, RegCM.

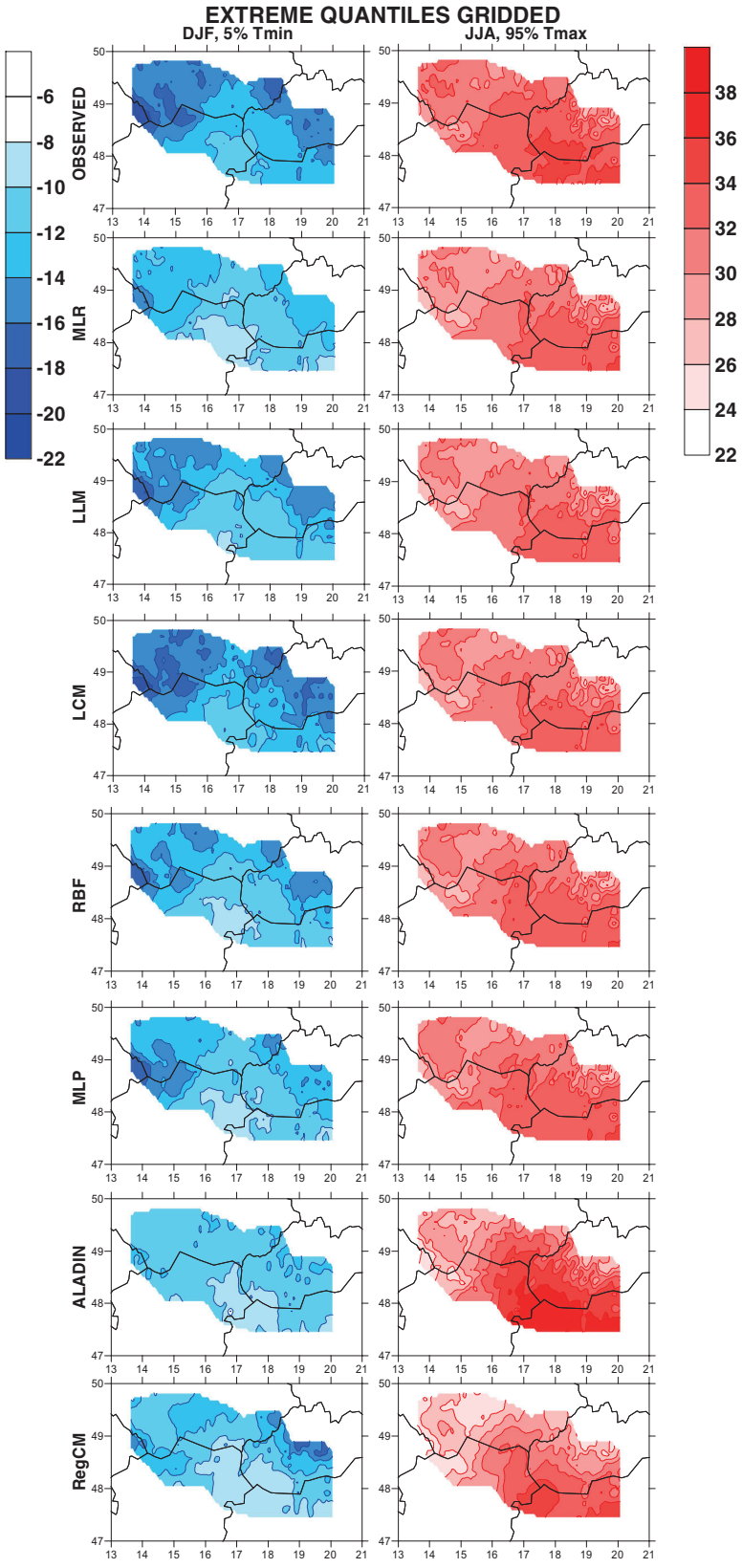


Fig. 8 Extreme temperature quantiles (in °C) for gridded data. *Left*: 5th percentile of minimum temperature, DJF; *right*: 95th percentile of maximum temperature, JJA

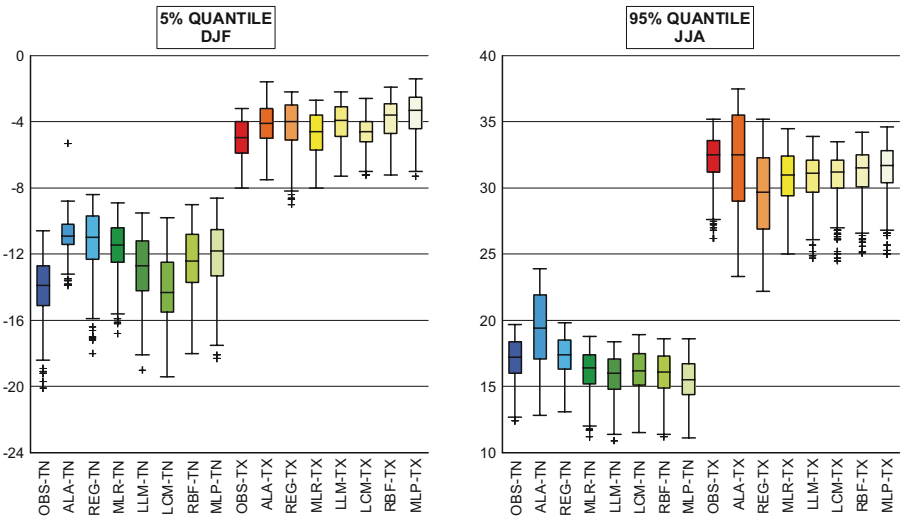


Fig. 9 *Box-and-whisker plots* for the 5th percentile in DJF (*left*) and 95th percentile in JJA (*right*) of minimum and maximum temperature (in °C). Otherwise, as in Fig. 4

5 Fit to observations

Maps of correlations of downscaled temperature series with gridded observations are presented in Figs. 12 and 13 for winter and summer, respectively. Equally to the previous sections, results for station data are almost identical and are therefore not discussed here. As in most other criteria, LCM is an outlier among the SDS models, performing much worse than other models. The seasonality of performance is opposite between the SDS models and RCMs: while the SDS models exhibit higher correlations in summer, the RCMs are more correlated with observations in winter. In general, the performance of SDS models and RCMs is comparable in winter whereas SDS models are much better in terms of the fit to observations in summer, especially for maximum temperature, for which correlations exceed 0.95 at some places. One can note an altitudinal dependency of correlations. SDS models tend to have better fit to observations in elevated areas, especially in Slovakian mountains and along the Czech south-western borders. The opposite is true for RCMs: they tend to correlate worse in mountainous areas. Continentality also plays a role, in particular for maximum temperature in both RCMs in both seasons and in all SDS models in winter: a tendency for a better performance in a more maritime (western) area can be observed.

The evaluation of trends in Fig. 14 should be considered as more or less tentative because only relatively short 10-year series for the 1991–2000 period are examined. Wintertime minimum temperature has increased, and this feature is reproduced by all the models; only RegCM exaggerates the increase rate by a factor of more than two. On the other hand, winter maxima almost do not change; however, all the models

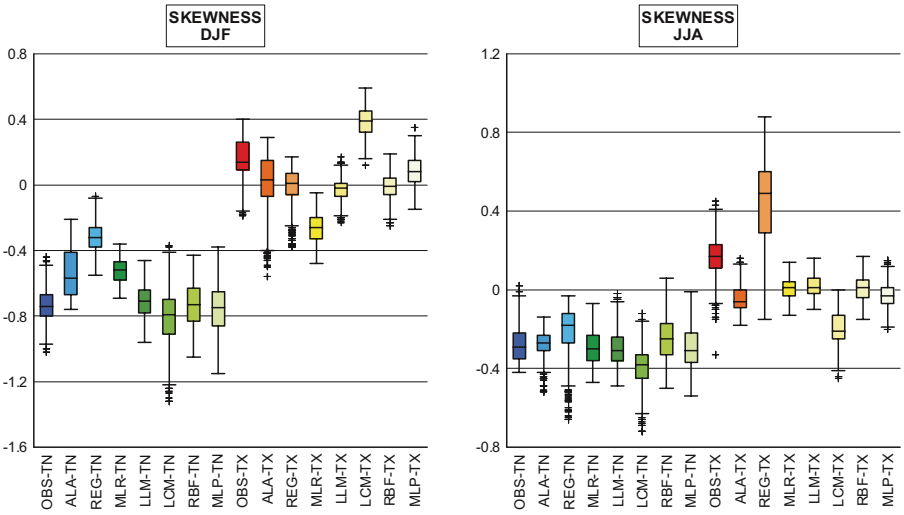


Fig. 10 Skewness for minimum and maximum temperature, DJF (*left*), JJA (*right*); otherwise, as in Fig. 4

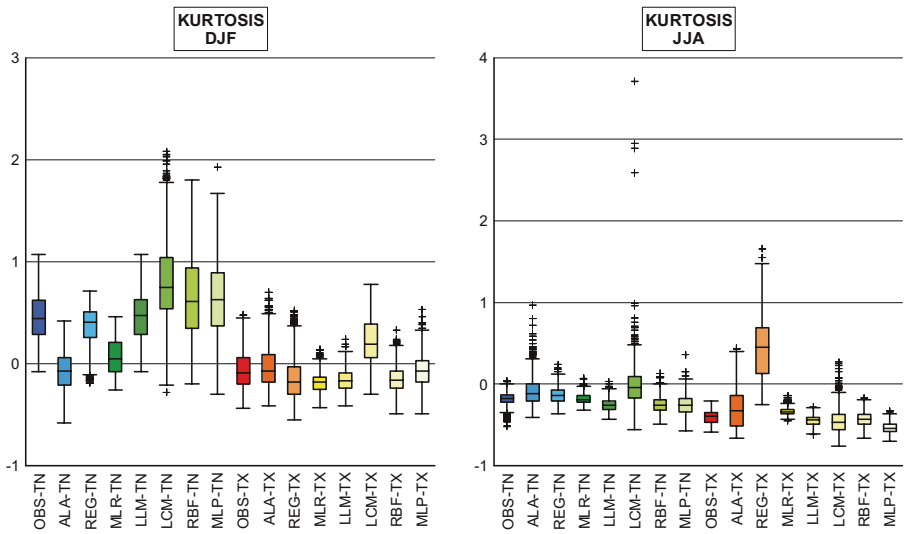


Fig. 11 Kurtosis; otherwise, as in Fig. 10

simulate warming of more than 1 °C per decade, except ALADIN where the warming is less fast. As a result, most models are not capable of reproducing the observed negative trend in daily temperature range. In summer, a near-zero trend in minimum temperature, with a slight tendency to cooling at the majority of stations, is more or less reproduced by most models, except ALADIN simulating a ubiquitous cooling of about −1 °C per decade and LCM simulating slight warming. A decrease of maximum temperature in summer by slightly less than 1 °C per decade is reproduced by all the models, although most of them slightly overestimate the cooling.

6 Discussion

A parallel validation against a dense station network and a regular grid demonstrates that results obtained on the gridded data bear great similarity (and, in many cases, are almost identical) to results obtained on the station data. Therefore, our results confirm that using a gridded dataset as a validation base is a valid approach and that the gridded dataset is equivalent to station data for the validation criteria employed.

In most of the following text, we leave the LCM SDS model aside because its performance is considerably inferior in many criteria, this issue being discussed briefly in a separate paragraph.

The fact that both SDS and RCMs were driven by the same reanalysis data made it possible to quantify the fit of downscaled data to observations. A strong seasonality appears in the correlations of downscaled temperature with observations: while the performance of RCMs and SDS models is of comparable accuracy in winter, SDS models are much more successful in reproducing day-to-day temperature variability in summer. The only previous study that conducted a similar analysis for temperature we are aware of is Kidson and Thompson (1998). That study for New Zealand, however, arrived to somewhat different conclusions: an RCM was slightly better than an SDS model, but only an annual analysis was performed there. The geographical dependence of the skill of SDS models agrees with results of previous studies (e.g. Huth 2002): SDS models tend to perform better where the locations are influenced by upper-air atmospheric circulation more directly, that is, at higher altitudes in particular. Although the correlations in RCMs exhibit strong gradients in mountainous areas, too, the tendency towards a better performance at higher elevations is weaker than in SDS models or is completely absent.

The climate change proceeds on time scales much longer than day-to-day; therefore, the reproduction of long-term behaviour of climate elements is also of considerable importance. The long-term behaviour was characterized by linear trends; the overall performance of RCMs and SDS models is of comparable quality in this respect. In light of a strong tendency to the underestimation of trends, noticed by Lorenz and Jacob (2010) in the ENSEMBLES RCMs driven by ERA-40 reanalysis, the most important finding is perhaps that the downscaling models do not underestimate the amplitude of trends wherever a non-zero trend is observed in reality. An opposite is sometimes true: some models exaggerate the magnitude of trends in one season for one temperature variable. Although the generality of these findings is limited because the analysis of trends is based on fairly short time series, our results agree with other studies validating trends in RCMs (Giorgi et al. 2004; Lorenz and Jacob 2010; Bukovsky 2012; Ceppi et al. 2012), which also point to an unsatisfactory reproduction of trends. However, it is not clear if it is the RCMs that can be blamed for the unsatisfactory reproduction

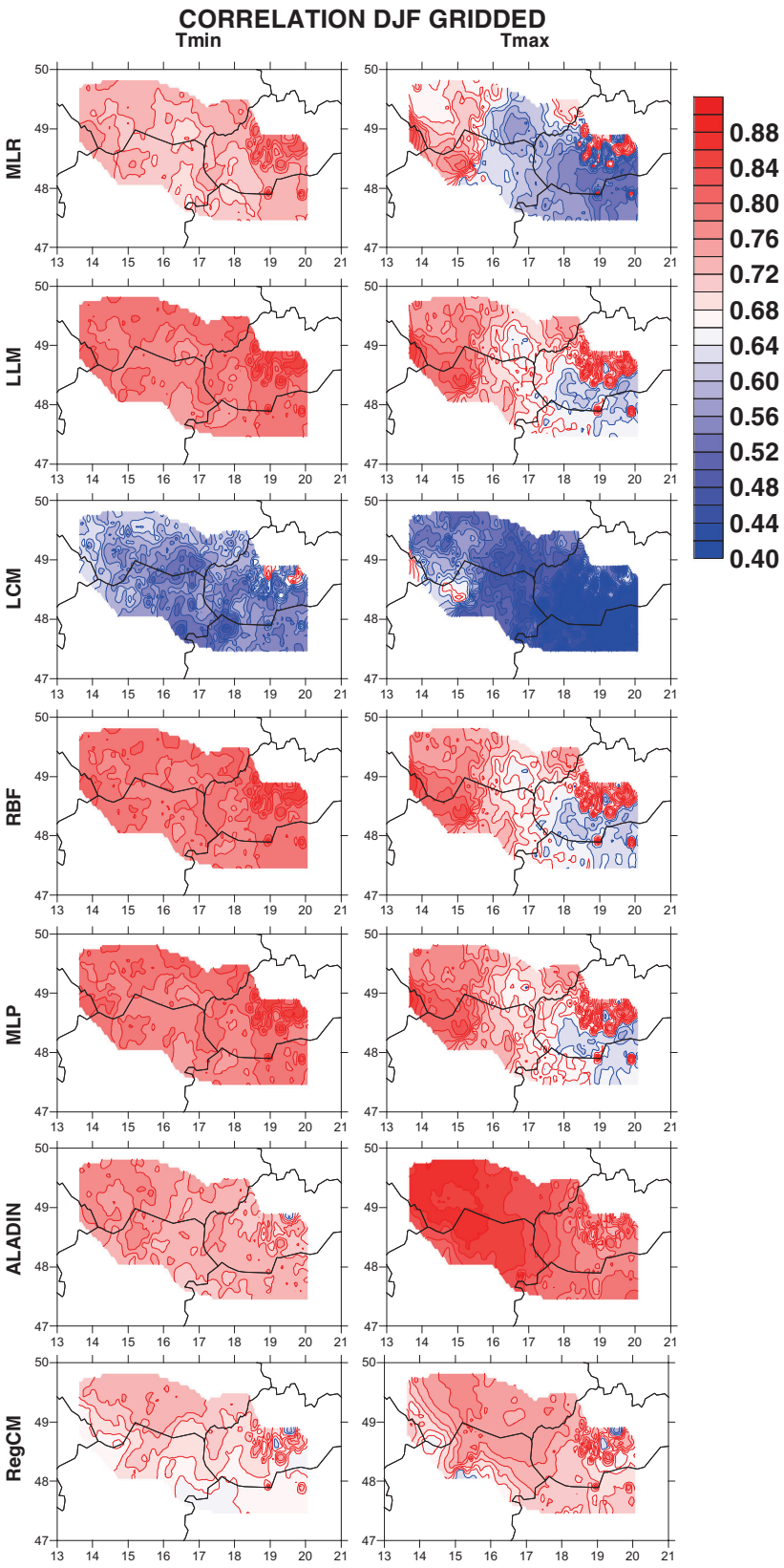


Fig. 12 Correlations (×100) with observations, DJF, gridded dataset, for minimum (*left*) and maximum (*right*) temperature

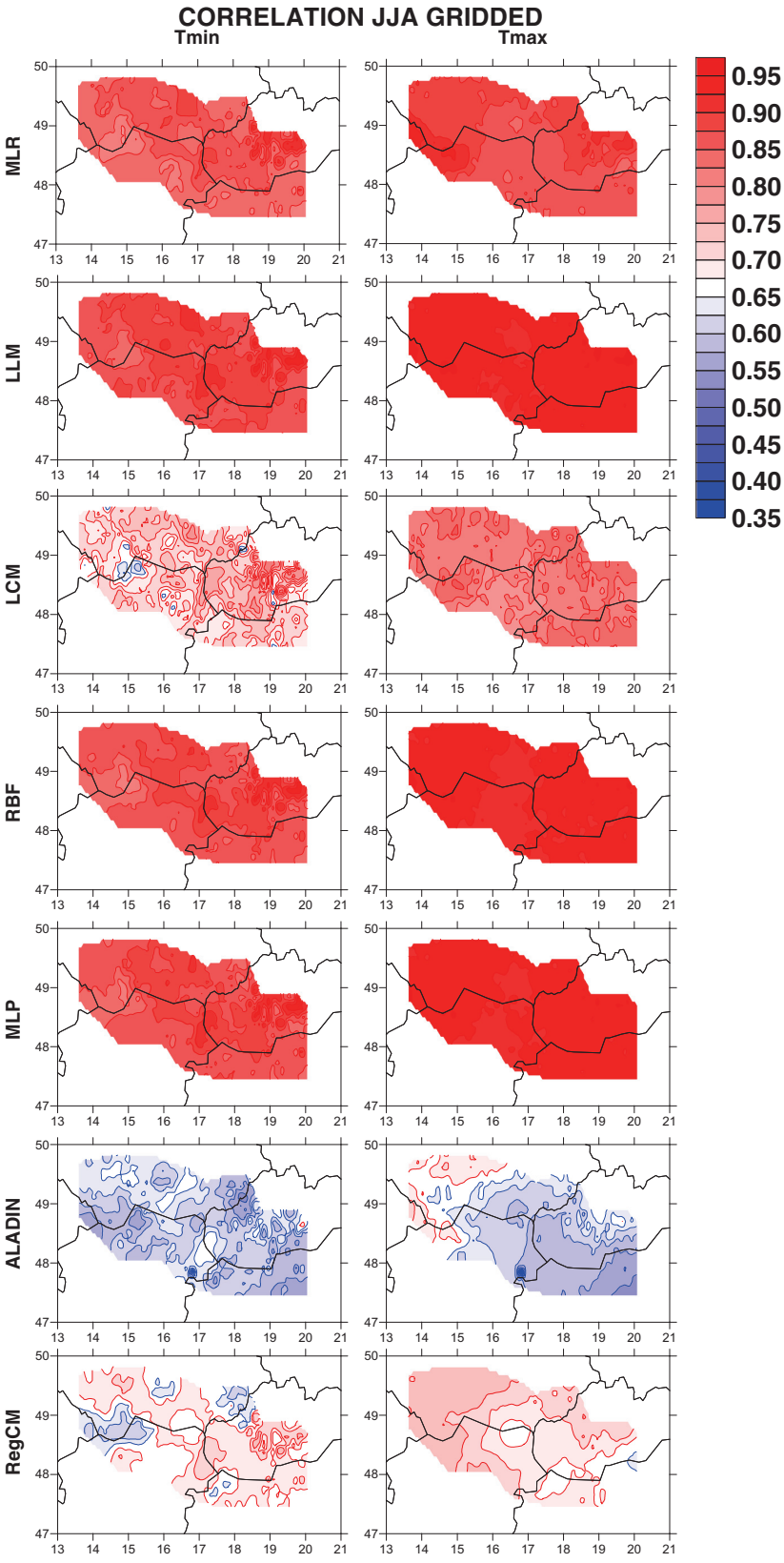


Fig. 13 As in Fig. 12 except for JJA

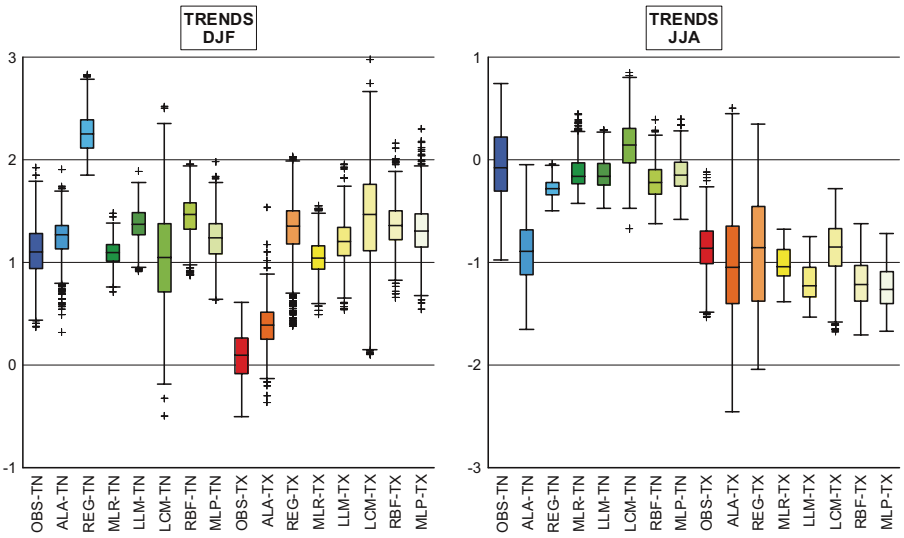


Fig. 14 Trends (in °C per 10 years); otherwise, as in Fig. 10

of trends. It is equally well possible that a wrong behaviour of trends is concealed in the driving reanalyses, and RCMs only transfer it to the surface climate variables.

We present results of two kinds of validation: one for a whole year and the other for specific seasons (winter and summer). Their comparison suggests that an annual validation is likely to conceal seasonal peculiarities and that its interpretation may be misleading. Let us take persistence of minimum temperature as an example: whereas annual results suggest its overestimation by all the models, the seasonal validation reveals that the overestimation takes place in summer only; in winter, no model overestimates it. Therefore, we strongly recommend to conduct the validation studies on a seasonal basis.

It should be emphasized that the SDS models were built for the whole year. It is very likely that SDS models built separately for each season would lead to further improvements in their skill because a better (season-specific) description of links between large-scale fields and local climate could be achieved. On the other hand, this may lead to problems in applications to future climates when seasonality will change.

It is also of interest to compare the performance of linear (MLR) and non-linear (LLM, RBF, MLP) SDS models. There is a great deal of similarity between them. Unlike Huth et al. (2008), we do not observe any superiority of non-linear methods in reproducing characteristics of temperature distributions: the accuracy of simulation of extremes, skewness and kurtosis does not differ between the two groups of methods. It is not clear to us what may be the source of different relative performances of linear and non-linear methods between this study and Huth et al. (2008). It may be worth noting that the linear model yields consistently higher spatial autocorrelations than the non-linear models, but this does not hold for persistence nor for any other validation criterion. Furthermore, we may notice that the validation results of the three non-linear methods are fairly close to each other in most cases.

The overestimation of the 5th percentile by SDS models in winter is accompanied by the underestimation of the 95th percentile (not shown); analogously, the underestimation of the 95th percentile in summer is accompanied by the overestimation of the 5th percentile (not shown). The main cause of this behaviour is the underestimation of the variance inherent in SDS models if variance is not corrected by inflation or another tool. The exaggeration of autocorrelations, both temporal and spatial by SDS models, reported in previous studies (Easterling 1999; Huth 2002; Huth et al. 2008), does not appear to be ubiquitous. While it is present in spatial autocorrelations in both seasons, persistence is overestimated in summer only, but simulated correctly or even underestimated in winter. The seasonality in the performance of persistence is in a good accord with the results of Huth et al. (2001). A commonly suggested remedy to the overestimated persistence in SDS models is the addition of white noise. In the light of our results, the noise to be added to downscaled time series should be spatially uncorrelated, but temporally correlated; only such a setting may reduce spatial autocorrelations to realistic values while not further reducing persistence. Recent SDS studies, going this direction, introduced combinations of deterministic and stochastic models, able to correctly reproduce temporal and spatial autocorrelations (Khalili et al. 2013). Nevertheless, such targeted models improve one or a few characteristics only, leaving other characteristics unchanged or even deteriorating them.

