

An analysis of global warming in the Alpine region based on nonlinear nonstationary time series models

Francesco Battaglia · Mattheos K. Protopapas

Accepted: 23 June 2012 / Published online: 21 July 2012
© Springer-Verlag 2012

Abstract The annual temperatures recorded for the last two centuries in fifteen European stations around the Alps are analyzed. They show a global warming whose growth rate is not however constant in time. An analysis based on linear Arima models does not provide accurate results. Thus, we propose threshold nonlinear nonstationary models based on several regimes both in time and in levels. Such models fit all series satisfactorily, allow a closer description of the temperature changes evolution, and help to discover the essential differences in the behavior of the different stations.

Keywords Climate change · Feedback · Regime change · Threshold model

1 Introduction

Climate is changing towards a global warming, whose consequences may be extremely serious for nature and mankind. Scientists are trying to study and explain the reasons of such change, and an authoritative Intergovernmental Panel on Climate Change (IPCC) was created in 1988. It was set up by the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP) as an effort by the United Nations to provide the governments of the world with a clear scientific view of what is happening to the world's climate.

The fourth IPCC assessment report (IPCC 2007) concludes that warming of the climate system is unequivocal; in the last hundred years the global mean surface temperatures have risen by about 0.8 °C, and most of the observed increase is very likely due to human activity.

F. Battaglia (✉) · M. K. Protopapas
Dipartimento di Scienze Statistiche, University La Sapienza, Rome, Italy
e-mail: francesco.battaglia@uniroma1.it

Though surface temperature is the result of a complex process with many factors, there is a general consensus that the recent temperature increase was induced mainly by man-made CO₂ and other greenhouse gas emissions (EEA 2008). Moreover, in addition to the direct effect of gas emissions, the complex interaction between the cycling of substances, radiation and other processes leads to feedback loops (EEA 2008). For example, a positive feedback is the melting of permafrost induced by an initial temperature increase, since the large amount of methane trapped in this soil will be released and contribute to further warming. Several feedback processes, both positive and negative, have been advocated, but their precise quantitative effect on warming is still uncertain.

On the average, in the last century the global land temperature increased by 1.0° (oceanic temperature increased less), but the northern hemisphere underwent a larger increase, and Europe warmed more, about 1.2°. Moreover, the Alps have undergone an exceptionally high temperature increase, around 2°, about twice the warming of the northern hemisphere (Zebisch et al. 2009).

Climate change in the Alps has been extensively studied by the project HISTALP (Historical instrumental climatological surface time series of the Greater Alpine Region (GAR), see Auer et al. 2007). A rich database was produced, containing homogenized records of several measures (pressure, precipitation, sunshine) and temperature, some of them dating back to 1760, for 131 sites in Alps and its wider surrounding, a region denominated GAR. One of the findings of the Histalp project is that the fundamental evolution of the temperature change was essentially similar in the whole region, regardless the altitude of the recording site, but the climate main features are different, and require a regionalization. Four climatological subregions were proposed: North–West (NW) and North–East (NE), north of the Alps, and South–West (SW) and South–East (SE), south of the Alps.

We have selected the annual temperature series at 15 sites, for which the longest data sets are available, and approximately evenly distributed among the subregions. Figure 1 shows the GAR map, the subregions and the stations location, while sites and dates are detailed in Table 1. The last two centuries were taken into account, with the aim of answering, on a statistical basis, the following questions:

- when did the temperature increase start?
- was the rate of increase uniform, or significantly varying over the recording period?
- are there relevant differences in the temperature evolution among the different stations?

A further relevant question is related to the evaluation of the effects of feedback processes. If a feedback loop has a substantial impact on the temperature, this should introduce nonlinearity in the time series.

For these reasons, we propose to fit to the temperature series a recently introduced nonlinear nonstationary model (Battaglia and Protopapas 2011, 2012). This kind of models is based on a two-fold threshold autoregressive structure, and allows for different regimes driven both by the time flow, and by the range of values of the series.