The only SDS model not suffering from the underestimated variance, manifested in systematic errors in the extreme percentiles, is LCM. The reason for this behaviour consists in the fact that it is an analog model, which is able to sample the

whole range of observed values. However, the sampling in LCM is not optimal because it does not reproduce correctly either of the higher-order statistical moments. The strong underestimation of both measures of autocorrelation by LCM is a direct consequence of the fact that it is an analog model: The analogues are selected independently for each day and for every site; the analog procedure does not contain any mechanism for keeping the autocorrelations realistic. Therefore, the analog models cannot be recommended whenever the temporal consistency (and also spatial consistency if the downscaling is conducted at each station (gridpoint) separately) is an issue.

Deficiencies of RCMs cannot be explained by simple statistical arguments as those of SDS models. We notice an overestimation of the 5th temperature percentile by both RCMs in winter, while in summer, RegCM underestimates the 95th percentile of maximum temperature, and ALADIN exaggerates its regional variations for both temperatures. A usual reason for such a kind of discrepancy, the difference in the elevation between a model's grid and station data against which the model is validated, does not seem to be a main factor because these differences are rather small for the RCMs thanks to their high spatial resolution, in which the model topography approximates the real one fairly well. Deficiencies in the performance of the RCMs can, in general terms, be attributed to the formulation of their dynamical cores and physical parameterizations. Since the two RCMs use different parameterizations for all major physical processes, it is hardly possible to trace the differences in their performance to a specific physical process. The relative lack of correspondence between observed and simulated temperatures in terms of low correlations suggests that the boundary forcing in summer is too weak to keep the simulated climate close to observations, allowing RCMs to develop their own small-scale climatology that deviates from reality. The fact that this effect appears in summer and not in winter, and is stronger in ALADIN than in RegCM, may facilitate the localization of the cause in the RCMs' codes. The underestimation of persistence by an RCM in winter and overestimation in summer was reported already almost 20 years ago by Mearns et al. (1995); RCMs do not seem to have improved in this respect since then. Smoother fields of persistence and autocorrelations in RegCM relative to ALADIN, as well as to all other models and observations, are difficult to interpret since both RCMs were subjected to the same procedure of double nesting, with the intermediate models having a 25-km resolution.

This study investigates differences between the downscaling models driven by reanalyzed data, that is, in current climate conditions. Of course, the structure of differences will likely be different when the downscaling models are embedded in GCM simulations of current or future climates.

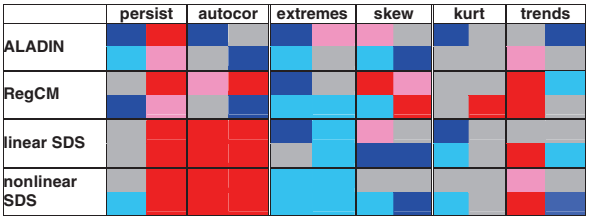


Fig. 15 Overview of the validation analysis. Each *row* corresponds to a model or a group of models; linear SDS includes MLR; non-linear SDS includes LLM, RBF and MLP. LCM is omitted as an outlier in most validation criteria. Each *column* corresponds to one validation criterion. Every entry of the table consists of four *squares*: left (*right*) for DJF (JJA); top (*bottom*) for minimum (*maximum*) temperature. *Red* (*blue*) denotes overestimation (underestimation), *dark* (*light*) *colour* indicating strong (*weak*) under-/overestimation. *Grey* denotes no or negligible error. For extremes, the underestimation (overestimation) means too small (too large) amplitude

7 Conclusions

In this study, we validate minimum and maximum temperature in two RCMs and five SDS models (one of which is linear, two are non-linear based on artificial neural networks and two are local models, which in fact are also non-linear, and one of the local models is effectively an analog model) according to a unified set of criteria that have a potential relevance for impact assessments. The validation is performed for the following three groups of criteria: spatio-temporal characteristics, characteristics of statistical distributions (extremes, higher-order moments), and the degree of fit to observed data. The validation is conducted both on a dense station network and a regular grid with a high resolution of 10 km in central Europe. Therefore, the innovative value of this study is fourfold: (1) it provides a mutual comparison between dynamical and statistical downscaling methods; (2) concentrates on characteristics that have only rarely been validated, but which are potentially relevant to impact studies; (3) the validation is conducted on a very dense network; and (4) station and gridded datasets are compared as validation benchmarks.

The validation of downscaling models in terms of various characteristics of minimum and maximum temperature is summarized in Fig. 15. One can see that all the models succeed in some criteria while fail in others. Their performance differs between seasons, and to a somewhat lesser extent, also between minimum and maximum temperature. Clearly, no model or group of models can be considered superior (or, alternatively, inferior) to others. And, more specifically and perhaps more importantly, our results provide no justification for preferring dynamical models at the expense of statistical models—and vice versa.

The users of climate change scenarios may wish to see clear recommendations as to which methods to use and which methods to condemn. Our result, however, do not allow an

unambiguous guidance to be provided, perhaps except an advice not to use the analog LCM model, which is inferior in most aspects. It is advisable that climate change impact studies utilize ensembles of scenarios coming from different downscaling methods, thereby partially eliminating their weak points or choosing a downscaling method best fitting their needs: e.g. in hydrological applications on a spatial scale of a catchment where spatial structure is relevant, downscaling models capable of replicating spatial autocorrelations are recommendable.

Finally, please recall that all results presented here hold for temperature; for other climate variables such as precipitation, the outputs and recommendations may be different.

Acknowledgments This study was initiated within project CECILIA (Central and Eastern Europe Climate Change Impact and Vulnerability Assessment) funded by the 6th Framework Programme of the European Union, contract 037005. The support by the Czech Science Foundation, project P209/11/2405, and by the CzechGlobe Centre, funded from the European Union and the national budget of the Czech Republic (project CZ.1.05/1.1.00/02.0073 "CzechGlobe–Centre for Global Climate Change Impacts Studies"), is highly acknowledged. The study benefited from networking within the COST ES1102 Action "Validating and Integrating Downscaling Methods for Climate Change Research" (VALUE) where the participation of RH is supported by the Ministry of Education, Youth, and Sports of the Czech Republic under contract LD12059. The authors thank an anonymous reviewer whose comments were very helpful to improving and clarifying the text.

References

Bachner S, Kapala A, Simmer C (2008) Evaluation of daily precipitation characteristics in the CLM and their sensitivity to parameterizations. Meteorol Zeit 17:407–419

Benestad RE, Hanssen-Bauer I, Førland EJ (2007) An evaluation of statistical models for downscaling precipitation and their ability to capture long-term trends. Int J Climatol 27:649–665

Benestad RE, Hanssen-Bauer I, Chen D (2008) Empirical-statistical downscaling. World Scientific, Singapore, 215 pp

Bougeault P (1985) A simple parameterization of the large-scale effects of cumulus convection. Mon Weather Rev 113:2108–2121

Bukovsky MS (2012) Temperature trends in the NARCCAP regional climate models. J Clim 24:3985–3991

Busuioc A, von Storch H (2003) Conditional stochastic model for generating daily precipitation time series. Clim Res 24:181–195

Busuioc A, Giorgi F, Bi X, Ionita M (2006) Comparison of regional climate model and statistical downscaling simulations of different winter precipitation change scenarios over Romania. Theor Appl Climatol 86:101–123

Ceppi P, Scherrer SC, Fischer AM, Appenzeller C (2012) Revisiting Swiss temperature trends 1959–2008. Int J Climatol 32:203–213

Coulibaly P, Dibike YB, Anctil F (2005) Downscaling precipitation and temperature with temporal neural network. J Hydrometeorol 6:483–496

Déqué M (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. Glob Planet Change 57:16–26

Dickinson RE, Henderson-Sellers A, Kennedy PJ (1993) Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR Community Climate Model. Tech. rep., National Center for Atmospheric Research

Easterling DR (1999) Development of regional climate scenarios using a downscaling approach. Clim Change 41:615–634

Eccel E, Ghielmi L, Granitto P, Barbiero R, Grazzini F, Cesari D (2007) Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models. Nonlin Proc Geophys 14:211–222

Elguindi N, Bi X, Giorgi F, Nagarajan B, Pal J, Solmon F, Rauscher S, Zakey A (2007) RegCM Version 3.1 User's Guide. Technical Report, Abdus Salam International Centre for Theoretical Physics

Farda A, Déqué M, Somot S, Horányi A, Spiridonov V, Tóth H (2010) Model ALADIN as regional climate model for central and Eastern Europe. Studia Geophys Geod 54:313–332

Flaounas E, Drobinski P, Vrac M, Bastin S, Lebeaupin-Brossier C, Stéfanon M, Borga M, Calvet J-C (2013) Precipitation and temperature space-time variability and extremes in the Mediterranean region: evaluation of dynamical and statistical downscaling methods. Clim Dyn 40:2687–2705

Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatol 27:1547–1578

Giorgi F, Bi X, Pal J (2004) Mean, interannual variability and trends in a regional climate change experiment over Europe. I. Present-day climate (1961–1990). Clim Dyn 22:733–756

Grell GA (1993) Prognostic evaluation of assumptions used by cumulus parameterizations. Mon Weather Rev 121:764–787

Halenka T, Kalvová J, Chládová Z, Demeterová A, Zemánková K, Belda M (2006) On the capability of RegCM to capture extremes in long term regional climate simulation—comparison with the observations for Czech Republic. Theor Appl Climatol 86:125–145

Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall, Upper Saddle River, New Jersey

Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM (2006) Downscaling heavy precipitation over the UK: a comparison of dynamical and statistical methods and their future scenarios. Int J Climatol 26:1397–1415

Hellström C, Chen D, Achberger C, Räisänen J (2001) Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. Clim Res 19:45–55

Huth R (2002) Statistical downscaling of daily temperature in central Europe. J Clim 15:1731–1742

Huth R, Kyselý J, Pokorná L (2000) A GCM simulation of heat waves, dry spells, and their relationships to circulation. Clim Change 46: 29–60

Huth R, Kyselý J, Dubrovský M (2001) Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series. J Clim 14:4047–4061

Huth R, Kyselý J, Dubrovský M (2003) Simulation of surface air temperature by GCMs, statistical downscaling and weather generator: higher-order statistical moments. Studia Geophys Geod 47:203–216

Huth R, Kliegrová S, Metelka L (2008) Nonlinearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? Int J Climatol 28:465–477

Kalvová J, Nemešová I (1998) Estimating autocorrelations of daily extreme temperatures in observed and simulated climates. Theor Appl Climatol 59:151–164

Kettle H, Thompson R (2004) Statistical downscaling in European mountains: verification of reconstructed air temperature. Clim Res 26:97–112

Khalili M, Van Nguyen VT, Gachon P (2013) A statistical approach to multi-site multivariate downscaling of daily extreme temperature series. Int J Climatol 33:15–32

Kidson JW, Thompson CS (1998) A comparison of statistical and model-based downscaling techniques for estimating local climate variations. J Clim 11:735–753

Kiehl JT, Hack JJ, Bonan GB, Boville BA, Breigleb BP, Williamson D, Rasch P (1996) Description of the NCAR Community Climate Model (CCM3). Tech. Rep. NCAR/TN-420+STR, National Center for Atmospheric Research

Lorenz P, Jacob D (2010) Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40. Clim Res 44:167–177

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user. Rev Geophys 48, RG3003

Mearns LO, Giorgi F, McDaniel L, Shields C (1995) Analysis of variability and diurnal range of daily temperature in a nested regional climate model: comparison with observations and doubled CO2 results. Clim Dyn 11:193–209

Mikšovský J, Raidl A (2005) Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures. Nonlin Proc Geophys 12:979–991

Mpelasoka FS, Mullan AR, Heerdegen RG (2001) New Zealand climate change information derived by multivariate statistical and artificial neural network approaches. Int J Climatol 21:1415–1433

Murphy J (1999) An evaluation of statistical and dynamical techniques for downscaling local climate. J Clim 12:2256–2284

Noilhan J, Mahfouf J-F (1996) The ISBA land surface parameterization scheme. Glob Planet Change 13:145–159

Osborn TJ, Hulme M (1998) Evaluation of the European daily precipitation characteristics from the atmospheric model intercomparison project. Int J Climatol 18:505–522

Ott E, Sauer T, Yorke JA (eds) (1994) Coping with chaos. Wiley, New York

Rauscher S, Coppola E, Piani C, Giorgi F (2010) Resolution effects on regional climate model simulations of seasonal precipitation over Europe. Clim Dyn 35:685–711

Ritter B, Geleyn J-F (1992) A comprehensive radiation scheme for numerical prediction models with potential applications in climate models. Mon Weather Rev 120:303–325

Schmidli J, Goodess CM, Frei C, Haylock MR, Hundecha Y, Ribalaygua J, Schmith T (2007) Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. J Geophys Res 112, D04105

Schoof JT, Pryor SC (2001) Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. Int J Climatol 21:773–790

Skelly WC, Henderson-Sellers A (1996) Grid box or grid point: what type of data do GCMs deliver to climate impacts researchers? Int J Climatol 16:1079–1086

Štěpánek P, Zahradníček P, Huth R (2011) Interpolation techniques used for data quality control and calculation of technical series: an example of central European daily time series. Idöjárás 115:87–98

Themeßl MJ, Gobiet A, Heinrich G (2012) Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. Clim Change 112:449–468

Trigo RM, Palutikof JP (1999) Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. Clim Res 13:45–59

Trigo RM, Palutikof JP (2001) Precipitation scenarios over Iberia: a comparison between direct GCM output and different downscaling techniques. J Clim 14:4422–4446

Uppala SM, Kållberg PW, Simmons AJ, Andrae U, da Costa BV, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen A, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Bethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J (2005) The ERA-40 re-analysis. Quart J Roy Meteorol Soc 131:2961–3012

von Storch H (1999) On the use of "inflation" in statistical downscaling. J Clim 12:3505–3506

von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6:1161–1171

Wilby RL, Hay LE, Gutowski WJ Jr, Arritt RW, Takle ES, Pan Z, Leavesley GH, Clark MP (2000) Hydrological responses to dynamically and statistically downscaled climate model output. Geophys Res Lett 27:1199–1202

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Clim 12:2474–2489

APPENDIX V

KRIŽAN, P., J. MIKŠOVSKÝ, M. KOZUBEK, W. GENGCHEN, AND B. JIANHUI (2011), Long term variability of total ozone yearly minima and maxima in the latitudinal belt from 20°N to 60°N derived from the merged satellite data in the period 1979-2008, *Advances in Space Research*, *48*(12), 2016-2022, doi:10.1016/j.asr.2011.07.010.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Advances in Space Research 48 (2011) 2016–2022

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Long term variability of total ozone yearly minima and maxima in the latitudinal belt from 20°N to 60°N derived from the merged satellite data in the period 1979–2008

Peter Krizan [a,*], Jiri Miksovsky [b], Michal Kozubek [a,b], Wang Gengchen [c], Bai Jianhui [c]

[a] Institute of Atmospheric Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
[b] Department of Meteorology, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
[c] Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Received 25 November 2010; received in revised form 18 July 2011; accepted 20 July 2011
Available online 27 July 2011

Abstract

This paper deals with the behavior of the annual cycle of total ozone (ACO3) and its amplitude ($O3_{AMP}$) in the latitudinal belt from 20°N to 60°N. The prominent feature of the $O3_{AMP}$ spatial pattern is the sharp maximum over the north-east coast of Asia. The spatial correlation of $O3_{AMP}$ with its highest/lowest value varies with location: in the middle latitudes it correlates predominantly with the values of annual maxima of total ozone, while in the lower latitudes, there is a strong negative correlation with the values of ACO3 minima. Regarding temporal evolution of $O3_{AMP}$ we detected distinct negative trend in the period of 1979–1995 which is caused by stronger negative trend of maxima than the negative trend of minima in ACO3. In the period 1995–2008 we found the positive trend of ACO3 in most regions due to stronger positive trend of maxima than that of minima in ACO3 in the middle latitudes (especially over the central and northern Europe and the north-east Asia). In the lower latitudes a weak negative trend of $O3_{AMP}$ was identified and linked to weaker positive trend of maxima than positive trend of minima in ACO3. The behavior of the temporal trends was linked to the changes in Brewer–Dobson circulation and to the trends of tropopause pressure.

Keywords: Annual cycle; Total ozone; Trends in annual cycle; Long-term trends

1. Introduction

Ozone is an important trace gas in the atmosphere. Chapman (1930) hypothesized that UV radiation plays a crucial role in ozone photochemistry. The atomic oxygen is formed by dissociation of oxygen by solar photons. It is very reactive and thus quickly combines with the molecular oxygen to form ozone. Ozone absorbs UV radiation and protects the life on the Earth. Ozone is destroyed not only by absorption of UV radiation but also by reactions with chlorine, bromine, hydrogen and nitrogen in the atmosphere (e.g. Mohanakumar, 2008).

The first studies concerning the influence of halocarbons on total ozone appeared in the first half of the 1970s (Molina and Rowland, 1974; Stolarski and Cicerone, 1974). The authors pointed out that the man-made ozone depleting substances (ODS) in the atmosphere could destroy ozone layer. In the late 1980s, a decreasing trend of total ozone was observed even in the middle latitudes of both hemispheres (Rowland et al., 1988). Similar results were obtained by Bialek (2006) who found the negative trend of total ozone to be strongest in winter/spring and weakest in autumn for period 1980–2003. This negative trend was related to increasing of ODS in the atmosphere. In 1985, the ozone hole over Antarctic region was discovered

* Corresponding author. Tel.: +420 267103055.
E-mail addresses: krizan@ufa.cas.cz (P. Krizan), jiri.miksovsky@mff.cuni.cz (J. Miksovsky), wgc@mail.iap.ac.cn (W. Gengchen).

0273-1177/$36.00 © 2011 COSPAR. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.asr.2011.07.010

(Farman et al., 1985) as the strongest manifestation of influence of ODS on ozone layer. This discovery led to the international effort which resulted in the Montreal Protocol in 1987 with amendments in the following years. In the mid-1990s a turnaround of trend in total ozone was observed in the northern middle latitudes. This change has a dynamical origin because ODS in the stratosphere peaked in the late 1990s, not in the mid-1990s (Dhomse et al., 2006). Due to this change a simple linear trend of total ozone is not proper from the mid-1990s. A piecewise linear trend is proposed to be used (Reinsel et al., 2002). Krzyścin (2011) compared classical regression model, piecewise regression model and flexible trend model and he concluded that all models give significant positive trend in the period 1996–2008 for the total ozone averaged over the globe in boreal winter.

The annual cycle is a major component of the global ozone variation. In the middle latitudes we observe a maximum of the annual cycle in late winter/early spring and a minimum in summer/fall. This annual cycle is substantially affected by the Brewer–Dobson circulation which transports ozone from the tropics to high latitudes. The Brewer–Dobson circulation is stronger in the winter hemisphere than in the summer one.

This paper deals with the spatial distribution of annual cycle of total ozone in latitudinal belt from 20°N to 60°N, the dependence of annual cycle on its annual extremes, temporal trends of annual cycle of total ozone and its components and possible links of the detected trends to changes of temperature at 100 hPa and tropopause pressure. We look for change in trends of annual cycle of total ozone and its components in the mid-1990s. Section 2 treats the data and method, Section 3 gives the main results, Section 4 provides the discussion of the results and Section 5 summarizes the conclusions.

2. Data and method

Monthly averages of total ozone are taken from the TEMIS datasets at http://www.temis.nl/protocols/O3global.html. More about the TEMIS datasets can be found in van der A et al. (2010). Monthly means of total ozone in the period 1978–2008 are available in this database but we took into account only period 1979–2008 because the regular satellite observations of total ozone started in October 1978 and we wanted to have observations of total ozone in each month of the year. The TEMIS ozone data are available in regularly positioned grid points, spaced 1.5° in longitude and 1° in latitude. Maximum (minimum) in the annual cycle was computed as maximum (minimum) from monthly means of total ozone from the given year and amplitude of annual cycle was determined as difference between these two extremes.

This paper is done in the framework of cooperation between the Czech and Chinese Academy of Sciences and thus we are predominantly interested in distribution and trends of the amplitude of the annual cycle of total ozone (ACO3) over Europe and the territory of China. Since the latitude of the southernmost point of continental China is about 20°N and the latitude of northernmost point is about 53°N, we select a belt bounded by 20°N and 60°N. We did not want to apply any artificial borders, so we performed the analysis for the whole range of longitudes.

In order to identify trends in the values of annual cycle extremes piecewise linear regression was employed with the breakpoint in 1995 reflecting the change in the total ozone trend in the mid-1990s. This procedure was performed not only for annual cycle of total ozone but also for its yearly maxima and minima as well as for maxima and minima of two potential covariates: temperature in the 100 hPa level and pressure at the tropopause, both of which were obtained in the 2.5 to 2.5° horizontal resolution from the NCEP/NCAR reanalysis dataset (Kalnay et al., 1996). The trend estimation procedure was carried out individually for each grid point.

3. Results

The geographic distribution of the amplitude of annual cycle of total ozone ($O3_{AMP}$) in the latitudinal belt from 20°N to 60°N is shown in Fig. 1a. The lowest values of $O3_{AMP}$ occur in the subtropics (about 40–60 D.U.). There is an increase towards higher latitudes where the amplitude typically exceeds 100 D.U. with strong longitudinal variation. The highest values are observed poleward of 45°N near 150°E (Okhotsk Sea region). In this maximum $O3_{AMP}$ reaches more than 140 D.U. compared to the minimum at the same latitude over the northern Atlantic ocean (less than 100 D.U.).

The amounts of ozone in the maximum of ACO3 ($O3_{MAX}$) are shown in Fig. 1b. The observed pattern strongly resembles that of Fig. 1a, i.e., the behavior of $O3_{AMP}$ is similar to behavior of ACO3 maximum. The lowest values of $O3_{MAX}$ are observed in low latitudes (about 300 D.U.), and they increase towards the pole as well as well as their longitudinal variations. The global maximum is found at about 55°N and 150°E (460 D.U.). The most profound longitudinal variations of $O3_{MAX}$ (about 80 D.U.) are detected along 55°N with the highest values observed near the Okhotsk Sea and the lowest ones over continental Europe and Asia and eastern Atlantic.

The lowest amounts of ozone in the ACO3 minima ($O3_{MIN}$) are again observed in subtropical regions (Fig. 1c). There is an increase with latitude but not so strong as in the case of $O3_{MAX}$. The Okhotsk Sea maximum is less pronounced, broader and shifted eastward. The maximum over Canada (55°N, 70°W) is better visible in the values of minima in ACO3 than in the case of its maxima.

Pearson correlation coefficient has been computed to quantify the strength of the link between $O3_{AMP}$ and values of its maxima and minima at each latitude (Fig. 2). The correlation between $O3_{AMP}$ and $O3_{MAX}$ is positive for all latitudes, within the range from 0.4 at 20°N to more than 0.9 north of approximately 40°N. The correlation is

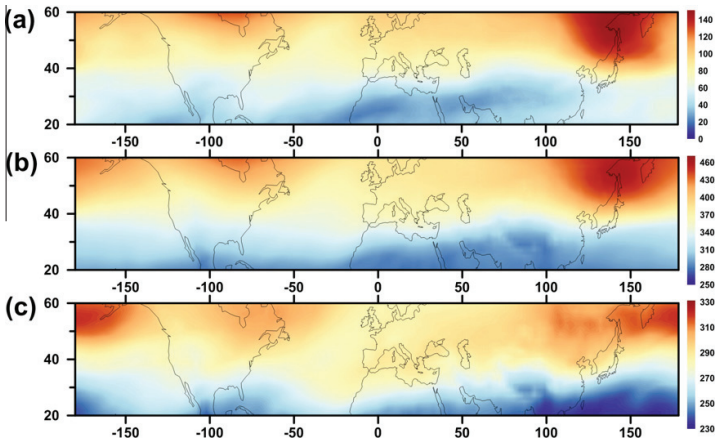


Fig. 1. Mean amplitude (a), maximum value (b) and minimum value (c) of the annual cycle of total ozone, for the period 1979–2008 (in Dobson units).

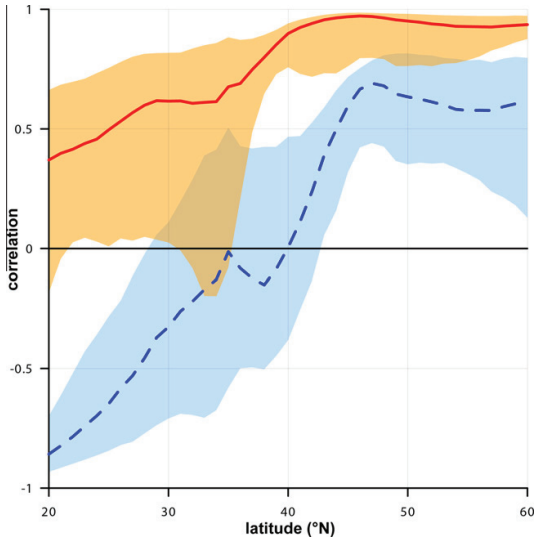


Fig. 2. Longitudinal correlation (vertical axis) of the mean amplitude of the total ozone annual cycle with the mean values of the annual cycle's maximum (full line) and minimum (dashed line), computed for grid points located at different latitudes (horizontal axis). The shaded areas show the 95% level confidence intervals.

statistically significant at the 95% confidence level northward of about 35°N.

The latitudinal dependence of correlation between $O3_{AMP}$ and $O3_{MIN}$ is more complicated. We observe significant negative correlations equatorward of 28°N and significant positive correlations poleward of 42°N. From 28°N to 42°N the correlation is weaker and statistically insignificant. The maximal positive correlation between these variables is observed near 45°N with slight poleward decrease similar to the case of maxima in ACO3. Confidence intervals of the correlation coefficient in Fig. 2 are constructed using bootstrap resampling; because of strong spatial autocorrelation in the series, a correction for serially correlated data was applied following the technique of Mudelsee (2003).

Shape and phase of O3 annual cycle vary with geographic location. In Fig. 3 we can see months in which ACO3 maxima (red circles) and minima (blue circles) are observed in selected grid points. The ACO3 maxima typically occur in late winter/early spring in the middle latitudes while in the subtropics they are shifted 2–4 months towards summer. The middle latitudes ACO3 minima are detected predominantly in fall while the subtropical ones occur in winter. This temporal pattern is caused mainly by the Brewer–Dobson circulation which transports ozone rich air from tropics and leads to accumulation of ozone in high latitudes in winter. This transport also results in occurrence of ACO3 minima in subtropical region.

The trend of $O3_{AMP}$ in the period 1979–1995 is shown in Fig. 4a. The grid points, in which the trend is significant at the 95% level, are highlighted by dots (blue for the negative trends, red for the positive ones; testing was done by bootstrap resampling). The majority of grid points with significant decrease of $O3_{AMP}$ are located north of approximately 40°N. The strongest negative trend is observed over the eastern Asia in the area of the highest amount of total ozone. Mostly insignificant change of $O3_{AMP}$ is observed southward of approximately 35°N. A positive trend of ACO3 can be seen at the majority of the middle latitudes after the break point in 1995 and it is statistically significant predominantly over Scandinavia, central Canada and eastern Siberia (Fig. 4b).

4. Discussion

The highest values of $O3_{AMP}$ are observed at the eastern coast of Asia due to very high values of total ozone in $O3_{MAX}$. We also observe the growth of $O3_{AMP}$, $O3_{MAX}$ and $O3_{MIN}$ with latitude. This increase can be explained

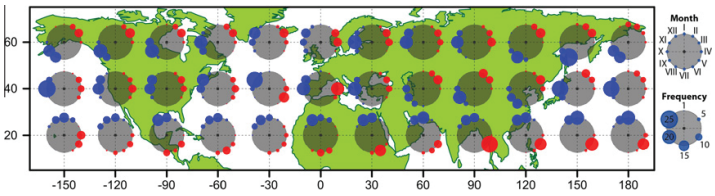


Fig. 3. Number of maxima (red) and minima (blue) of the annual cycle of total ozone occurring in individual months during the 1979–2008 period. The results are shown for grid points located in the center of the shaded discs, with size of the coloured circles on their perimeters proportional to the absolute frequency of the extremes.

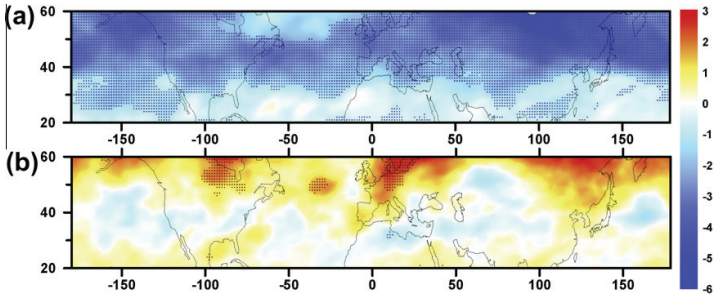


Fig. 4. Temporal trends of the amplitude of total ozone annual cycle in Dobson units per year, computed for the periods 1979–1995 (a) and 1995–2008 (b). Dots represent grid points with trends statistically significant at the 95% level.

by the Brewer–Dobson circulation which transports ozone rich air from low latitudes toward the pole. The circulation is stronger in the winter hemisphere therefore the ozone amount is greater in middle latitudes in late winter/early spring than in summer/fall and thus the latitudinal gradient of total ozone is stronger in winter than in summer. The most pronounced feature of spatial distribution of $O3_{AMP}$ is the strong maximum near the eastern coast of Siberia. This maximum can be linked to a high value of ozone in maximum of ACO3, already mentioned by Dutsch (1973). Holton (1972) stated that when Rossby waves propagate upward, we observe their westward tilting, and thus the elevated ozone near the eastern coast of Siberia is caused by the Aleutian low which is situated eastward. The Island low is less pronounced in $O3_{MAX}$. Lower values of $O3_{MAX}$ over central Asia and Europe are related to the presence of anticyclone in winter above Asia and to asymmetric spatial expansion of polar vortex (WMO, 2002). In summer the elevated values of $O3_{MIN}$ are shifted to the east, because no vertical propagation of Rossby waves takes place due to the east winds in the stratosphere (Charney and Drazin, 1961).

In Fig. 5a the trend of $O3_{MAX}$ during the years 1979–1995 is shown. There is strong statistically significant decrease except the subtropical latitudes. In the same period we observe weaker negative trend of ACO3 minima which is statistically significant at smaller number of grid points (Fig. 6a). The decrease of $O3_{AMP}$ is thus observed due to stronger negative trend in O3 maxima than that of O3 minima in 1979–1995.

In the period 1995–2008 the trend of $O3_{MAX}$ (Fig. 5b) is reverted to positive poleward of 40°N and it is statistically significant mostly in Europe. In central Asia the decrease of $O3_{MAX}$ continues, but it is weaker in magnitude and not statistically significant. In subtropical regions there are insignificant negative trends of ACO3 maxima at majority of grid points.

After 1995 we observe slight positive trend of $O3_{MIN}$ (Fig. 6b) which is significant over northern Canada and in some places in the subtropics. As the result of the evolutions of $O3_{MAX}$ and $O3_{MIN}$ after 1995, the trend of $O3_{AMP}$ is positive in middle latitudes due to stronger increase of ACO3 maxima than that of ACO3 minima, but there is a decrease of $O3_{AMP}$ in the subtropics due to negative trend of $O3_{MAX}$ and positive trend of $O3_{MIN}$.

The negative trend of ACO3 maxima in the period 1979–1995 is caused not only by the increasing level of ozone depleted substances. The influence of stratospheric dynamics, volcanic aerosols and solar activity must also be taken into account. Fusco and Salby (1999) found that the planetary wave activity in the stratosphere decreases in the period 1979–1990. This phenomenon contributes to the stable and cold polar vortex which produces more polar stratospheric clouds and larger ozone loss.

Solomon et al. (1996) pointed out that the large volcanic eruptions influence the lower stratospheric dynamics by the absorption and scattering of solar radiations. The Mt. Pinatubo eruption in 1991 caused low values of total ozone in the early 1990s (Parrish et al., 1999). These low values have large impact on trend analysis in the period

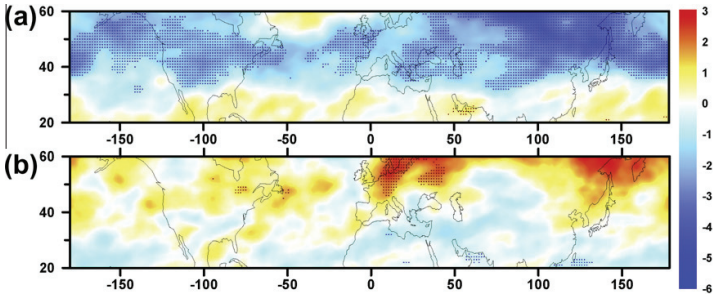


Fig. 5. The same as Fig. 4, for trends of maximum values of the total ozone annual cycle.

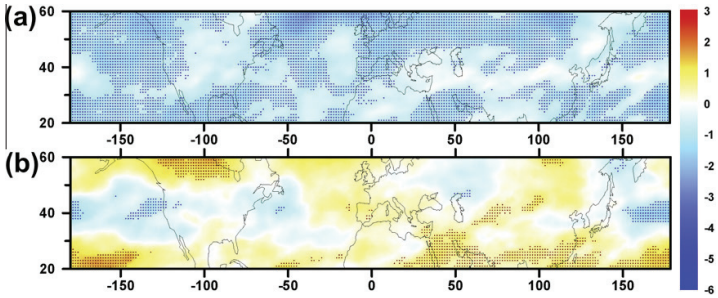


Fig. 6. The same as Fig. 4, for trends of minimum values of the total ozone annual cycle.

1979–1995, because they are situated near its end. According to Hood et al. (1997) the separation of solar and volcanic effects on total ozone is hard, because in the period 1979–1995 both large volcanic eruptions occur in solar maxima (El Chichón in 1982; Mt. Pinatubo in 1991). Steinbrecht et al. (1998) pointed out that there was upward trend of tropopause height in 1967–1997 at the station Hoheinpeissenberg which contributes to 25% of the decrease of total ozone. Similar tendency in the tropopause height was found at the station Munich (Hoinka et al., 1996).

The trend of $O3_{AMP}$ after 1995 is mainly positive, but mostly not statistically significant due to high variability of total ozone in winter and shorter period of observations. This change in trend of $O3_{AMP}$ is caused by dynamical reasons, because the decline of ozone depleting substances is too small for such high increase in total ozone after 1995. (Dhomse et al., 2006) found out that the increasing wave activity is responsible for majority of total ozone increase in the late 1990s. The strengthening of the Brewer–Dobson circulation is expected in the future due to global warming which enhances ozone recovery in the 21st century according to model study of Butchart et al. (2006).

$O3_{MIN}$ in the middle latitudes is controlled predominantly by solar radiation, because minima are observed in late summer/early fall. According to Hadjinicolau et al. (1997) the change during the year is influenced by variations of ozone in the same year, not by previous years, and that is why the changes in summer/fall ozone follow the changes during the winter of the same years. In the period after 1995 we observe positive trend of ACO3 maxima in middle latitudes and ACO3 minima in the subtropics. Both extremes occur during period with strong Brewer–Dobson circulation and we may speculate according to Butchart et al. (2006) that these positive trends are caused by the stronger Brewer–Dobson circulation.

The factors contributing to long-term changes in ozone concentration may also be related to the changes in the thermobaric field. In Fig. 7, the trend of yearly minima of temperature at the 100 hPa level is shown for periods 1979–1995 and 1995–2008. We can see that the negative temperature trend equatoward of 35°N is stronger in the period after 1995 and thus we observe accelerated cooling of the subtropics. In the middle latitudes the tendency toward warming occurs after 1995. Because of the opposite tendencies in the subtropics and the higher latitudes the latitudinal contrast increases. This phenomenon promotes the strengthening of Brewer–Dobson circulation after 1995.

In Fig. 8 the trend of yearly maxima of tropopause pressure is compared for the periods 1979–1995 and 1995–2008. Maxima in tropopause pressure occur in winter when chemical and dynamical influences on total ozone are the strongest. We can see the reverse trend of tropopause pressure before and after 1995 at many places over the world. In Europe, East Asia, north China, subtropical latitudes in Pacific and in the middle latitudes of Atlantic Ocean we observe negative trend of tropopause pressure before 1995 and positive one after 1995, while at Northern Canada and central Asia we observe change from positive to negative trend in 1995. In the same time we observe change in prevailing values of NAO index from negative to

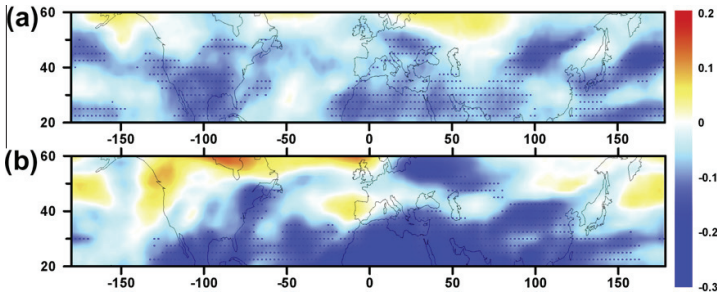


Fig. 7. Temporal trends (°C/year) of the yearly minima of temperature at 100 hPa computed for the period 1979–1995 (a) and for period 1995–2008 (b). Dots represent grid points with trends statistically significant at the 95% level.

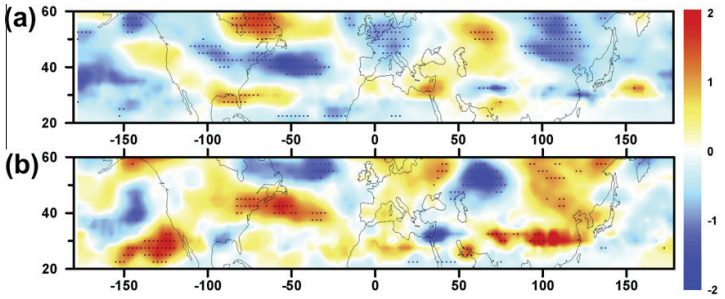


Fig. 8. The same as Fig. 7 for yearly maxima of tropopause pressure (hPa/year).

positive ones (Staehelin et al., 2001) and change in the number of ozone laminae in vertical profile of ozone (Krizan and Lastovicka, 2005; Tarasick et al., 2005).

5. Conclusion

By analyzing the amplitude of annual cycle of O3 ($O3_{AMP}$), we were able to confirm the existence of distinct temporal trends, typically different in sign between the periods 1979–1995 and 1995–2008. In particular, a strong decrease of $O3_{AMP}$, observed in most of the middle latitudes prior to 1995, was followed by an increase in the period 1995–2008 at the majority of locations. In the tropics, where the O3 annual variations are generally less pronounced, only weaker and mostly statistically insignificant tendencies were found. The long-term variations of $O3_{AMP}$ in the middle latitudes are related mostly to the changes in the maxima of the annual cycle, which undergo similar evolution as $O3_{AMP}$ throughout the analyzed period. Finally, our analysis suggests that the detected trends can be related to the evolution of the thermobaric field in the troposphere, especially to the increasing temperature gradient between the tropics and higher latitudes, which accelerates the Brewer–Dobson circulation and thus contributes to increase of annual maxima of total ozone in the middle latitudes since the mid-1990s.

Acknowledgements

This study was supported by the Ministry of Education, Youth and Sports of the Czech Republic through Grant ME 10077 of KONTAKT program and through research Project MSM0021620860 and partly by the Grant Agency of the Czech Republic through Grant 209/10/1792. We also thank to the Ministry of Science and Technology of People's Republic of China (2010-96) and to the Chinese Academy of Sciences (Bilateral exchange 2010). This work was supported by the Grant SVV-2010-261308.

The NCEP/NCAR reanalysis data were downloaded from NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.cdc.noaa.gov. The ozone data were taken from TEMIS database http://www.temis.nl/protocols/O3global.html.

References

Bialek, M. Long-term changes (1980–2003) in total ozone time series over Northern Hemisphere midlatitudes. Acta Geophysica 54, 60–70, 2006.

Butchart, N., Scaife, A.A., Bourqui, M., Grandpre, J., Hare, S.H.E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D., Sigmond, M. Simulations of anthropogenic change in the strength of the Brewer–Dobson circulation. Climate Dynamics 27, 727–741, 2006.

Chapman, S. A theory of upper atmospheric ozone. Mem. Royal Met. Soc. 3, 103–125, 1930.

Charney, J.G., Drazin, P.G. Propagation of planetary-scale disturbances from the lower into the upper atmosphere. Journal of Geophysical Research 66, 83–109, 1961.

Dhomse, S., Weber, M., Wohltmann, I., Rex, M., Burrows, J. On the possible causes of recent increases in northern hemispheric total ozone from a statistical analysis of satellite data from 1979 to 2003. Atmospheric Chemistry and Physics 6, 1165–1180, 2006.

Dutsch, H.U. The ozone distribution in the atmosphere. Canadian Journal of Chemistry 52, 1491–1504, 1973.

Farman, J.C., Gardiner, B.G., Shanklin, J.D. Large losses of total ozone in Antarctica reveal seasonal ClO X/NO X interaction. Nature 315 (6016), 207–210, 1985.

Fusco, A.C., Salby, M.L. Interannual variations of total ozone and their relationship to variations of planetary wave activity. Journal of Climatology 12, 1619–1629, 1999.

Hadjinicolau, P., Pyle, J.A., Chipperfield, M.P., Kettleborough, J.A. Effect of interannual meteorological variability on middle latitude O_3. Geophysical Research Letters 24, 2993–2996, 1997.

Holton, J.R. An Introduction to Dynamic Meteorology. Academic Press, New York, pp. 319, 1972.

Hoinka, K.P., Claude, H., Kohler, U. On the correlation between tropopause pressure and ozone above Central Europe. Geophysical Research Letters, 231753–231756, 1996.

Hood, L.L., McCormack, J.P., Labitzke, K. An investigation of dynamical contribution to midlatitude ozone trends in winter. Journal of Geophysical Research 102, 13079–13093, 1997.

Kalnay, E., Kanamitsu, M., Kistler, R., et al. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77, 437–471, 1996.

Krzyścin, J.W. Onset of the total ozone increase based on statistical analyses of global ground-based data for the period 1964–2008. International Journal of Climatology, doi:10.1002/joc.2264, n/a, 2011.

Krizan, P., Lastovicka, J. Trends in positive and negative ozone laminae in the Northern Hemisphere. Journal of Geophysical Research 110, D10107, doi:10.1029/2004JD005477, 2005.

Mohanakumar, K. Stratosphere Troposphere Interactions: An Introduction. Springer, Cochin, India, ISBN 978-1-4020-8216-0, 416pp., 2008.

Molina, M.J., Rowland, F.S. Stratospheric sink for chlorofluoromethanes, chlorine atom catalysed destruction of ozone. Nature 249, 810–812, 1974.

Mudelsee, M. Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. Mathematical Geology 35, 651–665, 2003.

Parrish, A., Connor, B.J., Tsou, J.J., Beyerle, G., McDermid, I.S. Hollandsworth. Microwave ozone and lidar aerosol profile observations at Table Mountain, California, following Pinatubo eruption. Journal of Geophysical Research 208, 20201–20208, 1999.

Reinsel, G.C., Weatherhead, E.C., Tiao, G.C., Miller, A.J., Nagatani, R.M., Wuebbles, D.J., Flynn, L.E. On detection of turnaround and recovery in trend for ozone. Journal of Geophysical Research 107, D10, doi:10.1029/2001JD000500, 2002.

Rowland, F.S., Harris, N., Bojkov, R.D., Bloomfield, P.B. Statistical error analysis of ozone trends – winter depletion in the northern hemisphere, in: Bojkov, R., Fabian, P. (Eds.), Ozone in the Atmosphere. A. Deepack, Hamptan, pp. 71–75, 1988.

Solomon, S., Portmann, R.W., Garcia, R.R., Thomason, L.W., Poole, L.R., McCormick, M.P. The role of aerosol variations in anthropogenic ozone depletion at northern midlatitudes. Journal of Geophysical Research 101, 6713–6727, 1996.

Staehelin, J., Harris, N.R.P., Appenzeller, C., Eberhard, J. Ozone trends: A review. Reviews of Geophysics 39, 231–290, 2001.

Steinbrecht, W., Claude, H., Kohler, U. Correlation between tropopause height and total ozone: implications for long-term changes. Journal of Geophysical Research 103, 19183–19192, 1998.

Stolarski, R.S., Cicerone, R.J. Stratospheric chlorine: a possible sink for ozone. Canadian Journal of Chemistry 52, 1610–1615, 1974.

Tarasick, D.W., Fioletov, V.E., Wardle, D.I., Kerr, J.B., Davies, J. Changes in vertical distribution of ozone over Canada from ozonesondes: 1980–2001. Journal of Geophysical Research 110, D02304, doi:10.1029/2004JD004643, 2005.

van der A, R.J., Allaart, M.A.F., Eskes, H.J. Multi sensor reanalysis of total ozone. Atmos. Chem. Phys. Discuss. 10, 11401–11448, doi:10.5194/acpd-10-11401-2010, 2010.

World Meteorological Organization. Scientific Assessment of Ozone Depletion. WMO Global Ozone Research and Monitoring Project Report No. 47, Geneva, 2002.

APPENDIX VI

MIKŠOVSKÝ, J., R. BRÁZDIL, P. ŠTĚPÁNEK, P. ZAHRADNÍČEK, AND P. PIŠOFT (2014), Long-term variability of temperature and precipitation in the Czech Lands: an attribution analysis, *Climatic Change*, *125*(2), 253-264, doi:10.1007/s10584-014-1147-7.

Climatic Change (2014) 125:253–264
DOI 10.1007/s10584-014-1147-7

Long-term variability of temperature and precipitation in the Czech Lands: an attribution analysis

Jiří Mikšovský · Rudolf Brázdil · Petr Štěpánek · Pavel Zahradníček · Petr Pišoft

Received: 17 December 2013 / Accepted: 4 May 2014 / Published online: 11 June 2014

Abstract Among the key problems associated with the study of climate variability and its evolution are identification of the factors responsible for observed changes and quantification of their effects. Here, correlation and regression analysis are employed to detect the imprints of selected natural forcings (solar and volcanic activity) and anthropogenic influences (amounts of greenhouse gases—GHGs—and atmospheric aerosols), as well as prominent climatic oscillations (Southern Oscillation—SO, North Atlantic Oscillation—NAO, Atlantic Multidecadal Oscillation—AMO) in the Czech annual and monthly temperature and precipitation series for the 1866–2010 period. We show that the long-term evolution of Czech temperature change is dominated by the influence of an increasing concentration of anthropogenic GHGs (explaining most of the observed warming), combined with substantially lower, and generally statistically insignificant, contributions from the sulphate aerosols (mild cooling) and variations in solar activity (mild warming), but with no distinct imprint from major volcanic eruptions. A significant portion of the observed short-term temperature variability can be linked to the influence of NAO. The contributions from SO and AMO are substantially weaker in magnitude. Aside from NAO, no major influence from the explanatory variables was found in the precipitation series. Nonlinear forms of regression were used to test for nonlinear interactions between the predictors and temperature/precipitation; the nonlinearities disclosed were, however, very weak, or not detectable at all. In addition to the outcomes

Electronic supplementary material The online version of this article (doi:10.1007/s10584-014-1147-7) contains supplementary material, which is available to authorized users.

J. Mikšovský (✉) · P. Pišoft
Department of Meteorology and Environment Protection, Faculty of Mathematics and Physics, Charles University, V Holešovičkách 2, 180 00 Prague, Czech Republic
e-mail: jiri.miksovsky@mff.cuni.cz

R. Brázdil
Institute of Geography, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic

R. Brázdil · P. Štěpánek · P. Zahradníček
Global Change Research Centre AS CR, Bělidla 986/4a, 603 00 Brno, Czech Republic

P. Štěpánek · P. Zahradníček
Regional Office Brno, Czech Hydrometeorological Institute, Kroftova 2578/43, 616 67 Brno, Czech Republic

of the attribution analysis for the Czech series, results for European and global land temperatures are also shown and discussed.

1 Introduction

As a part of the effort to improve the scientific understanding of the climate system and its temporal evolution, substantial attention has been paid to the issue of attribution, i.e. identification of the agents responsible for the variations observed. This problem is often addressed by means of complex numerical simulations, involving various configurations of general circulation models (GCMs). While such an approach benefits from physical consistency of method, the results still carry substantial uncertainty, even with regard to the effects of major climate forcings (e.g. Stocker et al. 2013, and the references therein). Alternatively, a methodology based on statistical analysis may be used, identifying connections between target and explanatory variables, and providing information about their significance and strength. Such statistical techniques do not directly consider the physical reality of interactions within the climate system; on the other hand, they are capable of revealing relationships that are omitted or misrepresented due to necessary simplifications in GCM simulations. Compared to GCMs, statistical models are also computationally less demanding, thus allowing fast recalculation when investigating different aspects of a given problem.

Statistical attribution analysis is often applied to data on a global scale (e.g. Walter et al. 1998; Schönwiese et al. 2010; Muller et al. 2013; Rohde et al. 2013). However, the pan-planetary results cannot simply be downscaled to regional or local levels. Just as the climate dynamics and manifestations of climate change vary with geographical location, local responses to external forcings and major climatic oscillations (also referred to as 'internal forcings' here) may differ substantially from their global equivalents (e.g. Walter and Schönwiese 2002; Staeger et al. 2003; Muller et al. 2013). To date, most studies dealing with statistical attribution analysis have used linear techniques, although several attempts have also been made to venture further and employ nonlinear mappings. Walter et al. (1998) demonstrated the potential of neural networks for improving the fraction of variance explained by various forcing factors in the global temperature signal, as did Pasini et al. (2006). Schönwiese et al. (2010) revisited this issue and confirmed the presence of mild nonlinearity. On the other hand, Brázdil et al. (2012a) identified no major nonlinearity in their preliminary assessment of the influence of the forcing factors on Czech climatic series.

This paper addresses the effects of climate forcings on the temporal variability of the mean annual and monthly series of temperature and precipitation for the Czech Lands (recently, the Czech Republic). Various aspects of attribution in the Czech climate have been studied previously, including the manifestations of volcanic activity (Písek and Brázdil 2006), the North Atlantic Oscillation (e.g. Brázdil et al. 2009) and the Southern Oscillation (Brázdil and Bíl 1998), although generally using data that encompass only a few decades during the modern instrumental era. Here, a multivariate approach to the problem of attribution of temperature/precipitation variability to various forcing factors is employed. The studied data span 145 consecutive years, covering the 1866–2010 period, and they are introduced in Sect. 2. While using multiple linear regression as a primary tool for our analysis, two forms of nonlinear mappings were also applied (described in Sect. 3, together with details of other methods and procedures). The results of the attribution analysis are presented and compared with outcomes for European and global temperatures in Sect. 4. Section 5 discusses the results obtained and gives concluding remarks.