The plan of the paper is as follows. In the next Section we present the data and a linear statistical analysis. In Sect. 3 we introduce the nonlinear nonstationary model



Fig. 1 Map of the Greater Alpine Region. Lines delimit the four subregions, recording stations *underlined*

Table 1 Details of the temperature time series

Site	Code	Subregion	Altitude	First year	Last year
Bern	BER	NW	565	1803	2007
Geneve	GEN	NW	380	1803	2007
Innsbruck	INN	NW	609	1777	2007
Karlsruhe	KAR	NW	112	1779	2007
Kremsmunster	KRE	NE	389	1768	2007
Milano	MIL	SW	103	1763	2007
Munchen	MUN	NE	525	1781	2007
Padova	PAD	SE	14	1774	2005
Regensburg	REG	NE	366	1773	2007
Strasbourg	STR	NW	150	1801	2007
Stuttgart	STU	NW	311	1772	2007
Torino	TOR	SW	275	1760	2007
Udine	UDI	SE	100	1803	2007
Verona	VER	SW	67	1788	2007
Wien	WIE	NE	209	1775	2007

and the meta-heuristic method used for identification and estimation. Our findings are illustrated in Sect. 4, and some conclusions are drawn in the final section.

2 Data and preliminary analysis

The annual average temperature data at the 15 stations detailed in Table 1 have been analyzed. Each series referred, for homogeneity, to the same time span, from year 1803 to 2005 (203 observations). The time series of the HISTALP project were accurately homogenized (details of the homogenization process are found in Auer et al. 2007), since the original measurements depend on several non climatic factors. In effect, “changes in location, surroundings of sites, instruments, time of observations, observers and a number of other factors produce inhomogeneities in the series” (Auer et al. 2007, p. 19). The homogenization process aimed at correcting for these anomalies, detecting and removing outliers, and filling gaps. However, the resulting data cannot be considered free of non climatic features. In particular, the urban effect may cause differences in the 15 stations. Though the urban influence was found decreasing in time (Böhm et al. 2001), and it is clear that “a possible systematic urban bias cannot be used as an argument to doubt climatic warming” (Auer et al. 2007, p. 34), it should be kept in mind that the 15 sites are located in cities with largely different dimensions, therefore the urban influence may be different in each series.

The data, expressed in degrees Celsius $\times 10$, are summarized in Fig. 2, the four panels show the time series graphs of the stations in the four subregions. Tables and figures will always be displayed in tenths of degree Celsius. In all cases there is a strong temperature increase at least for the last thirty years. Moreover, it is apparent (note that the upper and lower panels have different scale) that the overall temperature in the southern regions was larger than in the northern ones, while the general variability was larger in the northern stations. This is suggested also by a simple (mean, standard error) scatter diagram of the 15 stations (see Fig. 3), and essentially confirms the Histalp regionalization. However, differences between SW and SE are not so strong. Note that the different characteristics of our series may be also due to the specific location of the stations, and in particular to altitude: average height above sea level is 354 m for the six NW sites, 372 m for NE, while only 148 for the SW stations and 57 for SE.

There seems not to be any appreciable delay among the series behavior, and the largest cross-correlation for all series is at lag 0, with all positive estimates larger than 0.75. In order to evaluate the regionalization also with respect to dynamics, a hierarchical clustering exercise was performed, using zero-lag cross-correlation as similarity. Regardless the metric (single, average or complete linkage), the first breaking is between north of the Alps and south of the Alps sites, and, on decreasing the distance threshold, the northern cluster splits into the NW and NE subregions, while the southern stations are clustered in two parts: Milan-Turin and Padova-Udine-Verona, corresponding to the subregions SW and SE but with Verona associated to SE. Similar results are obtained if the first difference series are analyzed.

A linear analysis was also performed, fitting to each series a model of the ARIMA class, with the Schwarz order selection criterion. The results are shown in Table 2.