2 Data

Central Europe is among the regions with long histories of instrumental meteorological observations. Several of the records available extend back to the 19th century or even beyond, although consistency in series of this length is often compromised by non-climatological factors and homogenization may be required to obtain signals more suitable for analysis of long-term fluctuations (Brázdil et al. 2012b). Furthermore, even with careful quality control and pre-processing, individual signals may still contain residual errors and disturbances. For this analysis, a more robust signal, in the form of mean areal series of annual and monthly temperature and precipitation in the Czech Lands, is employed, created from homogenized series of anomalies from 10 (temperature) and 14 (precipitation) weather stations (for more details see Brázdil et al. 2012a).

To safeguard against the interference of statistical artefacts resulting from our technique of calculation of the mean areal values, an alternative series characterizing Czech temperature was adopted from the Berkeley Earth dataset by Rohde et al. (2013) (providing a long temperature series directly for the Czech Republic, at http://berkeleyearth.lbl.gov/regions/czech-republic). Additionally, in order to offer a comparison with the results on continental and global scales, the Berkeley Earth temperature series for Europe and for global land surface were used.

For our attribution analysis, we employed explanatory variables previously shown to represent natural forcings (related to solar activity and explosive volcanism), anthropogenic forcings (human-induced changes in the quantities of greenhouse gases—GHGs—and atmospheric aerosols) as well as phases of the major climatic oscillations that potentially influence weather patterns in central Europe (Southern Oscillation—SO, North Atlantic Oscillation—NAO, Atlantic Multidecadal Oscillation—AMO). These predictors, introduced below, are generally considered to be among the major causes of observed climatic patterns and related to their eventual temporal changes, although many questions still remain concerning the exact mechanisms of their involvement, and the precise strength of the resulting effects (e.g. Stocker et al. 2013).

Increasing concentrations of anthropogenic GHGs are recognized as one of the principal elements driving global temperature increase, particularly during the second half of the 20th century (Stocker et al. 2013). Here, the effect of GHGs is approximated by a series of annual means of CO_2-equivalent concentrations (hereinafter CO2EQ), representing aggregated radiative forcing of the Kyoto-protocol-controlled greenhouse gases (Meinshausen et al. 2011, data downloaded from http://www.pik-potsdam.de/~mmalte/rcps/).

Human activity also potentially influences the radiative properties of the atmosphere through the formation of anthropogenic aerosols, particularly of the sulphate aerosols, which have been linked to a distinct negative radiative forcing (e.g. Skeie et al. 2011). In similar fashion to Schönwiese et al. (2010), we approximate the effect of sulphate aerosols by SO_2 emissions. Our predictor series (further referenced as SO2) was derived from data published by Smith et al. (2011) for the European region and extended beyond the year 2005 by the rescaled values reported by Klimont et al. (2013).

Solar radiation is the predominant source of energy for the climate system, and even small variations in its intensity can perturb climate equilibrium. Aside from the most prominent 11-year cycle, other variations may be identified in the past values of solar irradiance, including long-term changes, such as the intermittent increase of solar activity during the 19th century and the first half of the 20th century (e.g. Wang et al. 2005). Here, we employ solar irradiance reconstruction data (SOLAR) after Wang et al. (2005), offset by a constant additive factor to respect the revised irradiance values reported by Kopp and Lean (2011) (data downloaded from http://lasp.colorado.edu/sorce/data/tsi_data.htm).

Major volcanic eruptions have been shown to leave a distinct imprint in the records of climatic variables, quite apparent in global temperature in particular (e.g. Rohde et al. 2013). Here, we considered the effects of explosive volcanism through a series of monthly and annual values of volcanic aerosol optical depth (VOLC) for the northern extra-tropical area, adapted from data by Crowley et al. (2008), available from http://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/climate-forcing and documented in more detail by Crowley and Unterman (2013).

Aside from the above-mentioned external forcing factors, the weather and climate in central Europe are also influenced by various large-scale oscillations in the climate system. While these do not constitute climate forcings in the usual sense, they are potentially responsible for a sizable fraction of observed variance. Three major climatic oscillations were included in our analysis; their effects are represented by their respective scalar indices:

Southern Oscillation (SO) is a dominant atmospheric variability mode in the central Pacific region, but SO-related influences have been reported as extending to the European area as well (e.g. Brázdil and Bíl 1998; Brönnimann et al. 2007). The monthly values of the SO index in the 1866–2012 period were downloaded from http://www.cru.uea.ac.uk/cru/data/soi/ (Ropelewski and Jones 1987).

North Atlantic Oscillation (NAO) is a major climatic mode in the northern Atlantic area, with substantial influence on European weather patterns, particularly during the winter season (e.g. Brázdil et al. 2009). Monthly values of NAO index for the 1825–2012 period were downloaded from http://www.cru.uea.ac.uk/cru/data/nao/ (Jones et al. 1997).

Atlantic Multidecadal Oscillation (AMO) is a sea-surface temperature-related climatic mode with impacts on the weather patterns around the Atlantic area (e.g. Enfield et al. 2001), but also on global temperature (Rohde et al. 2013; Zhou and Tung 2013), marking it as a potential contributor to the climate of central Europe as well. The monthly values of the AMO index for the 1856–2012 period were downloaded from http://www.esrl.noaa.gov/psd/data/timeseries/AMO/ (Enfield et al. 2001).

Mean Czech areal temperature and precipitation series by Brázdil et al. (2012a) date from the years 1800 and 1804 respectively; the Berkeley Earth series for Czech temperature extends even further into the past, albeit burdened with substantial uncertainty. For all signals, however, the numbers of source weather stations in the earlier parts of the 19th century are quite low. Further considering the limited availability and/or reliability of certain explanatory variables during the earlier periods of the instrumental era (particularly SO and AMO indices), we restricted our analysis to the 1866–2010 period. See Sect. E1 of the Electronic supplement to this paper for more details on the construction of the series of Czech temperature and precipitation as well as visualization of predictands and predictors.

3 Methods

Prior to the analysis itself, the predictands were converted to additive anomalies with respect to the 1951–1980 period (including the removal of the annual cycle from the monthly data), akin to the Berkeley Earth series. The analysis of Czech temperature was carried out for both the Brázdil et al. (2012a) and Berkeley Earth (Rohde et al. 2013) versions. Due to high similarity of the two series (mutual correlations of 0.99 for annual anomalies and 0.975 for monthly anomalies) and strong resemblance of the outcomes of the attribution analysis, results are presented for only the Brázdil et al. (2012a) series.

While basic correlation analysis can highlight the dominant sources of variability in the data (such as the strong relation between temperature and CO2EQ), it is generally unable to reliably reveal comparatively weaker, albeit statistically significant, links in the complex interactions of multiple variables. For this purpose, a multivariate approach is required, able to separate the contributions from various explanatory variables. Multiple linear regression (MLR) is perhaps the most common of such techniques. We apply it in its usual form here, with regression coefficients calculated to minimize the sum of the squared residuals (i.e. differences between the actual target values and their MLR-based estimates).

As the climate system is inherently nonlinear, it is reasonable to expect this nonlinearity to be reflected in the records of climatic variables and relations between them. In reality, however, the presence and magnitude of nonlinear links is highly variable in time and space and often insignificant from the perspective of time series analysis (e.g. Mikšovský et al. 2008). In order to test for the presence of nonlinear links between target and explanatory variables, we performed a regression analysis with two forms of nonlinear mapping: local linear models (LLMs) and neural networks in the form of multilayer perceptron (MLP). While both these techniques are capable of capturing connections beyond the direct proportionality considered by MLR, they differ in their mode of constructing and calibrating the respective transfer function. LLMs, described in more detail e.g. by Mikšovský et al. (2008), use individual linear mappings for smaller, potentially overlapping segments of the space of predictors, utilizing the fact that even complex nonlinear relations can often be linearized for small groups of similar cases. MLPs, on the other hand, are constructed as global mappings between the predictors and predictand(s), and can serve as universal approximators of nonlinear functions (e.g. Haykin 1999). Neither LLMs nor MLP assume a specific form of nonlinearity in the examined system, and they are thus suitable for the detection of a wide range of deviations from linear behaviour in the data.

By comparing the performance (quantified by root mean squared error, RMSE) of the regression mappings in the form of LLMs or MLP to MLR, the improvement achieved by application of nonlinear mappings was measured. The tests were carried out on data divided into calibration and validation parts, for a wide range of parameters defining the nonlinear mappings (such as the size of the local neighbourhood for LLMs, or the architecture descriptors for MLP). Even in the most favourable cases, the improvement from application of nonlinear techniques did not exceed a few per cent of MLR-based RMSE; for most combinations of target variable and time scale, the gain was undetectable. Considering the miniscule benefits deriving from nonlinear techniques in our analysis, combined with their greater demands regarding parameter optimization, higher computational requirements and more complicated interpretation of the outcomes, only the results from MLR and other linear techniques have been included in this paper.

Statistical significance of the regression coefficients was assessed by means of bootstrapping (with moving-block modification employed for the monthly data) and it pertains to a 95 % confidence level. Both annual and monthly series were investigated; the explanatory series available in annual time-steps (CO2EQ, SO2, SOLAR) were used without temporal interpolation in the monthly-based analysis, i.e. the annual value was considered representative of all months of the respective year.

4 Attribution analysis results

Figure 1a summarizes the normalized regression coefficients for the annual temperature in the Czech Lands. The strong association between temperature and concentration of greenhouse gases is apparent and statistically significant at a high level, with each unit of ppm increasing

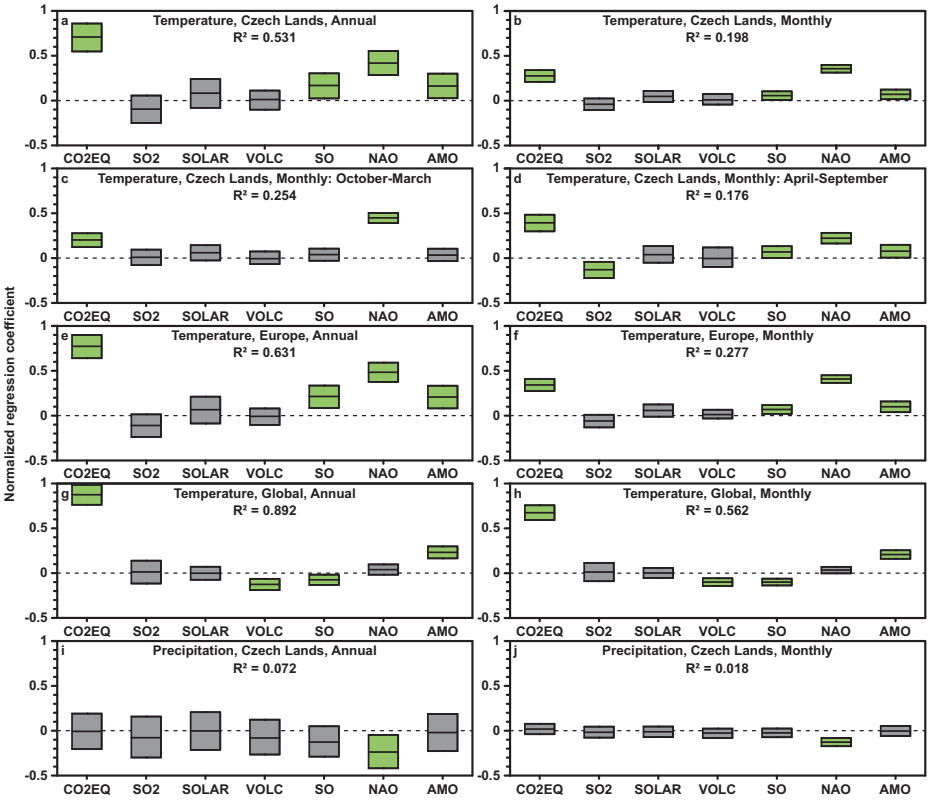


Fig. 1 Normalized regression coefficients (i.e. regression coefficients obtained for both the target and explanatory variables normalized to zero mean and standard deviation equal to one) between the temperature/precipitation series (predictand) and series representing various forcing factors (predictors). The *middle line* marks the value of the regression coefficient, the *box* represents its 95 % confidence interval; coefficients significantly different from zero are highlighted in *green*

the temperature by about 0.014 °C. The presence of sulphate aerosols contributes a mild cooling, but the effect is statistically insignificant. Increase in solar irradiance generates a minor temperature rise, also insignificant. The imprint of volcanic activity in the temperature signal is negligible. There is a relationship between the temperature signal and all three oscillation indices. A strong link, statistically significant, was detected for NAO. The components associated with SO and AMO were weaker, with positive regression coefficients at the edge of statistical significance. Contributions from all seven explanatory variables account for about 53 % of total variance in the Czech annual temperature signal. The analysis of monthly series revealed a similar pattern of relative importance for the predictors (Fig. 1b), with a somewhat lower relative contribution from CO2EQ, which may be ascribed to additional variability in the monthly series compared to the annual scale. With R^2=0.20, the regression model was able to explain a substantially lower fraction of total variance. The results for the European temperatures were quite similar to those for the Czech Lands (Fig. 1e and f). This follows from the similarity of Czech and European temperatures at both annual (mutual Pearson correlation of 0.93) and monthly (0.88) scales.

Compared to the Czech and European temperature signals, sources of variability are somewhat different for the global land temperature (Fig. 1g and h). The role of CO2EQ

becomes even more dominant, although the respective absolute temperature increase drops to about 0.009 °C per ppm. The contributions from sulphate aerosols (approximated in this case by the global SO_2 emissions after Smith et al. 2011) and solar activity are negligible. In contrast to the inconclusive effect of major volcanic events on the Czech/European temperatures, their imprint in the global temperature signal is prominent and statistically significant. The effects of SO and AMO are also well defined; on the other hand, the contribution from NAO is marginal. As the worldwide spatial averaging suppresses regional effects and reduces inter-annual variability in the series, the fraction of variance explained by the regression model rises to 89 % for annual values and 56 % for monthly values.

Most of the predictors considered do not exhibit any statistically significant relation to the series of Czech precipitation, either annual or monthly. NAO phase was the only factor with a significant contribution (Fig. 1i and j), confirming the tendency towards anticorrelation between NAO index and precipitation totals. Even so, the fraction of variance explained by the model is a mere 7 % for annual totals and 2 % for monthly totals.

No major, stable unexplained component was identified in the temperature or precipitation regression residuals (for more details, see Sect. E3 of the Electronic supplement).

4.1 Seasonal variations

A substantial contrast between weather patterns during the different phases of the year is typical of the climate of the mid-latitudes; this may then be reflected in seasonalized local responses to the predictors. To assess the seasonal variability of the regression outcomes, analyses of monthly data were carried out separately for the cold (October–March) and warm (April–September) half-years. While the basic character of the links remained the same as in the undivided data, some variations in the prominence of individual explanatory variables emerged. Effects of both GHGs (warming) and sulphate aerosols (cooling) are substantially more pronounced during the months of the warm half-year in the Czech temperature series (Fig. 1c and d), with the contribution from sulphates exceeding the threshold of statistical significance. This can probably be ascribed to a higher impact of factors modifying radiative balance during the period with more intense solar radiation. The effects of solar and volcanic activity remain small and statistically insignificant during both seasons. The influence of NAO is substantially stronger in the cold half-year, while the opposite holds for SO and AMO. For precipitation, NAO phase remains the only relevant predictor during both warm and cold seasons, though its effect is stronger during the summer half-year (not shown).

4.2 Time-delayed responses

Due to the complex nature of the processes involved in signal transmission within the climate system, a response in climatic variables does not necessarily need to be concurrent with the factors inducing it. This especially concerns predictors representing forcings with faster variability and projecting teleconnections over longer distances, such as SO (e.g. Tsonis et al. 2005). To reveal the possible delays in the response mechanisms, relations between the time-offset values of individual explanatory variables and their respective partial residuals were analyzed, i.e. residuals from regression mappings involving all predictors except for the one under investigation. Figure 2 shows the results for Czech monthly temperature and the four predictors available on a monthly scale: VOLC, SO, NAO and AMO. The correlations for volcanic activity index showed no distinct extremum, and the values remained well within the range consistent with the hypothesis of independent processes. For SO and AMO, a local maximum was detected at time-lag $\Delta t=0$, though just barely above the statistical significance threshold,

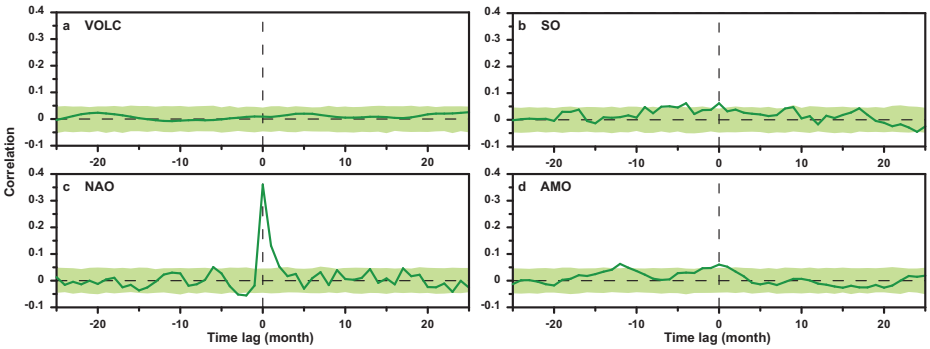


Fig. 2 Time-lagged Pearson correlation coefficients between partial regression residuals and selected explanatory variables, for Czech monthly temperature. The *shaded area* represents correlation values consistent with the null hypothesis of no linear relation between the signals (95% confidence level)

and accompanied by secondary maxima of comparable magnitude. The only variable with distinctly dominant and statistically significant time-delayed correlation maximum at $\Delta t=0$ was NAO. The experiments with time-delayed predictors were also conducted for the rest of the target variables, but no significant, unambiguous delayed responses were detected. All attribution analyses were therefore carried out on series without any time shift.

5 Discussion and conclusions

Our analysis highlighted the influences of several external and internal climatic forcings in the series of Czech temperature and precipitation. Since the normalized coefficients in Fig. 1 do not directly demonstrate the magnitude of individual contributions, the core results are summarized in Fig. 3 in the form of responses associated with selected representative variations of the explanatory variables.

Regardless of temporal or spatial scale, the concentration of GHGs proved to be the key temperature predictor, shaping the long-term variability of the series and providing most of the observed warming. This dominance is not surprising, considering the role of GHGs in establishing the radiative balance of the climate system.

As revealed by the conclusions of previous research (e.g. Skeie et al. 2011, and references therein), the presence of sulphate aerosols generally contributes to negative radiative forcing, thus cooling the affected area. Indeed, our results suggest a negative contribution to Czech and European temperature signals from the SO_2 emissions, although the link is quite weak and largely statistically insignificant. This ambiguity should not be surprising, as a single, emission-based scalar series may be insufficient to represent properly the complexity of aerosol-related mechanisms in the atmosphere. This may also be the reason for the lack of sulphate-related component in the global temperature series.

The variations of solar irradiation do not seem to be linked to any major component in the temperature series. Even so, the warming effect from an increased amount of incoming solar energy is noticeable in the Czech and European series, in contrast to the global land temperature signal, which shows almost no solar-related component (as demonstrated previously by Rohde et al. 2013).

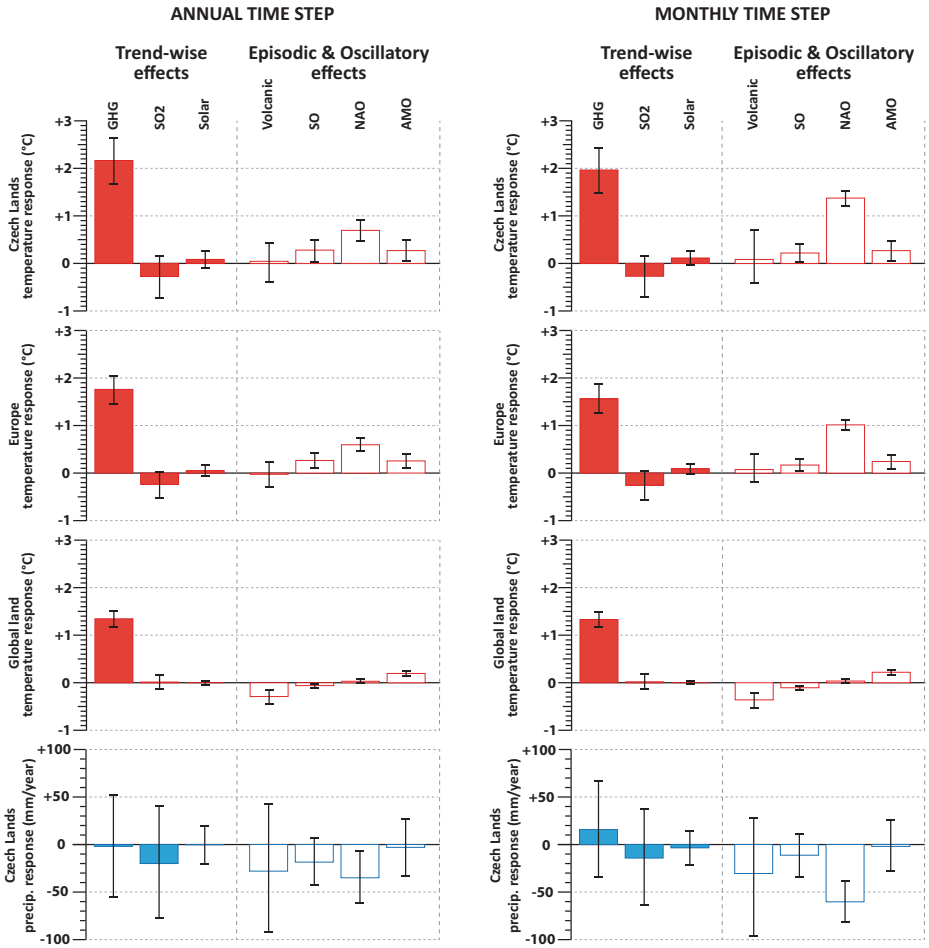


Fig. 3 Response of annual and monthly Czech, European and global land temperature and Czech precipitation to selected variations in the explanatory variables: change of the CO_2-equivalent concentration of GHGs between 1866 and 2010 (+152 ppm); peak value of SO_2 emissions during the 1866–2010 period (41 and 131 Tg SO_2 per year for Czech/European and global data, respectively); solar irradiance change between the periods 1866–1876 and 2000–2010 (+0.46 $W.m^{-2}$); Mt. Pinatubo-sized volcanic event; change of SO, NAO and AMO indices by double the value of their standard deviation. The *error bars* illustrate the 95 % confidence intervals of the response, based on the uncertainty of the respective regression coefficients

No systematic imprint of major volcanic eruptions was revealed in Czech or European temperatures. This finding is consistent with the results of earlier analyses targeting the manifestations of large volcanic eruptions in central Europe (Písek and Brázdil 2006). The global temperature, on the other hand, carries a clear pattern of distinct cooling following prominent volcanic events, as has previously been shown (e.g. Rohde et al. 2013). The reason for the reduced signature of volcanism in our Europe-based signals may be related to the generally higher unexplained variability of the series, masking the traces of individual volcanic events.

Of the internal climatic oscillations investigated, NAO is by far the most influential for both Czech and European temperatures. This reflects the well-established effects of North Atlantic

circulation patterns on (central) European weather (e.g. Trigo et al. 2002; Brázdil et al. 2009). The NAO influence is more pronounced during the cold half-year in which climatic patterns are predominantly influenced by strongly expressed circulation conditions compared to warm half-year with stronger radiative effects. The SO (or, more generally, ENSO), a relatively prominent contributor to the global temperature signal (as shown by numerous prior studies, such as Tsonis et al. 2005), projects comparatively less explicit influence into the European area. Even so, the SO-related component in the Czech temperature series was borderline statistically significant, in line with past studies highlighting the possible teleconnections (e.g. Brázdil and Bíl 1998; Brönnimann et al. 2007). Similar behaviour was also found for the AMO index: the relatively strong response in global land temperature (documented by Rohde et al. 2013) was less prominent, although still statistically significant, in the Czech/European data (though it should be noted that its strength may be subject to the method of AMO index detrending, which can potentially also affect detection of volcanic forcing – see Canty et al. 2013). Due to its faster temporal variability, NAO influence is comparatively stronger on the monthly than annual time scale. Effects from SO and AMO, dominated by lower frequency components, do exhibit only minor differences between time scales, as do contributions from CO2EQ, SO2, SOLAR and VOLC.

One of the limitations of our methodology stems from the application of linear mappings. As purely linear transfer functions cannot incorporate relations beyond simple proportionality, part of the deterministic relations between the analyzed series may remain unresolved. By employing two representatives of nonlinear regression, we hoped to mitigate this deficiency, following the prior studies exposing the presence of nonlinearities in analysis setups similar to ours (Walter et al. 1998; Pasini et al. 2006; Schönwiese et al. 2010). Alas, given the same set of predictors, none of the employed nonlinear transfer functions was able to provide significantly better approximation of the target variables than simple linear regression. While this certainly cannot be interpreted as a proof of an exclusively linear nature for the links studied, it suggests that nonlinearity may be too weak from the practical perspective of multivariate attribution analysis, at least for the specific combinations of series analysed here. Hence, in the light of our results and those of prior studies, the question of the presence, detectability and magnitude of nonlinearities in the responses to climate forcings remains incompletely answered. Additional tests, specifically tailored to reveal particular forms of nonlinearity associated with individual variables or their interactions, and dealing with effects potentially hampering the performance of nonlinear mappings (such as presence of inhomogeneities in long climatic records or vulnerability to low signal-to-noise ratios), may bring more definite conclusions.