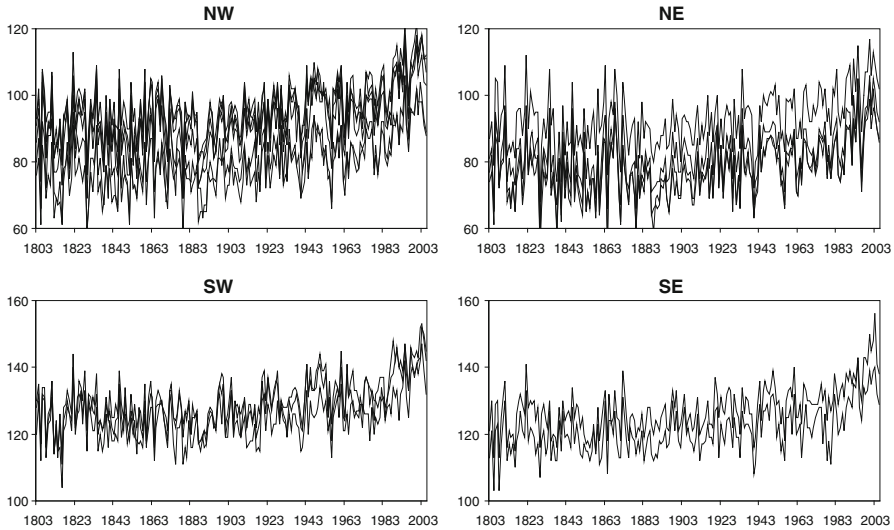


Fig. 2 Time series graphs of the stations in the four subregions

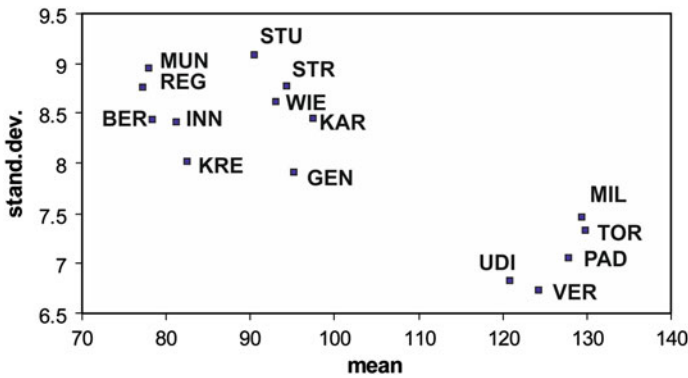


Fig. 3 (Mean, Standard error) scatter diagram

The diagnostic check by Box-Pierce, Jarque-Bera, Keenan nonlinearity and LM-Arch tests was carried out for all models, without finding any significant value at the 0.05 level. It may be immediately seen that the identified structure is the same for all series: integrated first order moving average

$$X_t = X_{t-1} + c + u_t - \theta u_{t-1}.$$

The estimated parameter values are much similar for all of them. The constant term estimates are never significantly different from zero (estimated values are approximately equal to their standard errors), thus implying no clear-cut conclusion about the trend; also, the R^2 values are generally rather small, and slightly different for each region: the average estimated R^2 is 0.3 for NW, 0.19 for NE, and 0.33 in the Southern

Table 2 ARIMA models fitted to the temperature time series

Site	Code	Fitted model	Parameters	R^2
Bern	BER	IMA(1,1)	$c = 0.09, \theta = -0.89^{**}$	0.36
Geneve	GEN	IMA(1,1)	$c = 0.09, \theta = -0.81^{**}$	0.39
Innsbruck	INN	IMA(1,1)	$c = 0.07, \theta = -0.85^{**}$	0.26
Karlsruhe	KAR	IMA(1,1)	$c = 0.09, \theta = -0.89^{**}$	0.32
Kremsmunster	KRE	IMA(1,1)	$c = 0.05, \theta = -0.91^{**}$	0.16
Milano	MIL	IMA(1,1)	$c = 0.09, \theta = -0.83^{**}$	0.39
Munchen	MUN	IMA(1,1)	$c = 0.08, \theta = -0.88^{**}$	0.22
Padova	PAD	IMA(1,1)	$c = 0.10, \theta = -0.81^{**}$	0.38
Regensburg	REG	IMA(1,1)	$c = 0.07, \theta = -0.91^{**}$	0.21
Strasbourg	STR	IMA(1,1)	$c = 0.09, \theta = -0.88^{**}$	0.28
Stuttgart	STU	IMA(1,1)	$c = 0.09, \theta = -0.86^{**}$	0.28
Torino	TOR	IMA(1,1)	$c = 0.08, \theta = -0.78^{**}$	0.35
Udine	UDI	IMA(1,1)	$c = 0.08, \theta = -0.87^{**}$	0.27
Verona	VER	IMA(1,1)	$c = 0.08, \theta = -0.81^{**}$	0.27
Wien	WIE	IMA(1,1)	$c = 0.07, \theta = -0.91^{**}$	0.18

** 1% Significant

regions. The resulting picture is that the temperature variation appears for a large part purely random, and the behavior is similar in structure for all stations.

Though the linear analysis may appear not particularly useful for the present data, especially for forecasting purpose, it may help to detect anomalous data. The recordings are very accurate and have been cleverly homogenized, but unexpected events or sudden temperature changes, both in the local climate itself and in the recording methodology, may lead to influential data that, though formally correct, may severely bias both the identification and the estimation stage of model building, especially if nonlinear and/or nonstationary types are considered. Thus, the outlier check procedure as proposed by [Tsay \(1988\)](#) and [Chen and Liu \(1993\)](#), based on ARIMA models, was applied to each series (using the models in Table 2). Some outliers were detected, but their importance was not so clear: the standardized estimated amplitudes were nearly always smaller than 3.0, therefore they would not be detected if a low to medium sensitivity procedure is chosen. The only three cases requiring attention are the 1822 recording at Turin, the 1833 recording at Geneve, and the 1940 recording at Wien. Both the Turin and Geneve figures are extremely large, a comparable temperature was reached only in the 1990s, and were corrected by subtracting the least squares outlier estimate, equal respectively to 1.60 and 1.65 °. No clear motivation was found in the literature for these old events. On the contrary, the average temperature recorded at Wien station in 1940 was unusually small, extremely inhomogeneous with the values of more than a century. The figure was corrected by adding 2.1° (the least squares outlier estimate). A possible explanation is that during the World War II many stations were relocated from city centre sites to airfields or airports ([Böhm et al. 2010](#), p. 47), and in effect several other stations suffer from unusually small figures in the war years,

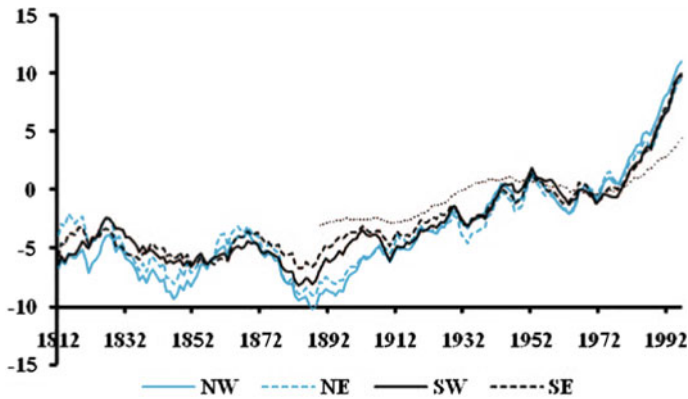


Fig. 4 Average temperature in each subregion and in northern hemisphere (dotted), filtered with a simple moving average, 19 terms. Data are anomalies from the 1961–1990 mean

though not so serious as in the Wien case. The linear models fitted to the Geneva, Turin and Wien corrected series are not much different from those appearing in Table 2.