While interpreting our results, it is necessary to take into account the limitations of the statistical approach to attribution analysis. Firstly, both predictor and predictand series come with a certain degree of uncertainty, not directly factored into the calculation of the results. It should also be considered that in the case of target variables averaged over larger areas (European and global temperature), different sub-regions may contribute mutually opposite responses, obscuring the resulting effect. Additional variations of the results may be related to the choice of predictor series (several of our explanatory variables represent reconstructions rather than direct measurements, and alternative datasets exist for some of them). Selected tests were therefore repeated with different variables characterizing solar irradiance and volcanic activity (using predictors employed by Schönwiese et al. (2010) as well as some of the series discussed in Schmidt et al. (2012) and its prior version). Only minor changes to the character and magnitude of the effects of the respective forcings resulted from such modifications. We also employed alternative versions of the NAO index (obtained from https://www.climatedataguide.ucar.edu/climate-data; Hurrell 1995), SO index (http://www.cgd.ucar.edu/cas/catalog/climind/soi.html; Trenberth 1984) and AMO index (http://climexp.knmi.nl/data/

iamo_ersst_ts.dat; Trenberth and Shea 2006), and although certain differences in the strength of the respective contributions emerged, their nature seems to be largely unaffected by the choice of the index variant. Additional tests were carried out separately for the first and second half of the 1866–2010 period as well, to ascertain the temporal invariance of the predictor-predictand links. While the information content in our dataset proved insufficient for a reliable sub-period analysis of components dominated by long-term variability, the outcomes suggest stability of the statistically significant contributions from faster variable predictors (SO and NAO).

Finally, it should be emphasized once again that regardless of the form of the transfer function employed, the statistical approach to attribution analysis only considers the formal similarities among the signals, unaware of the physical reality behind them. In this regard, use of GCM-based simulations represents a more sophisticated and less restricted alternative, despite its practical challenges. In the future continuation of our analysis, we therefore intend to combine our findings with conclusions based on long-term historical runs of GCM/RCM couples, currently in production.

Acknowledgments This study was supported by the Czech Science Foundation (GA ČR), through grant P209/11/0956. We would also like to express our gratitude to the authors and providers of all the datasets used. Tony Long (Svinošice) helped work up the English. Finally, we want to thank the three anonymous reviewers for their helpful comments on the manuscript.

References

Brázdil R, Bíl M (1998) Jev El Niño-Jižní Oscilace a jeho možné projevy v polích tlaku vzduchu, teploty vzduchu a srážek v Evropě ve 20. století. Geografie 103(2):65–87

Brázdil R, Chromá K, Dobrovolný P, Tolasz R (2009) Climate fluctuations in the Czech Republic during the period 1961–2005. Int J Climatol 29(2):223–242

Brázdil R, Bělínová M, Dobrovolný P et al (2012a) Temperature and precipitation fluctuations in the Czech Lands during the instrumental period. Masaryk University, Brno

Brázdil R, Zahradníček P, Pišoft P et al (2012b) Temperature and precipitation fluctuations in the Czech Republic during the period of instrumental measurements. Theor Appl Climatol 110(1–2):17–34

Brönnimann S, Xoplaki E, Casty C et al (2007) ENSO influence on Europe during the last centuries. Clim Dyn 28(2–3):181–197

Canty T, Mascioli NR, Smarte MD, Salawitch RJ (2013) An empirical model of global climate – Part 1: a critical evaluation of volcanic cooling. Atmos Chem Phys 13:3997–4031

Crowley TJ, Unterman MB (2013) Technical details concerning development of a 1200 yr proxy index for global volcanism. Earth Syst Sci Data 5:187–197

Crowley TJ, Zielinski G, Vinther B et al (2008) Volcanism and the Little Ice Age. PAGES News 16(2):22–23

Enfield DB, Mestas-Nuñez AM, Trimble PJ (2001) The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. Geoph Res Lett 28(10):2077–2080

Haykin S (1999) Neural networks: a comprehensive foundation (2nd ed). Prentice Hall, Upper Saddle River

Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. Science 269:676–679

Jones PD, Jonsson T, Wheeler D (1997) Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. Int J Climatol 17(13):1433–1450

Klimont Z, Smith SJ, Cofala J (2013) The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions. Environ Res Lett 8(1):014003

Kopp G, Lean JL (2011) A new, lower value of total solar irradiance: evidence and climate significance. Geophys Res Lett 38(1):L01706

Meinshausen M, Smith S, Calvin K et al (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Clim Change 109(1–2):213–241

Mikšovský J, Pišoft P, Raidl A (2008) Global patterns of nonlinearity in real and GCM-simulated atmospheric data. Lect Notes Earth Sci 112:17–34

Muller RA, Curry J, Groom D et al (2013) Decadal variations in the global atmospheric land temperatures. J Geophys Res Atmos 118(11):5280–5286

Pasini A, Lore M, Ameli F (2006) Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system. Ecol Model 191(1):58–67
Písek J, Brázdil R (2006) Responses of large volcanic eruptions in the instrumental and documentary climatic data over central Europe. Int J Climatol 26(4):439–459
Rohde R, Muller RA, Jacobsen R et al (2013) A new estimate of the average earth surface land temperature spanning 1753 to 2011. Geoinfor Geostat An Overview 1:1
Ropelewski CF, Jones PD (1987) An extension of the Tahiti-Darwin Southern Oscillation Index. Mon Weather Rev 115(9):2161–2165
Schmidt GA, Jungclaus JH, Ammann CM et al (2012) Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1). Geosci Model Dev 5:185–191
Schönwiese CD, Walter A, Brinckmann S (2010) Statistical assessments of anthropogenic and natural global climate forcing. An update. Meteorol Z 19(1):3–10
Skeie RB, Berntsen TK, Myhre G et al (2011) Anthropogenic radiative forcing time series from pre-industrial times until 2010. Atmos Chem Phys 11(22):11827–11857
Smith SJ, van Aardenne J, Klimont Z et al (2011) Anthropogenic sulfur dioxide emissions: 1850–2005. Atmos Chem Phys 11(3):1101–1116
Staeger T, Grieser J, Schönwiese CD (2003) Statistical separation of observed global and European climate data into natural and anthropogenic signals. Clim Res 24(1):3–13
Stocker TF, Qin D, Plattner GK et al (eds) (2013) Climate change 2013: the physical science basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge
Trenberth KE (1984) Signal versus noise in the Southern Oscillation. Mon Weather Rev 112:326–332
Trenberth KE, Shea DJ (2006) Atlantic hurricanes and natural variability in 2005. Geoph Res Lett 33:L12704
Trigo RM, Osborn TJ, Corte-Real JM (2002) The North Atlantic Oscillation influence on Europe: climate impacts and associated physical mechanisms. Clim Res 20(1):9–17
Tsonis AA, Elsner JB, Hunt AG, Jagger TH (2005) Unfolding the relation between global temperature and ENSO. Geoph Res Lett 32(9):L09701
Walter A, Schönwiese CD (2002) Attribution and detection of anthropogenic climate change using a backpropagation neural network. Meteorol Z 11(5):335–343
Walter A, Denhard M, Schönwiese CD (1998) Simulation of global and hemispheric temperature variations and signal detection studies using neural networks. Meteorol Z 7(4):171–180
Wang YM, Lean JL, Sheeley NR (2005) Modeling the sun's magnetic field and irradiance since 1713. Astrophys J 625(1):522–538
Zhou J, Tung KK (2013) Deducing multidecadal anthropogenic global warming trends using multiple regression analysis. J Atmos Sci 70:3–8

APPENDIX VII

BRÁZDIL, R., M. TRNKA, J. MIKŠOVSKÝ, L. ŘEZNÍČKOVÁ, AND P. DOBROVOLNÝ (2015B), Spring-summer droughts in the Czech Land in 1805-2012 and their forcings, *International Journal of Climatology*, *35*, 1405-1421, doi:10.1002/joc.4065.

INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **35**: 1405–1421 (2015)
Published online 11 June 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4065

RMetS
Royal Meteorological Society

Spring-summer droughts in the Czech Land in 1805–2012 and their forcings

Rudolf Brázdil,[a,b]* Miroslav Trnka,[b,c] Jiří Mikšovský,[b,d] Ladislava Řezníčková[a,b] and Petr Dobrovolný[a,b]

[a] *Institute of Geography, Masaryk University, Brno, Czech Republic*
[b] *Global Change Research Centre, Czech Academy of Sciences, Brno, Czech Republic*
[c] *Department of Agrosystems and Bioclimatology, Mendel University, Brno, Czech Republic*
[d] *Department of Meteorology and Environment Protection, Faculty of Mathematics and Physics, Charles University, Praha, Czech Republic*

ABSTRACT: Drought is an extreme meteorological phenomenon involving serious economic consequences. In the Czech Lands, it is reflected in significant reductions in agricultural productivity, lack of water for hygiene and industry, and impacts of forest management. Mean monthly temperature and precipitation series created for the Czech Lands for 1805–2012 were used to calculate spring (MAM) and summer (JJA) drought indices (SPI-1, SPI-12, SPEI-1, SPEI-12, Z-index and PDSI), which were then used for further analyses. Fluctuations in drought indices demonstrate an increasing long-term dryness in the Czech climate, statistically significant for SPEI-12 and PDSI in MAM and JJA (in MAM as well for SPEI-1 and Z-index). A significant concentration of drought episodes before 1880 may be attributed to a lack of precipitation, whereas the droughts of recent decades (particularly 2004–2012) are more strongly related to high temperatures. The effects of droughts are reflected in significant reductions in winter wheat and spring barley yields in the eastern province of Moravia. Regression analysis of drought forcings discloses the importance of the North Atlantic Oscillation phase and the aggregate effect of anthropogenic forcing (driven largely by increases in CO_2 concentration). Their magnitude of influence varies strongly with the type of drought index and season of the year. Other factors, such as solar irradiation and the Southern Oscillation phase make only minor contributions to drought variability. The effects of volcanic activity and the Atlantic Multidecadal Oscillation are even weaker and statistically insignificant.

KEY WORDS drought; drought indices; fluctuation; trend; forcing; Czech Lands

Received 17 March 2014; Revised 7 May 2014; Accepted 8 May 2014

1. Introduction

Drought is a climatological phenomenon that has important impacts on many aspects of human society and affects many of its more important activities. In agriculture, it leads to significantly smaller yields of agricultural crops than in normal years (e.g. Hlavinka *et al.*, 2009; Kolář *et al.*, 2013). Extended drought dries out commercially growing timber, forcing forestry managers into prematurely felling trees, well before their optimum size is reached (Brázdil, 1998). Low water levels in rivers and reduced groundwater reserves lead to problems in the management of water resources, usually for prolonged periods of time (e.g. Wilhite and Vanyarkho, 2000). Furthermore, droughts influence not only human society, but also lead to deterioration of natural ecosystem functions (Ciais *et al.*, 2005). All these criteria highlight the importance of drought research in Europe for the future, especially in the light of anthropogenically exacerbated climate change (Rowell and Jones, 2006) and also the need for proper and scientifically sound advance planning of reactions to drought (Wilhite *et al.*, 2007).

Drought may be defined as a negative deviation of water balance from the climatological norm over a given area. This implies that drought is a result of deficiency in precipitation over an extended period of time, whereas other meteorological elements (such as increased air temperature, global radiation and wind, as well as decreased air humidity) drive up water demand through increased evapotranspiration (e.g. Allen *et al.*, 1998), frequently intensifying the impact of precipitation deficit (e.g. Zahradníček P, 2014 (pers. comm.)). However, according to Lloyd-Hughes (2013), a universal description of drought requires reference to water supply, demand and management. With respect to various associated difficulties, he concludes that a workable generalized and objective definition of drought does not exist.

On the basis of timescales and impacts, droughts may be divided into four categories: meteorological, agricultural, hydrological and socio-economic (Heim, 2002; Dai, 2011b). To these, Mishra and Singh (2010) add groundwater drought. Meteorological drought is signalled by indicators intrinsic to weather data and precedes the onset of specific impacts, i.e. additional types of drought.

* Correspondence to: R. Brázdil, Institute of Geography, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic. E-mail: brazdil@sci.muni.cz

Agricultural drought may be measured in terms of duration in weeks to 6–9 months, whereas hydrological, groundwater and socio-economic impacts usually became apparent over longer time intervals. The socio-economic importance of droughts is reflected in the great number of studies devoted to this phenomenon on an international scale. Noteworthy among these are a global assessment by Wilhite *et al.* (2000), while the socio-economic benefits of drought preparedness are clearly summarized in Wilhite *et al.* (2007). This contribution concentrates on meteorological drought, and partly on agricultural drought, with SPI representing meteorological drought and Z-index and 1-month SPEI more representative of short-term agricultural drought and PDSI with 12-month SPEI representing the long-term drought events that impact primarily upon agriculture (and probably upon forestry and water sources as well).

Droughts may be compared with floods as the most serious hydrometeorological extremes in the Czech Republic (Brázdil and Kirchner, 2007). Particular attention to the study of temporal and spatial aspects of drought in the Czech Republic burgeoned after the year 2000. Možný (2004) investigated the intensity of droughts between 1891 and 2003. Blinka (2005) contributed a similar paper, providing a climatological analysis of droughts from 1876 to 2002. Dufková and Št'astná (2005) studied droughts in South Moravia and their influence on soil erosion. The most comprehensive analysis of droughts in Moravia, with particular attention to the 1961–2000 period, appeared in *A Climatic Atlas of Czechia* (Tolasz *et al.*, 2007) and *Selected Natural Extremes and Their Impacts in Moravia and Silesia* by Brázdil and Kirchner (2007), followed by a detailed analysis from Trnka *et al.* (2009). Brázdil *et al.* (2009b) concentrated on the variability of drought indices in the period 1881–2006 calculated from mean Czech temperature and precipitation series. Dubrovský *et al.* (2009) developed and applied the relative Palmer Drought Severity Index (rPDSI) and the Palmer Z-index to estimate the effect of changed climate on drought frequency and established that the risk is likely to increase sharply in the future as a result of projected rising temperatures. Potop *et al.* (2011) used the Standardized Precipitation Evapotranspiration Index (SPEI) to study drought variability as recorded at five lowland stations (1901–2010). Treml (2011) analysed the most severe drought events in the Czech Republic in the 1875–2010 period. Most recently, Brázdil *et al.* (2013) developed a long-term Czech drought chronology from AD 1500 onwards, combining information from documentary data with instrumental records. Drought trends between 1961 and 2012 were analysed by Trnka M, 2014a, 2014b (pers. comm.), using a high-resolution grid database. Zahradníček P, 2014 (pers. comm.) studied an extreme drought occurring between August 2011 and May 2012 with respect to its impacts on agriculture and hydrology. Potop *et al.* (2014) investigated the spatiotemporal characteristics of drought in the Czech Republic using SPEI calculated for 184 stations in the 1961–2000 period.

The aim of this article is to analyse the long-term variability of droughts in the Czech Lands, based on drought indices calculated from homogeneous long-term instrumental data for spring and summer in the 1805–2012 period. The concentration of analysis in spring and summer follows from the fact that these seasons generally include the vegetation period and the months during which drought (or lack of water) is known to have an effect on the agricultural production (Hlavinka *et al.*, 2009; Trnka *et al.*, 2012) as well as upon tree growth (Brázdil *et al.*, 2002; Büntgen *et al.*, 2011) in the Czech Lands.

This study employs data for the first 75 years of the 19th century not previously available, not only extending the period of instrumental records but also allowing suppositions to be made about possible drought trends, and analysis to be performed for their forcing factors. Basic temperature and precipitation series and drought indices calculated from these data appear in Section 2. Section 3 outlines the methods used in this study. Section 4 presents the results of the analysis of long-term fluctuations of drought and changes in their recurrence interval; provides an attribution analysis of their forcings, and describes the effects of droughts on crop yields. The overall results are discussed in Section 5 and some conclusions drawn in Section 6.

2. Data

2.1. Temperature and precipitation series

This study of droughts in the Czech Lands is based on mean areal temperature and precipitation series calculated by Brázdil *et al.* (2012a, 2012b). These were based on homogenized monthly values recorded at 10 secular stations for temperature and 14 stations for precipitation. Series of individual secular stations were further confronted with areal temperature and precipitation means for the Czech Republic calculated over the 1961–2000 period, for which the station coverage is the best (268 temperature and 878 precipitation series). A quantile mapping method was used to convert original and homogenenous secular series on the statistical properties of 40-year means. On the basis of comparison with the 1961–2000 calibration period, each of the secular series was statistically adjusted for its entire length. The median value of 10 such adjusted temperature series (14 series for precipitation) was then considered as the final mean areal temperature (precipitation) series for the Czech Lands. The series are biased by the number of stations used because all series were available from only 1883 for temperature and 1876 for precipitation. Only two stations (Prague-Klementinum and Brno) were available back to 1818 for temperatures and 1828 for precipitation.

2.2. Drought indices

2.2.1. Standardized Precipitation Index

The assessment of meteorological drought starts with analysis of precipitation. This may be addressed either

by analysing precipitation totals in terms of reliability (e.g. Laughlin *et al.*, 2003) or by using one of the many precipitation-based drought indices that have been developed (e.g. McKee *et al.*, 1993; Byun and Wilhite, 1999). The Standardized Precipitation Index (SPI) is one of the most widely employed drought monitoring indicators, allowing drought evaluation using only monthly or weekly precipitation data. Mathematically, SPI is based on the cumulative probability of a specific rainfall event occurring at a given station (McKee *et al.*, 1993). The monthly (or weekly) sums of precipitation at such a station are usually fitted by means of gamma distribution, which has been found quite appropriate to precipitation distribution across most timescales. The fitted cumulative probability function is then transformed by an inverse normal function. A low or high probability on the cumulative probability function related to a particular rainfall total then indicates the likelihood of dry or wet events, respectively. In summary, SPI can effectively represent the amount of precipitation over a given timescale in relation to the median. This enables the user to state whether a station is experiencing drier than usual conditions relative to the station climatology. The near-optimum range of SPI could be approximated by an interval of ± 0.9 whereas the usual range of SPI values is from -3 to $+3$, with negative values describing periods of precipitation below the median. SPI enables the user to assess the occurrence of short-term (duration of 1 month – SPI-1), medium-term (from three to 12 months – SPI-3 and SPI-12) and long-term droughts (12 months and longer). As SPI does not reflect changes in evaporation or warming, it is a good descriptor of deficiency in precipitation (a key precursor of drought), but not of the changes in other water-balance components. In the light of the latter, more comprehensive indices were further employed.

2.2.2. *Palmer Drought Severity Index and Palmer Z-index*

The Palmer Drought Severity Index (PDSI; Palmer, 1965) is another approach widely used to quantify drought all over the world (e.g. Szinell *et al.*, 1998; Lloyd-Hughes and Saunders, 2002; Ntale and Gan, 2003; Dai *et al.*, 2004; van der Schrier *et al.*, 2006, 2007; Dai, 2011a). In general, this index is based on the supply-and-demand concept of a water-balance equation and thus incorporates antecedent precipitation, moisture supply and demand at the surface as calculated according to the Thornthwaite (1948) potential evapotranspiration (PET) method. It applies a two-layer bucket-type model for soil moisture computations with three assumptions related to soil profile characteristics:

(i) the water-holding capacity of the surface layer is set at a maximum of 25 mm,
(ii) the water-holding capacity of the underlying layer has a maximum value that depends on soil type,
(iii) water transfer into or out of the lower layer only occurs when the surface layer is full or empty. The PDSI itself can be described as an accumulative departure relative to local mean conditions in atmospheric moisture supply and demand at the surface (Palmer, 1965) and it is considered a good representation of episodes of prolonged drought.

PDSI calculation includes an intermediate term known as the Palmer moisture anomaly index (or Z-index), which is a measure of surface moisture anomaly for a given month without the consideration of the antecedent conditions so characteristic of PDSI. It is basically the moisture departure, adjusted by a weighting factor known as the climatic characteristic. The Z-index can be used to track drought events on a monthly basis as it responds relatively quickly to changes in soil moisture (Karl, 1986). The capacity of the Z-index to rank the dryness or wetness of individual months makes it especially useful as one of the indicators of short spells of drought.

The original monthly PDSI relied on empirical constants, soil property assumptions and climate characteristics derived by Palmer (1965) using data from nine stations in Kansas and Iowa (United States). In the study in hand, a self-calibrated version of the Z-index and PDSI (Wells *et al.*, 2004) was used. Wells *et al.* (2004) modified the original Palmer model in order to adjust the previously empirical constants automatically in response to input data uniquely derived from each studied location. The self-calibrated PSDI adjusts itself to produce a range of PDSI values for any location between -4.0 and $+4.0$ with drought represented by negative values and mean PDSI value equal to zero.

2.2.3. *Standardized Precipitation Evapotranspiration Index*

The SPEI was introduced by Vicente-Serrano *et al.* (2010) to improve the original SPI concept. SPEI uses the monthly difference between precipitation and PET, a simple climatic water balance (Thornthwaite, 1948), i.e. the same method as that used in PDSI and Z-index calculations, and can be calculated at different timescales. The PET calculation can be made either through physically based methods (e.g. the Penman–Monteith method, PMM) or models based on empirical relationships, where PET is calculated with fewer data requirements. While PMM should be a preferred choice, it requires data (e.g. solar radiation, wind speed and humidity) not available for some of the sites analysed during the 20th century and completely absent before the 1920s. Mavromatis (2007) showed that the use of simple or complex methods to calculate PET provides similar results when a drought index, such as the PDSI, is calculated, although concerns exist about using PDSI for climate projections (Hoerling *et al.*, 2012). While PMM may be used in SPEI calculations, this contribution follows the original procedure established by Vicente-Serrano *et al.* (2010) based on the Thornthwaite's (1948) method, requiring only mean monthly temperatures. With a value for PET, the difference between the precipitation and PET for the given month is calculated, which provides a simple measure of water surplus or deficit

for it. The differences are aggregated at various timescales, following the same procedure as that for SPI. The average value of SPEI is zero and the standard deviation is one. SPEI is a standardized variable, and it can therefore be compared with other SPEI values over time and space. A SPEI of zero indicates a value corresponding to 50% of the cumulative probability of precipitation and PET difference, according to log-logistic distribution.

2.3. Forcing factors

Considering the complex relations between drought characteristics and a number of climatic variables, it may be of interest to identify the underlying factors responsible for the observed variability. These may include anthropogenic forcings (such as human-induced changes in atmospheric composition) as well as natural ones (variations in solar and volcanic activity), and also the effects of major circulation modes (such as North Atlantic Oscillation, the Southern Oscillation and the Atlantic Multidecadal Oscillation). To carry out an attribution analysis, quantifying the contributions of such external and internal climate forcings to drought indices, the following six quantities are used as potential explanatory variables:

(a) Annual CO_2-equivalent concentrations (hereinafter denoted as CO2) representing aggregated radiative forcing of greenhouse gases (GHGs) and other anthropogenic agents (such as aerosols and tropospheric ozone) in the 1805–2012 period (Meinshausen *et al.*, 2011; http://www.pik-potsdam.de/~mmalte/rcps/).
(b) Annual solar irradiance estimate (SOLAR) in the 1805–2012 period. The data, based on a reconstruction by Wang *et al.* (2005), and offset to respect the revised irradiance values derived by Kopp and Lean (2011), were downloaded from http://lasp.colorado.edu/sorce/data/tsi_data.htm.
(c) Monthly series of volcanic aerosol optical depth (VOLC), adapted for the 1805–2012 period from data by Crowley *et al.* (2008), included in the supplementary material of a paper by Schmidt *et al.* (2012).
(d) Monthly values for the North Atlantic Oscillation Index (NAOI), in the version provided by CRU (http://www.cru.uea.ac.uk/cru/data/nao/; Jones *et al.*, 1997) for the 1825–2012 period, and extended with the rescaled series by Luterbacher *et al.* (1999) for the early part of the 19th century.
(e) Monthly values for the Southern Oscillation Index (SOI) for the 1866–2012 period, obtained from http://www.cgd.ucar.edu/cas/catalog/climind/soi.html (Trenberth, 1984).
(f) Monthly values of the Atlantic Multidecadal Oscillation Index (AMOI) for the 1856–2012 period. The data were downloaded from http://www.esrl.noaa.gov/psd/data/timeseries/AMO/ (Enfield *et al.*, 2001).

The available series of some of these covariates (particularly SOI and AMOI) do not cover the entire period studied. While various reconstructions of related quantities exist for earlier parts of the 19th century, the preliminary tests found none of them to be compatible enough to provide a reliable extension (unlike NAOI, where the match between the index based on instrumental data by Jones *et al.*, 1997 and the reconstruction by Luterbacher *et al.*, 1999 was better). The primary target period for the attribution part of this analysis was therefore set to 1867–2012, starting 1 year after the beginning of the SOI series. Additional tests for the 1806–2012 period were carried out as well, but with a reduced number of explanatory variables.

2.4. Yield levels

To improve the understanding of the relation between drought intensity and drought impacts, we collated a unique dataset of spring barley and winter wheat yields for the south-eastern region of the Czech Lands covering the periods 1869–1913 and 1961–2012. This is the longest yield record analysed in the region to date. The data for the 1869–1913 period was extracted from annual reports: *Mittheilungen der kaiserlich-königlichen Mährisch-Schlesischen Gesellschaft zur Beförderung des Ackerbaues, der Natur- und Landeskunde* ['Reports of the Imperial-Royal Moravian-Silesian Society for the Improvement of Agriculture, Natural Science and Regional Geography']. Despite considerable effort, no yield data for the period prior to 1869 could be found, and this year appears, to the best of our knowledge, to be when official and systematic recording of agricultural statistics in the region started. The years 1881, 1890 and 1891 lack yield reports. Organizational and/or financial matters intervened and either the yield statistics for these years were not collected, or simply not published. Very little is available for the period between 1913 and 1960, with only 8 years in the late 1920s and early 1930s giving district yield data. There is nothing for the periods covering the two world wars and for a period of 15 years after World War II. Data for the 1961–2012 period were compiled from mean annual yields of winter wheat and spring barley at district level using the archives of the Czech Statistical Office (year 2000), the Ministry of Agriculture (2001–2007) and the Chamber of Agriculture (2008–2012). Original data were extracted, digitized and quantitatively checked.