As a conclusion of the preliminary analysis, an idea of the long-run evolution of the temperature in the four subregions may be obtained by simply considering a 19-term simple moving average of the subregional average temperatures, shown in Fig. 4, where it is apparent that the long-run dynamics have substantially similar features, though with some differences, especially in the past. The northern hemisphere annual average is also shown in Fig. 4 (dotted, dataset Hadcrut3nh of the Climatic Research Unit, University of East Anglia, <http://www.cru.uea.ac.uk>). In particular, it is interesting to note that the slight cooling phase 1950–1970, found for the northern hemisphere, is not evident in the GAR annual data (it was observed that while the summer temperature did undergo that cooling, the winter temperature continued to grow uniformly, resulting in overall increase for annual average data, see e. g. Böhm et al. 2010).

3 Methods

The analysis of the temperature data, in order to highlight possible nonlinear and nonstationary behavior, has been conducted by fitting to each of them a multi-regime threshold model. It assumes that the series is generated by several alternative linear autoregressive equations (the regimes) and the generating process switches from one to another according to the value of an indicator, that may be related to time or to another time series (the driving variable, also a possibly delayed value of the series itself), or both. In the first case we have a nonstationary but linear model (also called structural change, general references are Bai and Perron 1998; Lin and Teräsvirta 1994), while in the second case the model is nonlinear but stationary under suitable choices of the parameter values and is usually called a threshold model; a general reference is Tong (1990). It has also been proposed to allow regime changes determined by a random unobservable indicator, as first suggested by Tong and Lim

(1980) and later formalized by [Hamilton \(1989\)](#); and more recently, to assume that the probability distribution of the random indicator depends on the previous observations of the series, as in [Wu and Chen \(2007\)](#), [Kamgaing et al. \(2009\)](#), [Dueker et al. \(2007\)](#). A non parametric alternative, where the autoregressive coefficients are functions of the driving variable, was proposed by [Chen and Tsay \(1993\)](#), [Cai et al. \(2000\)](#). However for time series with complicated dynamic structure, such models should be considered only as approximations or partial tools for describing only a part of the features of the series. We assume, as in [Rissanen \(2007\)](#) that there is no absolutely correct model, but only models with a better or worse fit. In other words, what we are searching for is not a formalization of the data generating process, but only a parsimonious and efficient representation of the dynamic behavior of the observed data. For this reason we shall not consider tests for nonlinearity or nonstationarity, but will leave the model choice to a fit evaluation. However, tests such as those proposed in [Chan and Tong \(1990\)](#), [Lin and Teräsvirta \(1994\)](#), [Lundbergh et al. \(2003\)](#) may be used for detecting a nonlinear or nonstationary behavior.

Since the temperature series show an instability in time, a possible option is considering models which are simultaneously regime changing both according to time and to a driving variable. This allows the choice of a model inside a set containing the traditional ARIMA, the threshold autoregression, the nonstationary structural change models, and the combination of both. Obviously, identification is much more complicated and this is perhaps the reason why such nonlinear, nonstationary models have been rarely addressed.

[Lundbergh et al. \(2003\)](#) proposed a model based on a first order autoregressive structure with the parameter changing both according to time and to a driving variable alternating between two regimes only, with smooth change (in the same way as smooth transition autoregressive models (STARs) of [Teräsvirta 1994](#)). [Battaglia and Protopapas \(2011\)](#) extended this framework to allow also change of the autoregressive parameters in a piecewise linear fashion (piecewise linear threshold multi-regime model, [Baragona et al. 2004](#)) and proposed a genetic algorithm for building such a model, but also limited to two regimes. Finally, [Battaglia and Protopapas \(2012\)](#) completed this procedure to allow for more than two regimes.

3.1 The models

Consider an autoregressive model where the coefficients change both depending on time and on the level of a delayed observation, in the form of a State Dependent Model of [Priestley \(1988\)](#):

$$X_t = \phi_0(t, X_{t-d}) + \sum_{j=1}^p \phi_j(t, X_{t-d})X_{t-j} + \epsilon_t. \quad (1)$$

We assume that the autoregressive coefficients alternate among a finite number of different regimes, according to the threshold principle.