3. Methods

The temporal variability of six drought indices was investigated separately for March–May (MAM) and June–August (JJA) in the 1805–2012 period. Long-term trends were described by linear regression against time, with statistical significance of the trend determined by t-test at a significance level of $\alpha = 0.05$. Box-plots were used for comparison of basic statistical characteristics in non-overlapping 30-year periods.

The extremeness of droughts as characterized by drought indices was expressed via calculation of the return period for mean N-year recurrence intervals ($N = 2$, 5, 10, 20,

50 and 100 years). First, values of drought indices were modelled with Gaussian normal distribution and the suitability of this theoretical distribution was verified with Shapiro–Wilks goodness-of-fit test. Probability of occurrence (p) was then estimated from Gaussian frequency distribution. Finally, return periods (N) were derived as an inverse of probability estimates, that is, $N = 1/p$ (Wilks, 2006).

For the attribution analysis, various linear and nonlinear regression mappings were tested to provide a transfer function between drought indices and explanatory variables. Multiple linear regression (MLR) was chosen as a primary tool for this task – a stable and robust, although completely linear technique, approximating the target quantity by a weighted mean of the explanatory variables. As even short-term drought characteristics are not necessarily tied to only the concurrent values of predictors, and time-delayed and time-accumulated responses may take place in the (local) climate system, several variants of series with faster temporal variability (VOLC, NAOI, SOI and AMOI) were prepared as non-weighted means of values from a sequence of consecutive months, with starting month ranging from December to June, and length of the averaging period between three and nine months. Using a forward stepwise regression procedure, the version with highest (although not necessarily statistically significant) contribution to the explained variance was identified for each of the VOLC, NAOI, SOI and AMOI series. These series, together with annual CO2 and SOLAR signals, make up the six predictors used as inputs for the regression. The selection procedure was carried out separately for each short-term and medium-term drought index (SPI-1, SPEI-1, Z-index and PDSI), and both target periods. An example of the resulting set of predictors is shown in Figure 1. For long-term drought indices (SPI-12 and SPEI-12), the predictors were averaged from the 15 months preceding the end of the target period. The long-term memory components in the PDSI were first addressed by conducting experiments with predictors derived as weighted means of monthly values, with weights progressively decreasing for more distant months. However, such predictors proved inferior to their stepwise selected alternatives, which were therefore preferred for the analysis.

The climate system is inherently nonlinear and the same can be expected for signals characterizing its behaviour. However, this nonlinearity is not always detectable from climatic series, and its magnitude varies with location, season and the matter investigated (e.g. Mikšovský and Raidl, 2006; Mikšovský *et al.*, 2008, and references therein). To reveal the possible manifestations of nonlinear behaviour in the relations analysed, results obtained by MLR were compared with those obtained by the method of local models (e.g. Mikšovský and Raidl, 2006) and from multilayer perceptron neural networks (e.g. Haykin, 1999 and references therein). Although mild nonlinearity was detected for some combinations of the target drought index and time interval, the improvement from application of nonlinear mappings was slight in most cases. This finding is consistent with Brázdil *et al.* (2012a) and Mikšovský *et al.* (2014), who found no major nonlinearity in the relations of Czech temperature and precipitation to various external and internal forcings. For this reason, and in the light of easier interpretation for the outcomes of linear regression, only MLR-based results are presented in this article.

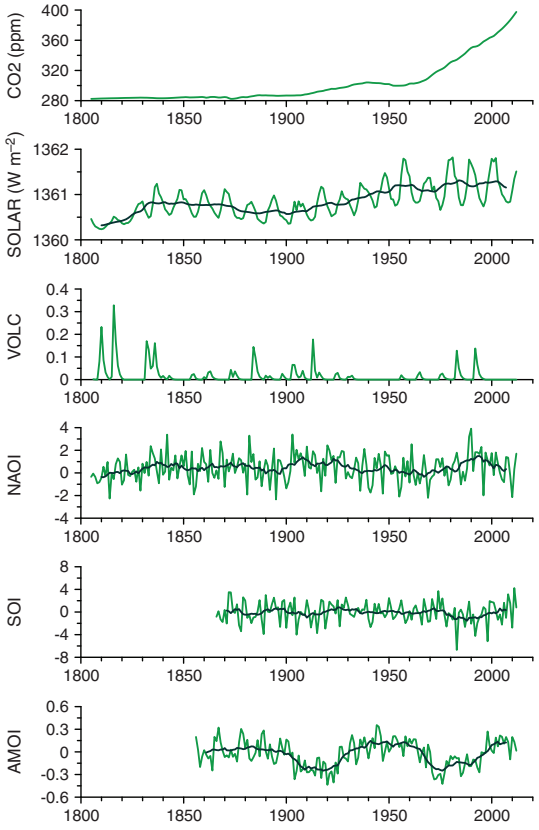


Figure 1. Series of predictors used as explanatory variables for the MAM Z-index; the darker line represents 11-year running means for series with fast, non-episodic variability (see Section 2.3 and Table 3 for predictor details and regression results).

The statistical significance of the results of the attribution analysis was tested through bootstrapping based on ten thousand randomized versions of the dataset. To consider the effect of residual autocorrelations in the MLR outputs (prominent particularly for PDSI), the moving-block variant of bootstrap was employed.

4. Results

4.1. Long-term fluctuations

Fluctuations in the MAM and JJA drought indices of the Czech Lands in the 1805–2012 period appear in Figure 2. Differences in inter-decadal variability between individual indices follow from their character, reflecting only precipitation patterns (SPI) or more complex effects (SPEI, Z-index and PDSI). While SPI indices show no trend from the 1940s to recent times, decreasing trends in other indices (increasing dryness) are clearly expressed,

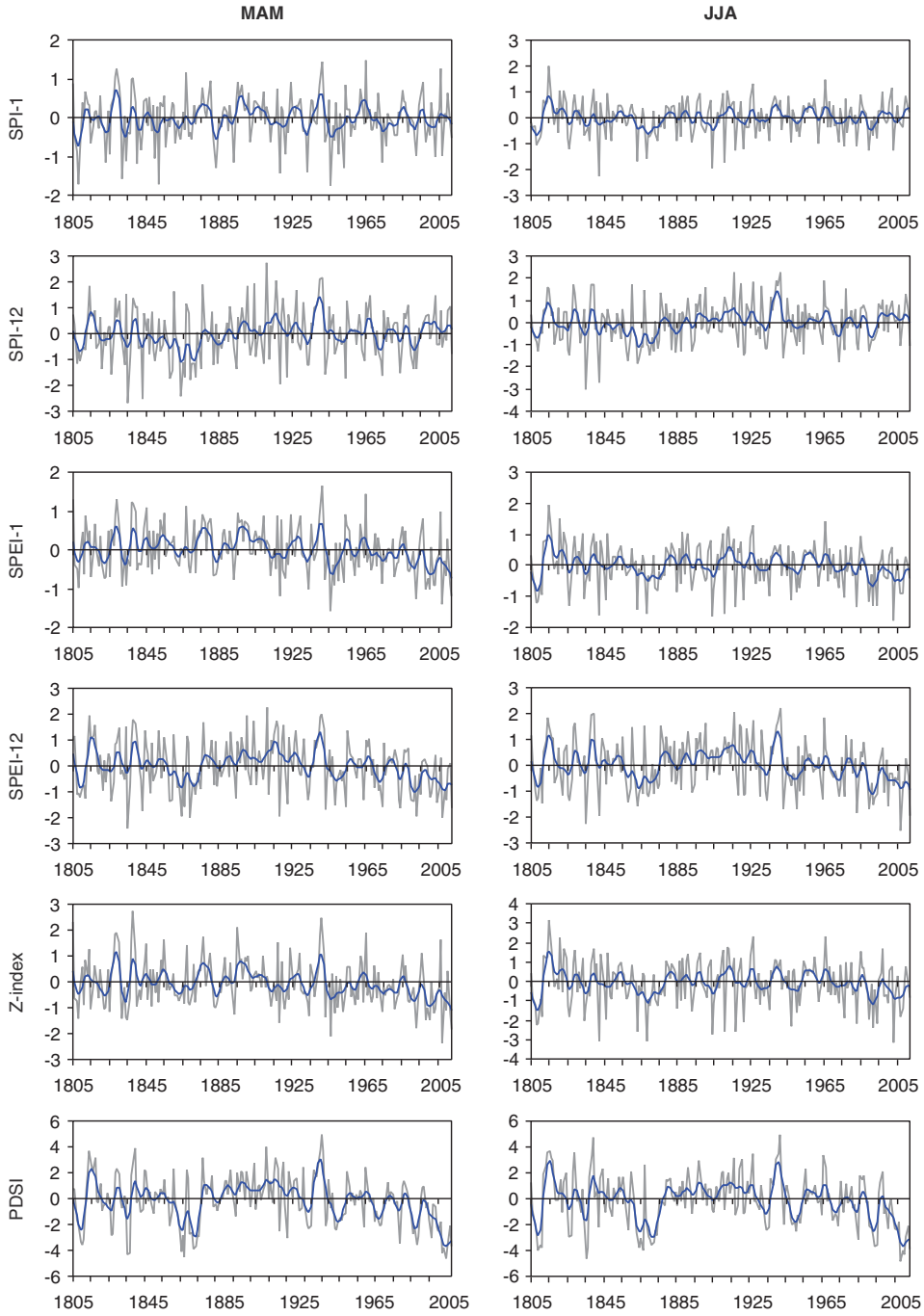


Figure 2. Fluctuations of MAM and JJA drought indices for the Czech Lands in the 1805–2012 period. Series are smoothed by 10-year Gaussian filter.

particularly for PDSI. There are also no important differences between drought indices for MAM and JJA. Positive insignificant linear trends in 1805–2012 are associated with SPI. All other drought indices are characterized by negative linear trends, statistically significant apart from SPEI-1 and Z-index for JJA (Table 1).

The box plots of MAM and JJA drought indices in the 30-year periods from 1805–1834 to 1985–2012 that appear in Figure 3 indicate some important features of drought variability, as well as differences between indices. While SPI indices reflect the deepest droughts in 1835–1864, the four remaining indices clearly signal the driest patterns in 1985–2012, when SPI-12 values, in particular, are positive. Prevailing wet weather patterns in 1895–1924 are confirmed by positive values of all drought indices.

Table 1. Hundred-year linear trends of selected MAM and JJA drought indices (SPI-1, SPI-12, SPEI-1, SPEI-12, Z-index and PDSI) for the Czech Lands in the 1805–2012 period. Statistically significant values, at $\alpha = 0.05$, appear in bold. [Correction added 6 February 2015 after original online publication: the values in Table 1 have been corrected to show hundred-year linear trends.]

Season	SPI-1	SPI-12	SPEI-1	SPEI-12	Z-index	PDSI
MAM	0.01	0.15	**−0.20**	**−0.23**	**−0.29**	**−0.54**
JJA	0.05	0.16	−0.11	**−0.24**	−0.12	**−0.49**

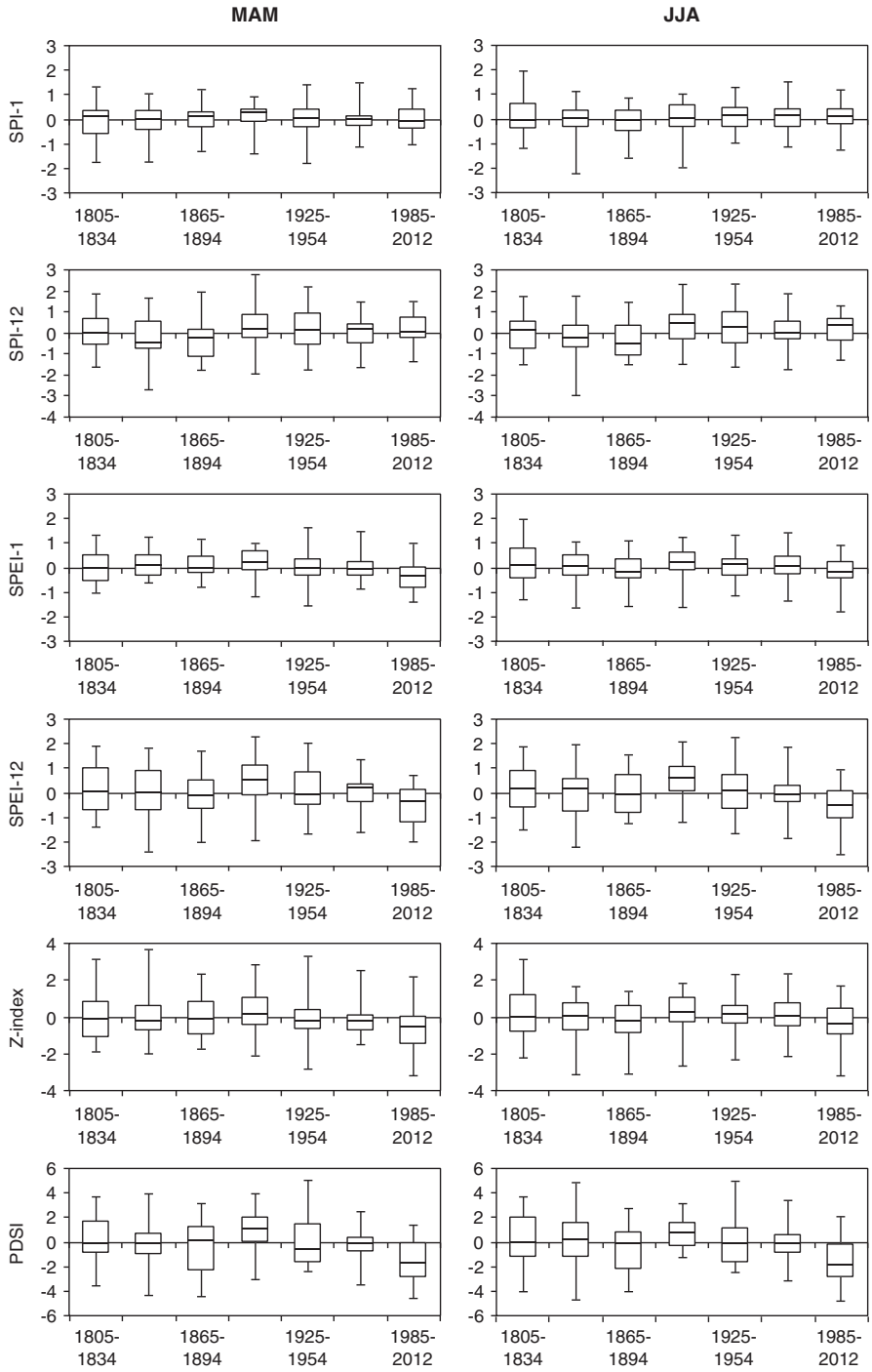


Figure 3. Box-plots (maximum, 75, 50 and 25% quartiles, minimum) of MAM and JJA drought indices for the Czech Lands in 30-year periods for 1805–2012.

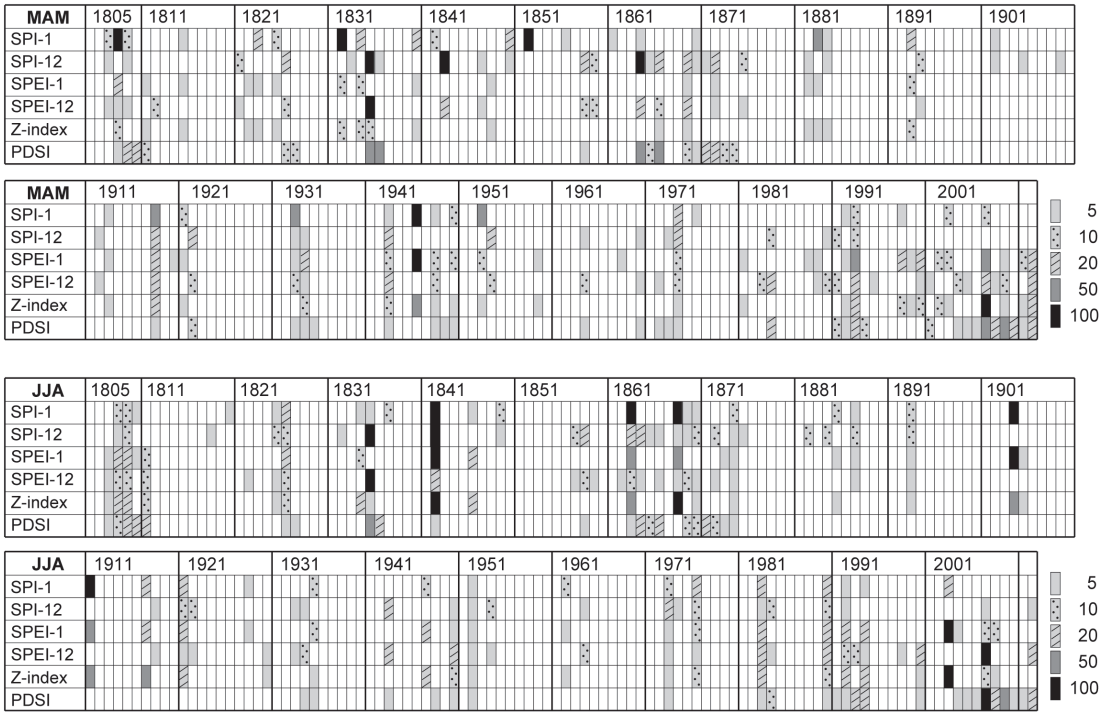


Figure 4. MAM and JJA droughts expressed by SPI-1, SPI-12, SPEI-1, SPEI-12, Z-index and PDSI with a recurrence interval of $N \geq 5$ years in the Czech Lands during the 1805–2012 period.

4.2. Fluctuations and recurrence intervals

Czech series of six drought indices for MAM and JJA in the 1805–2012 period were further analysed with respect to threshold values of N-year recurrence intervals ($N = 2$, 5, 10, 20, 50 and 100). An overview of such indices at a recurrence interval of $N \geq 5$ years appears in Figure 4. The 5-year return period was based on agro-climatological studies (e.g. Hlavinka *et al.*, 2009; Trnka *et al.*, 2012) showing harvests significantly influenced by drought in approximately 20% of cases over the past 130 years. The longest sequence of droughts with $N \geq 5$ years (Figure 4) covers 9 years in 2004–2012 for PDSI, with no close similarity in other indices or in other parts of the period studied. As mentioned above, it appears that drought events around the generally cooler 19th century should be considered as a reflection of lower precipitation totals, while in recent decades the key role is played by increasing temperatures (Brázdil *et al.*, 2009b, 2012a, 2012b; Trnka *et al.*, 2012).

Figure 5 shows the decadal frequencies of MAM and JJA droughts expressed by six selected drought indices at a recurrence interval of $N \geq 5$ years during the 1805–2012 period. There exists a clear difference in dominant drought decades between the indices based only on precipitation (SPI-1 and SPI-12) and the other more complex indices. Spring droughts occurred most frequently in 1861–1870 in terms of SPI-12 and in 1941–1950 in terms of SPI-1 whereas the other four indices indicated the highest frequency in 2001–2012. For PDSI, the 1861–1870 decade was also important, and for SPEI-1 with Z-index 1991–2000 as well. The decadal frequencies of summer droughts presented a more variable situation. While 1861–1870 dominated for SPI-1, SPI-12 and SPEI-12 (PDSI also important), a clear drought prevalence in 2001–2012 appeared only in PDSI. Although SPEI-1 also gave its highest drought frequency in 2001–2012, it indicated the same frequency of droughts in 1805–1810; Z-index did the same, together with the 1805–1810, 1861–1870 and 1991–2000 periods.

The occurrence of decades without drought (Figure 5) is notable mainly for PDSI as an indicator of long-term drought. No drought year with $N \geq 5$ was recorded between 1881 and 1910 for MAM and for as long as 1881–1930 for JJA. For the other indices, this did not occur for longer periods than a decade.

All the above statements may be supplemented by consideration of drought severity in particular years (Figure 4, Table 2). In terms of 100-year drought, the year 1835 was exceptional in that MAM and JJA experienced 100-year (SPI-12 and SPEI-12) or 50-year (PDSI) drought. A similar situation was also observed in 2007. Severe drought in another 6 years was recorded for MAM as well as for JJA, but without any overlap either season. For short-term drought, JJA 1842 and JJA 1868 were notable for their recurrence interval of $N \geq 100$ years.

4.3. Attribution of droughts

Table 3 summarizes the results of attribution analysis in the form of individual regression coefficients between target and explanatory variables, as well as the percentage of variance explained as coefficient of determination (R^2).

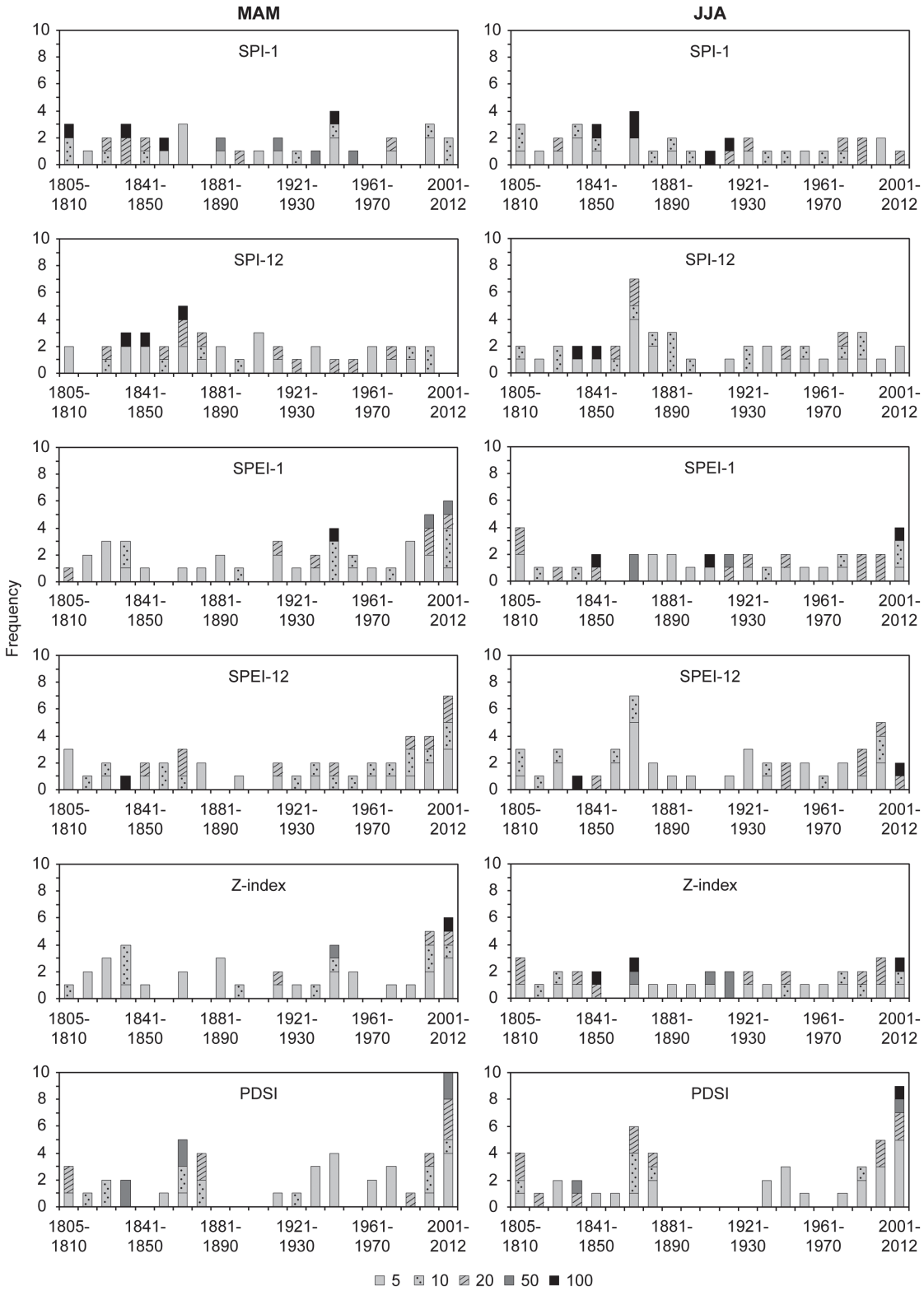


Figure 5. Decadal frequencies of MAM and JJA droughts expressed by SPI-1, SPI-12, SPEI-1, SPEI-12, Z-index and PDSI in order of their N-year recurrence interval ($N = 5$, 10, 20, 50, 100) for the Czech Lands in the 1805–2012 period.

Table 2. Years in which at least one drought index of the six employed – SPI-1, SPI-12, SPEI-1, SPEI-12, Z-index and PDSI – in the Czech Lands exhibited a recurrence interval of $N \geq 100$ years for MAM and JJA in the 1805–2012 period (for selected years, drought indices with $N \geq 50$ years are also added in brackets).

Year	MAM	JJA
1808	SPI-1	–
1832	SPI-1	–
1835	SPI-12, SPEI-12 (PDSI)	SPI-12, SPEI-12 (PDSI)
1842	–	SPI-1, SPEI-1, Z-index, SPI-12
1843	SPI-12	–
1852	SPI-1	–
1863	–	SPI-1 (SPEI-1, Z-index)
1864	SPI-12 (PDSI)	–
1868	–	SPI-1, Z-index (SPEI-1)
1904	–	SPI-1, SPEI-1 (Z-index)
1911	–	SPI-1 (SPEI-1, Z-index)
1946	SPI-1, SPEI-1 (Z-index)	–
2003	–	SPEI-1, Z-index
2007	Z-index (SPEI-1, PDSI)	SPEI-12, PDSI

To provide an easier inter-comparison of the regression outcomes for different drought indices, the regression coefficients are presented for the standardized versions of predictors and predictands, i.e. time series transformed in linear fashion to zero mean and standard deviation equal to one.

Of the explanatory variables considered, the most influential predictor is typically NAO index, the optimum source period of which usually covers the winter months preceding the target period, particularly for the spring drought indices. The NAO-related regression coefficients are generally statistically significant at the $\alpha = 0.05$ level and always negative. This implies a tendency towards more substantial droughts during positive NAO-index episodes, which result in warmer winter weather in Central Europe (e.g. Wanner *et al.*, 2001; Trigo *et al.*, 2002; Cahynová, 2005; Brázdil *et al.*, 2009a, 2012a).

A further influential explanatory variable was identified as CO_2-equivalent anthropogenic forcing, which is largely shaped by increasing concentrations of GHG in the atmosphere. This change is generally considered the main reason for long-term global temperature increase (Stocker *et al.*, 2013), particularly prominent in the second half of the 20th century. Consequently, the drought indices most sensitive to temperature (particularly PDSI and SPEI-12) exhibit a negative trend in this period, anticorrelated with the CO2 series. The corresponding regression coefficients are negative and statistically significant. Precipitation-based drought indices (SPI-1 and SPI-12) have only weak and statistically insignificant relations to the CO2 series, consistent with their lack of a temperature-related component.

The influence of the Southern Oscillation was much weaker than that of NAO, although statistically significant

Table 3. Regression coefficients for standardized series of drought indices (rows) and explanatory variables (columns) in the 1867–2012 period.

Index	Season	CO2	SOLAR	VOLC	NAOI	SOI	AMOI	R^2
SPI-1	**MAM**	0.015	**−0.180**	−0.079	**−0.338**	−0.017	−0.063	**0.152**
		Year	**Year**	**DJF**	**JFM**	**JFM**	**JFM**	
	JJA	−0.016	0.033	−0.075	**−0.176**	−0.111	0.089	**0.072**
		Year	**Year**	**DJFM**	**MJJ**	**MAM**	**JFM**	
SPI-12	**MAM**	−0.048	0.036	−0.044	**−0.247**	−0.127	0.034	**0.072**
		Year	**Year**	**15 m**	**15 m**	**15 m**	**15 m**	
	JJA	−0.077	−0.003	−0.107	**−0.327**	**−0.174**	−0.023	**0.123**
		Year	**Year**	**15 m**	**15 m**	**15 m**	**15 m**	
SPEI-1	**MAM**	**−0.206**	**−0.220**	−0.063	**−0.442**	−0.027	−0.061	**0.343**
		Year	**Year**	**MAM**	**JFM**	**JFM**	**JFM**	
	JJA	**−0.231**	0.051	−0.013	**−0.234**	−0.087	−0.108	**0.083**
		Year	**Year**	**JJA**	**MJJ**	**MJJA**	**JJA**	
SPEI-12	**MAM**	**−0.371**	0.014	−0.026	**−0.308**	**−0.181**	−0.092	**0.190**
		Year	**Year**	**15 m**	**15 m**	**15 m**	**15 m**	
	JJA	**−0.410**	−0.047	−0.083	**−0.355**	**−0.179**	−0.084	**0.236**
		Year	**Year**	**15 m**	**15 m**	**15 m**	**15 m**	
Z-index	**MAM**	**−0.195**	**−0.192**	−0.073	**−0.490**	−0.016	−0.097	**0.364**
		Year	**Year**	**DJFMAM**	**JFM**	**JFM**	**JFM**	
	JJA	−0.185	−0.005	−0.093	**−0.231**	−0.123	−0.098	**0.093**
		Year	**Year**	**DJFMAM**	**JFMAMJJ**	**MJJA**	**JJA**	
PDSI	**MAM**	**−0.402**	−0.054	−0.080	**−0.226**	−0.138	−0.096	**0.265**
		Year	**Year**	**DJF**	**JFM**	**DJFM**	**DJF**	
	JJA	**−0.395**	−0.035	−0.108	**−0.197**	−0.171	−0.123	**0.233**
		Year	**Year**	**DJF**	**JFMAMJ**	**DJFMAMJJA**	**JJA**	

Values statistically significant for $\alpha = 0.05$ are indicated by bold. The abbreviations below the coefficient values indicate the predictor's source period, chosen during pre-processing (with 15 m indicating 15-month period preceding the end of the target drought period). R^2 (coefficient of determination) shows the fraction of variance explained by the regression mapping.

links were indicated in some cases. This is not surprising, considering that the effects of SO on Central European weather are rather limited, if occasionally detectable (e.g. Brázdil and Bíl, 1998; Brönnimann, 2007; Brönnimann *et al.*, 2007). Even so, the prevalence of negative regression coefficients between various drought indices and SOI series suggests a possibility of a deterministic link, although of varying strength. However, Piervitali and Colacino (2001), analysing drought events for the 1565–1915 period in Western Sicily, showed that a reduction in El Niño Southern Oscillation (ENSO) events took place in periods when many drought events occurred and *vice versa*.

The solar irradiance series exhibited a statistically significant, but relatively weak, relation to some of the drought indices (SPEI-1, SPI-1 and Z-index) in MAM. While this may be a sign of an actual physical relation between the variables in question (with higher irradiation increasing temperature and reducing drought indices), there exists the possibility of intermixing between the influences of the CO2 and SOLAR signals, which share some of their basic long-term variability (lower values at the start of series and higher towards the end).

There is no statistically significant relation between the series of volcanic aerosol optical thickness (VOLC) and any of the drought indices. Such an outcome is consistent with the findings of Písek and Brázdil (2006), who identified no clear and systematic imprint of major volcanic episodes in the Central European temperature series, and it confirms the low level of influence of the global volcanic activity on the Czech climate, at least in the target period. However, the systematically negative value of the respective regression coefficients may suggest the presence of some deterministic link, perhaps extractable by other means of statistical analysis.

The Atlantic Multidecadal Oscillation has been shown to explain a portion of global temperature variance (Rohde *et al.*, 2013). In this analysis, however, this predictor did not appear to contribute in a significant way.

Even when R^2 was at its highest (SPEI-1 and Z-index in the MAM target period), just a relatively small fraction of variance in the drought indices was explained and many features of the target series remain unresolved (Figure 6). The lack of skill in the statistical model was at its most apparent for JJA. This suggests that although some of the explanatory variables included in this analysis provide well-defined, statistically significant contributions to the series of drought descriptors, most of the drought temporal variability is connected to other factors, probably related to regional climate/weather and local influences. Such findings have already been reported by Trnka *et al.* (2009), who identified a set of general circulation patterns closely linked to drought events over the territory.

The tests repeated for the full period analysed (1806–2012) with just CO2, SOLAR, VOLC and NAOI predictors gave results qualitatively similar to the full regression for 1867–2012, but with a lower fraction of variance explained. This decrease cannot be fully ascribed to the reduced number of predictors, as it also takes place when comparing the four-predictor experiments for the periods 1806–2012 and 1867–2012. The reduction of model skill for the longer-period setup thus implies weaker (or different) relations between the target and explanatory variables in the early periods of the signals, possibly due to the uncertainties discussed in Section 5. The potential non-stationarity of the predictor-predictand relations was also investigated, particularly for NAOI as the dominant fast-variable predictor. The results indicate weakened (although mostly still detectable) links to NAO during much of the 19th century for MAM (similar to the findings of Todd *et al.*, 2013). More complicated temporal patterns of NAO influence emerged for JJA, especially for the short-term drought indices. This can probably be ascribed to the relatively less pronounced effects of NAO phase during the Central European summer.

4.4. Droughts and crop yields

The effects of drought on agriculture are reflected mainly in the reduction of crop yields. There are long series of yields available for the Czech Lands, facilitating study and quantification of impacts. The yields show quite significant temporal changes, with more than fourfold mean increases between the end of the 19th century and the end of the 20th century (Figure 7(a) and (d)). This increasing trend is attributed to technological advances in agriculture (Trnka *et al.*, 2012). Rather than analysing absolute yields, we addressed their deviation in individual years from the four closest years of the record for which data were available. This allowed us to remove the 'technological trend' while retaining information on the interannual yield variability that is thought to be most influenced by climate parameters.

Previous analysis of yield records around the station with the most complete data record (i.e. Brno) confirmed that the water-balance *versus* yield relationship is best estimated by a second-order polynomial (Brázdil *et al.*, 2009b). This closely represents the nature of the crop yield–water relationship (Ash *et al.*, 1992) as crop yields may be inhibited not only through water stress but also by low global radiation, below-normal temperatures, root anoxia and higher infestation pressure of fungal diseases, all factors that tend to be associated with unusually wet seasons. However, the relationship between water-balance indicators and yield changes is relatively weak, explaining less than 15% of yield variability in the case study region (around the area of the Brno meteorological station). In the assessment of this relationship for the lowest-yield quartile, Z-index explained up to 37% of the low-yield variability for winter wheat and over 68% for spring barley (Figure 7(c) and (f)). There was no such response for the other three quartiles. This indicates that lack or surplus of water is the key factor explaining most of the lowest yields of spring barley and is quite important for wheat yields (Figure 7). A similar direct relationship between SPI or precipitation and crop yields was not confirmed.

The highest influence of drought on Czech national crop yields may be found in the years 1922, 1934, 1947, 1976,

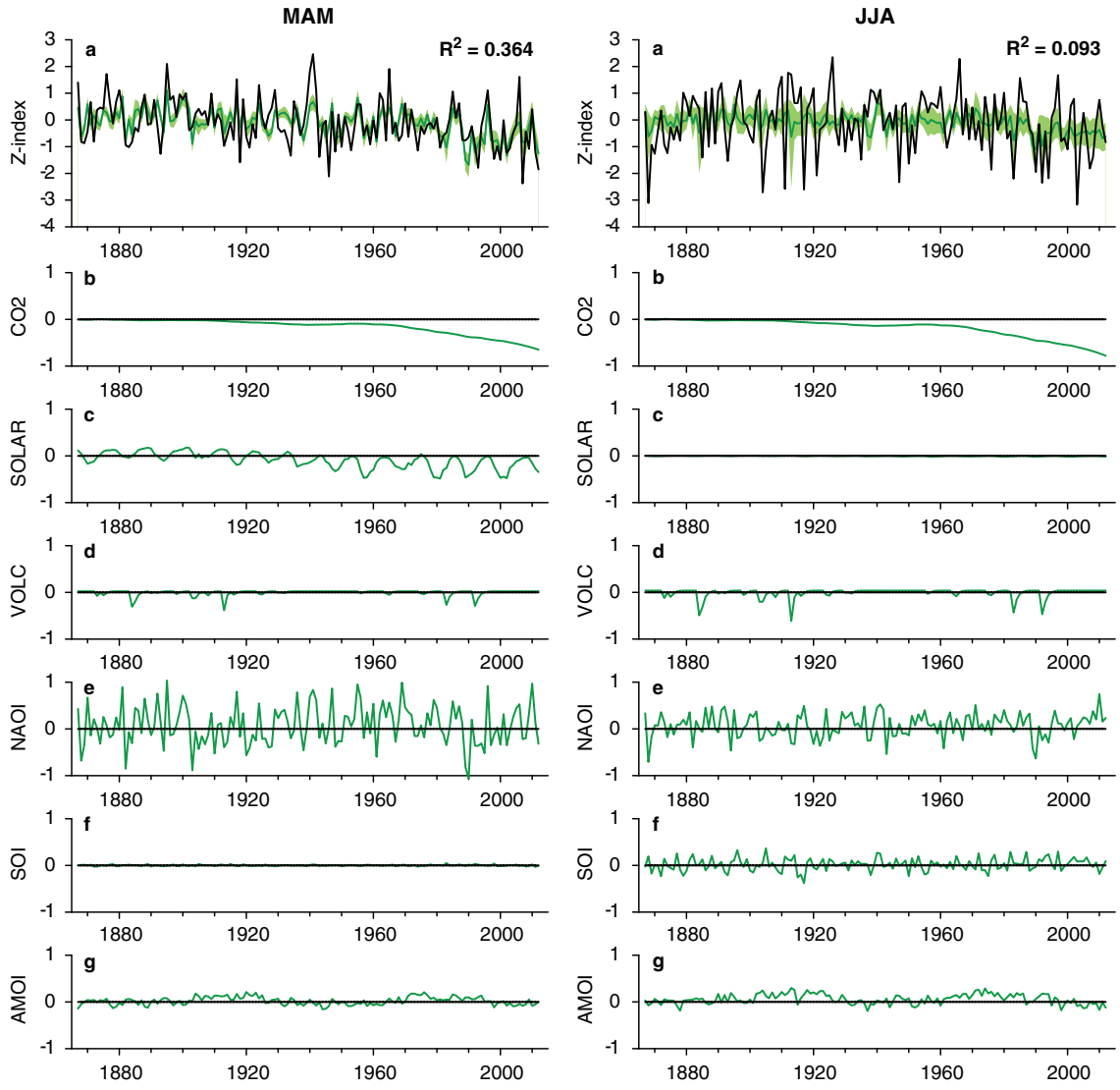


Figure 6. MAM and JJA Z-index series (black line), together with its estimate by linear regression model (grey line, with 95% confidence interval displayed by shading) (a), and contributions from individual explanatory variables (b–g), shown relative to their 1867–1877 means.

1988, 1992, 1993, 2000, 2003, 2007 and 2012. At the level of regional yields in the Brno area, the years most affected were 2012 (−40%/−64% spring barley/wheat yield reduction), 2000 (−49%/−34%), 1993 (−24%/−27%), 2007 (−31%/−14%), 2003 (−17%/−32%) and 1869 (−28%/−19%). Interestingly, the year of 1879 was extremely wet, leading to yield decreases of 26%/34%. The years 2000 and 2012 are especially noteworthy. Decreases in yields do not include thousands of hectares of crops that had to be re-sown or were not harvested at all.

5. Discussion

Despite the use of homogenized meteorological data in this study, calculation of drought indices may be slightly biased by uncertainty in mean Czech temperature and precipitation series employing the full numbers of ten stations used from 1883 (temperature) and 14 stations from 1876 (precipitation). This does not create a problem for temperature, as there are high spatial correlations between even relatively distant stations. A different situation arises for precipitation totals because precipitation-forming synoptic processes are of quite limited territorial extent, as reflected in the fact that spatial correlations decrease sharply with increasing distances between stations (Brázdil *et al.*, 2012a, 2012b). This may partly diminish the validity of Czech precipitation series before 1876.

Precipitation-based drought indices SPI-1 and SPI-12 show slightly increasing linear trends for the Czech Lands in the long term, although statistically non-significant. This corresponds to non-significant trends in secular Czech precipitation series, indicating cyclic features in

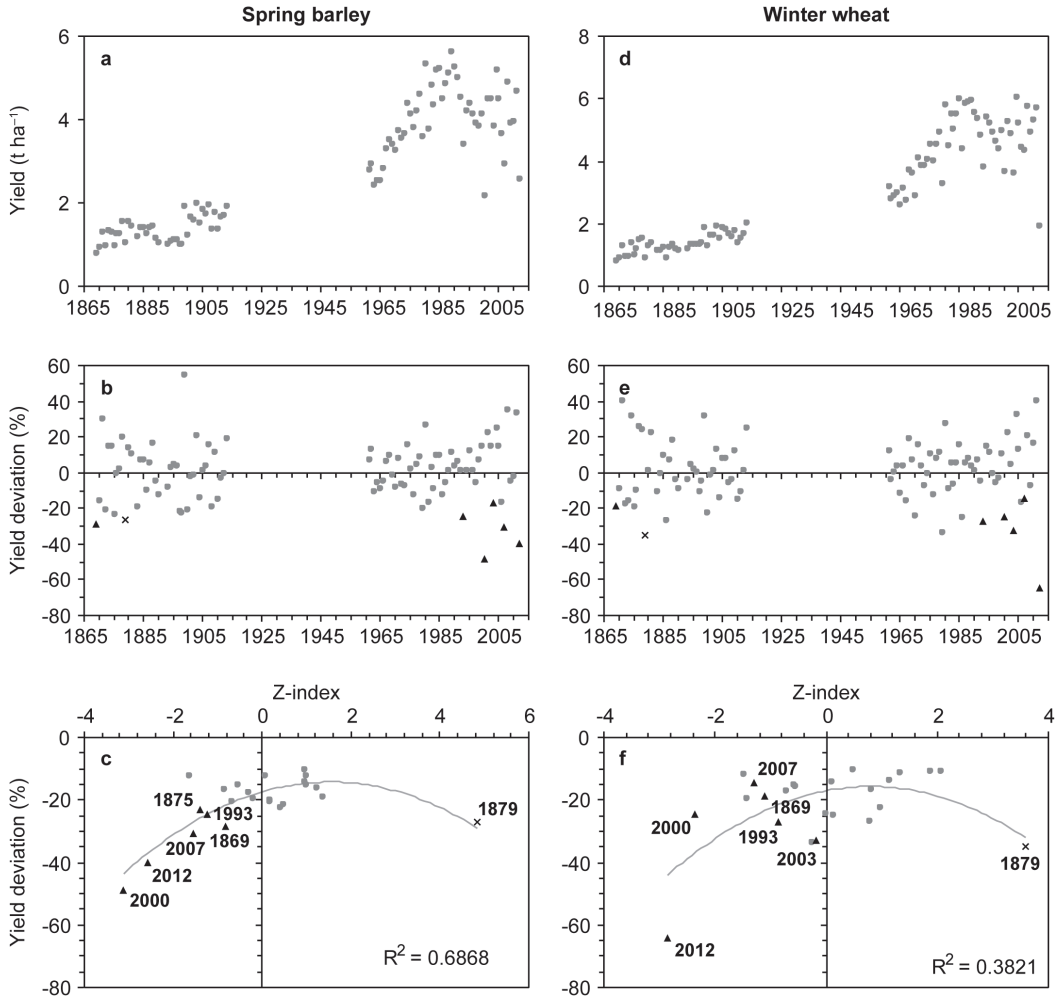


Figure 7. (a) Mean annual yields of spring barley (dots) around the Brno station (Brno-venkov and Břeclav districts) between 1869 and 2012; (b) yield deviation from four closest years with years affected by extremely dry conditions marked as triangles (cross indicates extremely wet conditions); (c) relationship between drought intensity expressed as Z-index and yield deviation in 25% of the lowest-yielding years from April to June; (d–f) as (a–c) but for winter wheat. N.B. Years marked at triangles were characterized by extremely dry conditions in at least two of the growing season months and a yield deviation of at least 25% compared to the four closest years for at least for one of the crops.

precipitation fluctuations (Brázdil *et al.*, 2012a, 2012b). In contrast, the four remaining drought indices (SPE-1, SPEI-12, Z-index and PDSI) show clearly negative linear trends for MAM and JJA that were mainly statistically significant. This means that the 208-year Czech series exhibits a significant tendency to increasing MAM and JJA dryness. While lack of precipitation was a primary reason for drought episodes in the greater part of the 1805–2012 period (see, e.g. frequent droughts in 1861–1870 in Figure 4), recent decades highlight the importance of significantly increasing temperatures as causative of medium-term and long-term droughts, documented in the 1990s and particularly in 2004–2012. This is largely in agreement with independent studies by Trnka M, 2014a, 2014b (pers. comm.) that employed a process-based soil-water-balance model. This indicated that most of the Czech Republic exhibits a tendency towards decreasing soil-water content between April and September, whereas the period from October to March typically displays a tendency towards increased soil-water content, particularly at higher altitudes. Trnka M, 2014b (pers. comm.) conclude that drying trends are strongest in May (over 44.1% of all Czech territory) followed by June (36.4%). Drought trends in other months are far less pronounced. A trend towards lower soil moisture content was also found in the eastern part of the Czech Republic during July–September.

In the wider area of Central Europe Dai *et al.* (2004) and Dai (2011a) reported the existence of a notable drying trend since the beginning of the 20th century, which they linked to increasing temperatures. Lloyd-Hughes and Saunders (2002) and van der Schrier *et al.* (2006) reported only insignificant changes in the areas experiencing moderate-to-extreme drought conditions during the 20th century in Europe as a whole. However, if the results of the two studies are analysed for only the region

delimited by 48°–51°N and 13°–18°E, a drying trend can be identified. Sheffield *et al.* (2012) argue that the simplified model of PET used in PDSI overestimates increase in global drought as a reflection of temperature rise. Applying more realistic calculations including available energy, humidity and wind speed show little change in drought over the past 60 years (for physical aspects of PDSI, see Seneviratne, 2012 and for changes in droughts due to global warming see Trenberth *et al.*, 2014). Nevertheless, calculation of soil moisture content using the SoilClim model based on the Penman–Monteith method (Hlavinka *et al.*, 2011) has shown an increased tendency towards lower soil moisture content (especially from April to June) during the 1961–2010 period (Trnka M, 2014a, 2014b (pers. comm.)). Moreover, Trnka M, 2014a, 2014b (pers. comm.) reported a decline in measured soil moisture content and a significant increase in measured pan evaporation for the same period. Todd *et al.* (2013) analysed scPDSI series for three stations in southeast England. They identified multiple drought-rich periods in 1730–1760 and 1890–present. The latter period shows an increasing tendency towards more severe droughts, in agreement with the results of this study.

Sheffield *et al.* (2009) identified the five top-ranked drought events for Europe in the 1951–2000 period based on duration (over three months) and extent: 1959–1961, 1976–1977, 1975–1976, 1951–1952 and 1995–1996. Although this analysis evaluates droughts over a long timescale and is limited only to MAM and JJA droughts, any coincidence with these periods is very weak or even random (compare Figure 4): MAM – 1976 (SPI-1), 1995 (SPEI-12); JJA – 1976 (all indices except PDSI), 1952 (all indices). This is, firstly, due to inter-model differences in soil depth and water-holding capacities that impact soil moisture retention and persistence between the Sheffield *et al.*'s (2009) modelling approach and the indices employed in this study. Secondly, the global coverage of the study means that the continental-scale drought events frequently miss the area of the Czech Lands, as is depicted in the Sheffield *et al.*'s (2009) drought episodes of 1950 and 1951. Finally, the Sheffield *et al.* (2009) findings in terms of specific year with particular drought exposure also differ from the results of Lloyd-Hughes and Saunders (2002) and van der Schrier *et al.* (2006), and absolute agreement between studies using different sets of modelling approaches, thresholds and assumptions is in any event unlikely.

Trnka *et al.* (2013) showed that, given current climate conditions, 95.5% of the Czech territory is dominated by quite 'wet' soil climate regimes. However, whatever the differences among global circulation models, these wet-dominant regimes will be greatly reduced in area or diminished in intensity under future climate conditions, i.e. by the year 2100. Such a changed climate will inevitably and considerably affect the probability of very dry years (i.e. years when the soil is dry for most of the summer throughout its profile), increasing it more than tenfold.

The time series of most drought indices carry the imprints of one or more large-scale climate forcing factors, although the relative strength and significance of such influences varies with the type of the index as well as the season. Of particular prominence is the long-term component associated with the intensification of radiative forcing arising out of increases in the concentration of GHGs. Considering the statistically strong connection between GHG quantities and Czech temperature, and an almost non-existent link to precipitation (Brázdil *et al.*, 2012a; Mikšovský *et al.*, 2014), it is reasonable to assume that GHG-induced temperature increase represents the primary transfer mechanism introducing the trend component to drought indices.

Of the faster-variable forcings, NAO phase is by far the most influential, with high values of NAO index coinciding with lowered values for all drought indices. The mechanisms underlying the drought–NAO interaction appear to be linked to variability in Czech temperature (typically increased during positive NAO phases) and precipitation (generally somewhat reduced during positive NAO phases, at least for the precipitation dataset employed here). The effects of other forcing factors are substantially weaker, although this analysis suggests a possible connection between certain drought descriptors and solar activity and/or SO phase, in line with prior studies that have indicated their possible influence on European climate (e.g. Brönnimann, 2007; Brönnimann *et al.*, 2007; Bice *et al.*, 2012; Gray *et al.*, 2013; Mikšovský *et al.*, 2014).

The employment of drought indices calculated for the whole territory of the Czech Republic may smooth out quite important differences between its various parts, differences that may have significant economic consequences, quite apart from the timing of drought with respect to existing crop phenophases. For example, grain production suffers when drought affects the most productive regions, such as the Elbe lowland in Bohemia or southern Moravia. Some spatial differences in precipitation distribution with respect to drought indices were shown by Brázdil *et al.* (2013). Moreover, Zahradníček P, 2014 (pers. comm.) documented an interesting case of dipole structure in droughts, with more-or-less normal patterns in the western part of the Czech Republic (Bohemia) and severe drought in its eastern part (Moravia) for August 2011–May 2012. Such knowledge has always to be borne in mind and then carefully applied in impact studies, particularly in agriculture.

It is a well-established fact that crops change in their sensitivity to stress at the various stages of their growth (e.g. Chmielewski and Köhn, 2000). We have therefore used the response of regional crop yields at district level as presented by Hlavinka *et al.* (2009) and Zahradníček P, 2014 (pers. comm.) as measures of drought index correspondence with drought impacts. Analysis comparing the relationship between drought indices (i.e. Z-index, PDSI and SPI with SPEI at various time steps) demonstrated that most crops (e.g. winter wheat and spring barley) are susceptible to short-term drought within the April–June period and they show their closest relationship with Z-index and SPEI-1 month. Maize and sugar beet have a higher sensitivity to drought events between May

and August, whereas winter oilseed rape and winter rye (winter crops) also demonstrated particular sensitivity to water stress in the months before emergence, and during it, in autumn. This analysis also helped to establish that yield losses attributable to drought episodes might be considerable. The most severe events involve drops of 30% or more, even 65%, at regional level – and even higher when the increased rate of crop failure (i.e. sown but non-harvested area) is factored in.

However, in some cases the impact of drought on regional yield is partly hidden within the statistics, because the mean yield data do not account for whatever fraction of sown area is ploughed back and re-sown as a result of crop failure. In extremely dry years (e.g. 2000 or 2012), drought indicators enable explanation of spatial variability of yield departures and thus helps, for example, to assess total damage attributable to a particular drought event. Hlavinka *et al.* (2009) showed that the difference in available soil moisture (expressed in terms of rZ-index) could explain 65% of yield variability for spring barley between individual districts, 22% for winter wheat, 43% for winter oilseed rape, 25% for winter rye, 41% for oat and 21% for potatoes.

The most recent drought episode, which started in late summer 2011 and lasted until July 2012, also showed the importance of autumn drought to winter crops (Zahradníček P, 2014 (pers. comm.)). The yields of winter wheat in the south-eastern lowlands of the Czech Republic were the worst in 15 years (in some regions even as low as in the early 1900s), mainly due to very dry conditions all autumn and spring, exacerbated by unfavourable conditions for over-wintering (low-precipitation leading to lack of snow cover to protect against intense frosts). Moreover, there was also a substantial acreage reduction during autumn and winter 2011/2012; allowing for this would lower the mean yield by another 10%, making the 2012 drought in some areas comparatively the worst drought event in the past five decades, with the sharpest year-to-year decline in available statistical records.

6. Conclusions

This article presents an analysis of 208 years of droughts in the Czech Lands and their forcings. In comparison with other Czech drought studies (e.g. Brázdil and Kirchner, 2007; Tolasz *et al.*, 2007; Brázdil *et al.*, 2009b; Potop *et al.*, 2011, 2012, 2014; Treml, 2011), it employs the longest-yet homogeneous instrumental series for calculation of areal means and attribution analysis of the variability of drought indices, to investigate the influence of various forcing factors. This new series has enabled the calculation of long-term series of six drought indices for two basic periods – MAM and JJA, generally covering the vegetation period. The analysis demonstrates increasing dryness throughout the territory of the Czech Republic, peaking in 2004–2012, documented by significant linear trends in several drought indices. Similar long periods of drought, but of lower severity, were also recorded in 1864–1866 and 1869–1874. While the drought periods of the 19th century were attributed mainly to low-precipitation totals, the most recent ones were more related to a significant increase in temperatures.

Regression analysis of factors that may be linked to the variance in the observed drought indices revealed that the North Atlantic Oscillation phase and the aggregated effects of anthropogenic forcing (expressed through CO_2-equivalent concentration) are the most prominent predictors, although the magnitude of their influence varied strongly with the type of target index as well as the season of the year. Minor contributions were also detected from solar irradiation and Southern Oscillation phase, whereas no significant links were found to volcanic activity and Atlantic Multidecadal Oscillation. This tallies with the results of Pongrácz *et al.* (2003) who used a fuzzy rule-based technique for similar analysis of droughts in Hungary. Using Hess-Brezowsky circulation pattern types and ENSO events on monthly PDSI, they refer to their influence on drought occurrence. However, the ENSO signal is relatively weak in a statistical sense. The fraction of MAM and JJA drought index variance in the Czech Lands explained by the climate forcings and circulation oscillations under consideration was, however, relatively small, and the influence of regional and local factors appears to be dominant in the variability of the drought indices studied. Furthermore, while this analysis suggests that relations between drought descriptors and forcings are largely linear, some indications of nonlinear links were detected, suggesting a potential for further extension of the attribution analysis.

Acknowledgements

The authors gratefully acknowledge the support of the following grants in the preparation of this article: RB, JM – Grant Agency of the Czech Republic, ref. no. P209/11/0956; MT – KONTAKT LH11010 and the 'Building up a multidisciplinary scientific team focused on drought' project, ref. no. CZ.1.07/2.3.00/20.0248; LŘ – Grant Agency of the Czech Republic, ref. no. 13-19831S; PD – Grant Agency of the Czech Republic, ref. no. 13–04291S. We would also like to thank Tony Long (Svinošice, Czech Republic) for helping to work up the English.

References

Allen RG, Pereira LS, Raes D, Smith M. 1998. Crop evapotranspiration. Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56, FAO: Rome.

Ash GHB, Shaykewich CF, Raddatz RL. 1992. Moisture risk assessment for spring wheat on the eastern prairies: a water-use simulation model. *Climatol. Bull.* **26**: 65–78.

Bice D, Montanari A, Vučetić V, Vučetić M. 2012. The influence of regional and global climatic oscillation on Croatian climate. *Int. J. Climatol.* **32**: 1537–1557, DOI: 10.1002/joc.2372.

Blinka P. 2005. Klimatologické hodnocení sucha a suchých období na území České republiky v letech 1876–2002 (Climatological evaluation of drought and dry periods on the territory of the Czech Republic in the years 1876–2002). *Meteorol. zprávy* **58**: 10–18.

Brázdil R. 1998. Meteorological extremes and their impacts on forests in the Czech Republic. In *The Impacts of Climate Variability on Forest*, Beniston M, Innes JL (eds). Springer: Berlin and Heidelberg, Germany New York, NY, 19–47.

Brázdil R, Bíl M. 1998. Jev El Niño-Jižní Oscilace a jeho možné projevy v polích tlaku vzduchu, teploty vzduchu a srážek v Evropě ve 20. století (El Niño-Southern Oscillation and its effects on air pressure, air temperature and precipitation in Europe in the 20th century). *Geografie* **103**: 65–87.

Brázdil R, Kirchner K (eds). 2007. *Vybrané přírodní extrémy a jejich dopady na Moravě a ve Slezsku (Selected Natural Extremes and Their Impacts in Moravia and Silesia)*. Masarykova univerzita, Český hydrometeorologický ústav, Ústav geoniky Akademie věd České republiky, v.v.i.: Brno, Praha, Ostrava.

Brázdil R, Štěpánková P, Kyncl T, Kyncl J. 2002. Fir tree-ring reconstruction of March–July precipitation in southern Moravia (Czech Republic), 1376–1996. *Clim. Res.* **20**: 223–239, DOI: 10.3354/cr020223.

Brázdil R, Chromá K, Dobrovolný P, Tolasz R. 2009a. Climate fluctuations in the Czech Republic during the period 1961–2005. *Int. J. Climatol.* **29**: 223–242, DOI: 10.1002/joc.1718.

Brázdil R, Trnka M, Dobrovolný P, Chromá K, Hlavinka P, Žalud Z. 2009b. Variability of droughts in the Czech Republic, 1881–2006. *Theor. Appl. Climatol.* **97**: 297–315, DOI: 10.1007/s00704-008-0065-x.

Brázdil R, Bělínová M, Dobrovolný P, Mikšovský J, Pišoft P, Řezníčková L, Štěpánek P, Valášek H, Zahradníček P. 2012a. *Temperature and Precipitation Fluctuations in the Czech Lands During the Instrumental Period*. Masaryk University: Brno, Czech Republic.

Brázdil R, Zahradníček P, Pišoft P, Štěpánek P, Bělínová M, Dobrovolný P. 2012b. Temperature and precipitation fluctuations in the Czech Lands during the period of instrumental measurements. *Theor. Appl. Climatol.* **110**: 17–34, DOI: 10.1007/s00704-012-0604-3.

Brázdil R, Dobrovolný P, Trnka M, Kotyza O, Řezníčková L, Valášek H, Zahradníček P, Štěpánek P. 2013. Droughts in the Czech Lands, 1090–2012 AD. *Clim. Past* **9**: 1985–2002, DOI: 10.5194/cp-9-1985-2013.

Brönnimann S. 2007. Impact of El Niño-Southern Oscillation on European climate. *Rev. Geophys.* **45**: RG3003, DOI: 10.1029/2006RG000199.

Brönnimann S, Xoplaki E, Casty C, Pauling A, Luterbacher J. 2007. ENSO influence on Europe during the last centuries. *Clim. Dyn.* **28**: 181–197, DOI: 10.1007/s00382-006-0175-z.

Büntgen U, Brázdil R, Dobrovolný P, Trnka M, Kyncl T. 2011. Five centuries of Southern Moravian drought variations revealed from living and historic tree rings. *Theor. Appl. Climatol.* **105**: 167–180, DOI: 10.1007/s00704-010-0373-9.

Byun H, Wilhite DA. 1999. Objective quantification of drought severity and duration. *J. Clim.* **12**: 2747–2756, DOI: 10.1175/1520-0442(1999)0122.0.CO;2.

Cahynová M. 2005. Vliv Severoatlantské oscilace na sezonní teploty vzduchu ve střední Evropě (Influence of the North Atlantic Oscillation on seasonal temperatures in Central Europe). *Meteorol. zprávy* **58**: 41–46.

Chmielewski F-M, Köhn W. 2000. Impact of weather on yield and yield components of winter rye over 30 years. *Agric. For. Meteorol.* **102**: 253–261, DOI: 10.1016/S0168-1923(00)00125-8.

Ciais P, Reichstein M, Viovy N, Granier A, Ogée J, Allard V, Aubinet M, Buchman N, Bernhofer C, Carrar A, Chevallier F, De Noblet N, Friend AD, Fiedlingstein P, Grünwald T, Heinesch B, Keronen P, Knohl A, Krinner G, Loustau D, Manca G, Matteucci G, Miglietta F, Ourcival JM, Papale D, Pilegaard K, Rambal S, Seufert G, Soussana JF, Sanz MJ, Schulze ED, Vesala T, Valentini R. 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* **437**: 529–533, DOI: 10.1038/nature03972.

Crowley TJ, Zielinski G, Vinther B, Udisti R, Kreutz K, Cole-Dai J, Castellano E. 2008. Volcanism and the Little Ice Age. *PAGES News* **16**: 22–23.

Dai A. 2011a. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. *J. Geophys. Res.* **116**: D12115, DOI: 10.1029/2010JD015541.

Dai A. 2011b. Drought under global warming: a review. *WIREs Clim. Change* **2**: 45–65, DOI: 10.1002/wce.61.

Dai A, Trenberth KE, Taotao Q. 2004. A global dataset of Palmer Drought Severity Index for 1870–2002: relationship with soil moisture and effects of surface warming. *J. Hydrometeorol.* **5**: 1117–1130, DOI: 10.1175/JHM-386.1.

Dubrovský M, Svoboda MD, Trnka M, Hayes MJ, Wilhite DA, Žalud Z, Hlavinka P. 2009. Application of relative drought indices in assessing climate-change impacts on drought conditions in Czechia. *Theor. Appl. Climatol.* **96**: 155–171, DOI: 10.1007/s00704-008-0020-x.

Dufková J, Št'astná M. 2005. Determination of drought occurrence trend in the area of Southern Moravia and its influence on soil erosion. *Meteorol. časopis* **8**: 131–136.

Enfield DB, Mestas-Nuñez AM, Trimble PJ. 2001. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res. Lett.* **28**: 2077–2080, DOI: 10.1029/2000GL012745.

Gray LJ, Scaife AA, Mitchell DM, Osprey S, Ineson S, Hardiman S, Butchart N, Knight J, Sutton R, Kodera K. 2013. A lagged response to the 11 year solar cycle in observed winter Atlantic/European weather patterns. *J. Geophys. Res. Atmos.* **118**: 13405–13420, DOI: 10.1002/2013JD020062.

Haykin S. 1999. *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice-Hall: Upper Saddle River, NJ.

Heim RR. 2002. A review of twentieth-century drought indices used in the United States. *Bull. Am. Meteorol. Soc.* **83**: 1149–1165.

Hlavinka P, Trnka M, Semerádová D, Dubrovský M, Žalud Z, Možný M. 2009. Effect of drought on yield variability of key crops in Czech Republic. *Agric. For. Meteorol.* **149**: 431–442, DOI: 10.1016/j.agrformet.2008.09.004.

Hlavinka P, Trnka M, Balek J, Semerádová D, Hayes M, Svoboda M, Eitzinger J, Možný M, Fischer M, Hunt E, Žalud Z. 2011. Development and evaluation of the SoilClim model for water balance and soil climate estimates. *Agric. Water Manage.* **98**: 1249–1261, DOI: 10.1016/j.agwat.2011.03.011.

Hoerling MP, Eischeid JK, Quan X-W, Diaz HF, Webb RS, Dole RM, Easterling DR. 2012. Is a transition to semipermanent drought conditions imminent in the U.S. Great Plains? *J. Clim.* **25**: 8380–8386, DOI: 10.1175/JCLI-D-12-00449.1.

Jones PD, Jonsson T, Wheeler D. 1997. Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and South-West Iceland. *Int. J. Climatol.* **17**: 1433–1450, DOI: 10.1002/(SICI)1097-0088(19971115)17:13<1433::AID-JOC203>3.0.CO;2-P.

Karl TR. 1986. The sensitivity of the Palmer Drought Severity Index and Palmer's Z index to their calibration coefficients including potential evapotranspiration. *J. Clim. Appl. Meteorol.* **25**: 77–86, DOI: 10.1175/1520-0450(1986)025<0077:TSOTPD>2.0.CO;2.

Kolář P, Trnka M, Brázdil R, Hlavinka P. 2013. Influence of climatic factors on the low yields of spring barley and winter wheat in Southern Moravia (Czech Republic) during the 1961–2007 period. *Theor. Appl. Climatol.*, DOI: 10.1007/s00704-013-1037-3.

Kopp G, Lean JL. 2011. A new, lower value of total solar irradiance: evidence and climate significance. *Geophys. Res. Lett.* **38**: L01706, DOI: 10.1029/2010GL045777.

Laughlin GP, Zuo HP, Walcott J, Bugg AL. 2003. The rainfall reliability wizard – a new tool to rapidly analyse spatial rainfall reliability with examples. *Environ. Model. Softw.* **18**: 49–57, DOI: 10.1016/S1364-8152(02)00037-3.

Lloyd-Hughes B. 2013. The impracticality of a universal drought definition. *Theor. Appl. Climatol.*, DOI: 10.1007/s00704-013-1025-7.

Lloyd-Hughes B, Saunders MA. 2002. A drought climatology for Europe. *Int. J. Climatol.* **22**: 1571–1592, DOI: 10.1002/joc.846.

Luterbacher J, Schmutz C, Gyalistras D, Xoplaki E, Wanner H. 1999. Reconstruction of monthly NAO and EU indices back to AD 1675. *Geophys. Res. Lett.* **26**: 2745–2748, DOI: 10.1029/1999GL900576.

Mavromatis T. 2007. Drought index evaluation for assessing future wheat production in Greece. *Int. J. Climatol.* **27**: 911–924, DOI: 10.1002/joc.1444.

McKee TB, Doesken NJ, Kleist J. 1993. The relationship of drought frequency and duration to time scales. In *Preprints, Eighth Conference on Applied Climatology*. American Meteorological Society: Anaheim, CA, 179–184.

Meinshausen M, Smith SJ, Calvin K, Daniel JS, Kainuma MLT, Lamarque J-F, Matsumoto K, Montzka SA, Raper SCB, Riahi K, Thomson A, Velders GJM, van Vuuren DPP. 2011. The RCP GHG concentrations and their extension from 1765 to 2300. *Clim. Change* **109**: 213–241, DOI: 10.1007/s10584-011-0156-z.

Mikšovský J, Raidl A. 2006. Testing for nonlinearity in European climatic time series by the method of surrogate data. *Theor. Appl. Climatol.* **83**: 21–33, DOI: 10.1007/s00704-005-0130-7.

Mikšovský J, Pišoft P, Raidl A. 2008. Global patterns of nonlinearity in real and GCM-simulated atmospheric data. In *Nonlinear Time Series Analysis in the Geosciences: Applications in Climatology,*

Geodynamics and Solar-Terrestrial Physics, Donner R, Barbosa S (eds). Springer: Berlin and Heidelberg, Germany, Lecture Notes in Earth Sciences **112**, 17–34.

Mikšovský J, Brázdil R, Štěpánek P, Zahradníček P, Pišoft P. 2014. Long-term variability of temperature and precipitation in the Czech Lands: an attribution analysis. *Clim. Change* (in press).

Mishra AK, Singh VP. 2010. A review of drought concepts. *J. Hydrol.* **391**: 202–216, DOI: 10.1016/j.jhydrol.2010.07.012.

Možný M. 2004. *Vymezení a intenzita sucha na území ČR v letech 1891–2003 (Delimitation and Intensity of Drought Over the ČR Territory in 1891–2003)*. Český hydrometeorologický ústav: Praha.

Ntale HK, Gan TY. 2003. Drought indices and their application to East Africa. *Int. J. Climatol.* **23**: 1335–1357, DOI: 10.1002/joc.931.

Palmer WC. 1965. Meteorological drought. Research Paper No. 45, U.S. Department of Commerce, Weather Bureau, Washington, DC.

Piervitali E, Colacino M. 2001. Evidence of drought in Western Sicily during the period 1565–1915 from liturgical offices. *Clim. Change* **49**: 225–238, DOI: 10.1023/A:1010746612289.

Písek J, Brázdil R. 2006. Responses of large volcanic eruptions in the instrumental and documentary climatic data over Central Europe. *Int. J. Climatol.* **26**: 439–459, DOI: 10.1002/joc.1249.

Pongrácz R, Bogardi I, Duckstein L. 2003. Climatic forcing of droughts: a Central European example. *Hydrol. Sci. J.* **48**: 39–50, DOI: 10.1623/hysj.48.1.39.43480.

Potop V, Soukup J, Možný M. 2011. Drought at various timescales for secular lowland climatological stations in the Czech Republic. *Meteorol. zprávy* **64**: 177–187.

Potop V, Možný M, Soukup J. 2012. Drought evolution at various time scales in the lowland regions and their impact on vegetable crops in the Czech Republic. *Agric. For. Meteorol.* **156**: 121–133, DOI: 10.1016/j.agrformet.2012.01.002.

Potop V, Boroneant C, Možný M, Štěpánek P, Skalák P. 2014. Observed spatiotemporal characteristics of drought on various time scales over the Czech Republic. *Theor. Appl. Climatol.* **115**: 563–581, DOI: 10.1007/s00704-013-0908-y.

Rohde R, Muller RA, Jacobsen R, Muller E, Perlmutter S, Rosenfeld A, Wurtele J, Groom D, Wickham C. 2013. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinform. Geostat.: An Overview* **1**: 1, DOI: 10.4172/2327-4581.1000101.

Rowell PD, Jones RG. 2006. Causes and uncertainty of future summer drying over Europe. *Clim. Dyn.* **27**: 281–299, DOI: 10.1007/s00382-006-0125-9.

Schmidt GA, Jungclaus JH, Ammann CM, Bard E, Braconnot P, Crowley TJ, Delaygue G, Joos F, Krivova NA, Muscheler R, Otto-Bliesner BL, Pongratz J, Shindell DT, Solanki SK, Steinhilber F, Vieira LEA. 2012. Climate forcing reconstructions for use in PMIP simulations of the last millennium. *Geosci. Model Dev.* **5**: 185–191, DOI: 10.5194/gmd-5-185-2012.

van der Schrier G, Briffa KR, Jones PD, Osborn TJ. 2006. Summer moisture variability across Europe. *J. Clim.* **19**: 2818–2834, DOI: 10.1175/JCLI3734.1.

van der Schrier G, Efthymiadis D, Briffa KR, Jones PD. 2007. European Alpine moisture variability 1800–2003. *Int. J. Climatol.* **27**: 415–427, DOI: 10.1002/joc.1411.

Seneviratne SI. 2012. Historical drought trends revisited. *Nature* **491**: 338–339, DOI: 10.1038/491338a.

Sheffield J, Andreadis KM, Wood EF, Lettemnaier DP. 2009. Global and continental drought in the second half of the twentieth century: severity-area-duration analysis and temporal variability of large-scale events. *J. Clim.* **22**: 1962–1981, DOI: 10.1175/2008JCLI2722-1.

Sheffield J, Wood EF, Roderick ML. 2012. Little change in global drought over the past 60 years. *Nature* **491**: 435–438, DOI: 10.1038/nature11575.

Stocker TF, Qin D, Plattner G-K, Tignor MMB, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). 2013. *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, UK.

Szinell CS, Bussay A, Szentimrey T. 1998. Drought tendencies in Hungary. *Int. J. Climatol.* **18**: 1479–1491, DOI: 10.1002/(SICI)1097-0088(19981115)18:13<1479::AID-JOC325>3.0.CO;2-P.

Thornthwaite CW. 1948. An approach towards a rational classification of climate. *Geogr. Rev.* **38**: 55–94, DOI: 10.2307/2107309.

Todd B, Macdonald N, Chiverrell RC, Caminade C, Hooke JM. 2013. Severity, duration and frequency of drought in SE England from 1697 to 2011. *Clim. Change* **121**: 673–687, DOI: 10.1007/s10584-013-0970-6.

Tolasz R, Míková T, Valeriánová A, Voženílek V (eds). 2007. *Climatic Atlas of Czechia*. Český hydrometeorologický ústav: Praha; Olomouc.

Treml P. 2011. Největší sucha na území České republiky v období let 1875–2010 (The largest droughts in the Czech Republic in the 1875–2010 period). *Meteorol. zprávy* **64**: 168–176.

Trenberth KE. 1984. Signal versus noise in the Southern Oscillation. *Mon. Weather Rev.* **112**: 326–332, DOI: 10.1175/1520-0493(1984)112<0326:SVNITS>2.0.CO;2.

Trenberth KE, Dai A, van der Schrier G, Jones PD, Barichivich J, Briffa KR, Sheffield J. 2014. Global warming and changes in drought. *Nat. Clim. Change* **4**: 17–22, DOI: 10.1038/nclimate2067.

Trigo RM, Osborn TJ, Corte-Real JM. 2002. The North Atlantic Oscillation influence on Europe: climate impacts and associated physical mechanisms. *Clim. Res.* **20**: 9–17, DOI: 10.3354/cr020009.

Trnka M, Dubrovský M, Svoboda M, Semerádová D, Hayes M, Žalud Z, Wilhite D. 2009. Developing a regional drought climatology for the Czech Republic. *Int. J. Climatol.* **29**: 863–883, DOI: 10.1002/joc.1745.

Trnka M, Brázdil R, Olesen JE, Eitzinger J, Zahradníček P, Kocmánková E, Dobrovolný P, Štěpánek P, Možný M, Bartošová L, Hlavinka P, Semerádová D, Valášek H, Havlíček M, Horáková V, Fischer M, Žalud Z. 2012. Could the changes in regional crop yields be a pointer of climatic change? *Agric. For. Meteorol.* **166–167**: 62–71, DOI: 10.1016/j.agrformet.2012.05.020.

Trnka M, Kersebaum KC, Eitzinger J, Hayes M, Hlavinka P, Svoboda M, Dubrovský M, Semerádová D, Wardlow B, Pokorný E, Možný M, Wilhite D, Žalud Z. 2013. Consequences of climate change for the soil climate in Central Europe and the central plains of the United States. *Clim. Change* **120**: 405–418, DOI: 10.1007/s10584-013-0786-4.

Vicente-Serrano SM, Beguería S, López-Moreno JI. 2010. A multi-scalar drought index sensitive to global warming: the Standardized Precipitation Evapotranspiration Index – SPEI. *J. Clim.* **23**: 1696–1718, DOI: 10.1175/2009JCLI2909.1.

Wang Y-M, Lean JL, Sheeley NR Jr. 2005. Modeling the sun's magnetic field and irradiance since 1713. *Astrophys. J.* **625**: 522–538, DOI: 10.1086/429689.

Wanner H, Brönnimann S, Casty C, Gyalistras D, Luterbacher J, Schmutz C, Stephenson DB, Xoplaki E. 2001. North Atlantic Oscillation – concepts and studies. *Surv. Geophys.* **22**: 321–382, DOI: 10.1023/A:1014217317898.

Wells N, Goddard S, Hayes M. 2004. A self-calibrating Palmer Drought Severity Index. *J. Clim.* **17**: 2335–2351, DOI: 10.1175/1520-0442(2004)017<2335:ASPDSI>2.0.CO;2.

Wilhite DA, Vanyarkho O. 2000. Drought: pervasive impacts of a creeping phenomenon. In *Drought: A Global Assessment*, Wilhite DA (ed). Routledge Publishers: London, 245–255.

Wilhite DA, Hayes MJ, Knutson C, Smith KH. 2000. Planning for drought: moving from crisis to risk management. *J. Am. Water Resour. Assoc.* **36**: 697–710, DOI: 10.1111/j.1752-1688.2000.tb04299.x.

Wilhite DA, Svoboda M, Hayes MJ. 2007. Understanding the complex impacts of drought: a key to enhancing drought mitigation and preparedness. *J. Water Resour. Manage.* **5**: 763–774, DOI: 10.1007/s11269-006-9076-5.

Wilks DS. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Academic Press: Amsterdam.