The original self-exciting autoregressive threshold model (SETAR, see e.g. Tong 1990) is:

$$X_t = \phi_0^{(k)} + \phi_1^{(k)} X_{t-1} + \dots + \phi_p^{(k)} X_{t-p} + \epsilon_t, \quad X_{t-d} \in R_k \tag{2}$$

where $\{R_k\}$ is a partition of the real line, and d is called the delay. A generalization originally proposed by Chan and Tong (1986), and later considered by Teräsvirta (1994, 1998), avoiding discontinuities in the autoregressive parameters, is called STAR, since the transition from one regime to the next is driven by a continuous function (generally a logistic). If r_L denotes the number of regimes and $R_k = (l_{k-1}, l_k]$, $k = 1 \dots, r_L$, the STAR equation may be written:

$$\phi_j(t, X_{t-d}) = \sum_{k=1}^{r_L} \phi_j^{(k)} G_{k-1}(X_{t-d}) \tag{3}$$

where $G_0(x) = 1$ and $G_k(x) = [1 + \exp\{-\gamma_L(x - l_k)\}]^{-1}$. The behavior of the autoregressive coefficients is essentially constant in each regime, with a continuous smooth change between regimes, whose speed is controlled by the constant $\gamma_L (> 0)$. The SETAR model is obtained as a special case of the STAR model, when γ_L tends to infinity.

A different proposal, where the autoregressive coefficients change linearly and continuously with the driving variable X_{t-d} , but with different slope in each regime, is the piecewise linear threshold model (PLTAR, Baragona et al. 2004), described by

$$\phi_j(t, X_{t-d}) = \phi_j^{(0)} + \phi_j^{(1)} X_{t-d} + \sum_{k=2}^{r_L} \phi_j^{(k)} \max(0, X_{t-d} - l_{k-1}) \tag{4}$$

Here the autoregressive coefficient behaves like a linear spline across regimes. The PLTAR may be written in a similar fashion to the STAR letting $S_k(x) = \max(0, x - l_{k-1})$, $S_1(x) = x$, $S_0(x) = 1$ and

$$\phi_j(t, X_{t-d}) = \sum_{k=1}^{r_L+1} \phi_j^{(k)} S_{k-1}(X_{t-d}) \tag{5}$$

Note however that there is an additional parameter here (the linear term in X_{t-d}) for each lag. Therefore the sum over k ranges from 1 to the number of regimes plus 1.

In order to take into account additional nonstationarity, we allow each of the coefficients $\phi_j^{(k)}$ to depend on time also, according to a STAR or PLTAR structure. Let r_T denote the number of regimes in time, and t_j denote the thresholds so that the regimes are defined by the partition $R'_k = (t_{k-1}, t_k]$, where $1 = t_0 < t_1 < \dots < t_{r_T} = N$ (where N is the series length).

We may allow STAR dependence on time using

$$\phi_j^{(k)} = \sum_{i=1}^{r_T} \beta_j(i, k) G'_{i-1}(t), \quad j = 0, \dots, p \tag{6}$$

with $G'_0(t) = 1$, $G'_i(t) = [1 + \exp\{-\gamma_T(t - t_i)\}]^{-1}$, $i = 0, \dots, r_T - 1$.

Alternatively, a time nonstationarity following a PLTAR structure may be defined as

$$\phi_j^{(k)} = \sum_{i=1}^{r_T+1} \beta_j(i, k) S'_{i-1}(t), \quad j = 0, \dots, p \tag{7}$$

where $S'_i(t) = \max(0, t - t_{i-1})/N$, $S'_1(t) = t/N$, $S'_0(t) = 1$.

On combining the different types of dependence on levels of the driving variable, and time, nine different kinds of models result: Stationary, STAR or PLTAR in time, combined with linear, STAR or PLTAR in levels. Denoting $r_L^* = r_L$ for STAR models in levels, $r_L^* = r_L + 1$ for PLTAR, and $r_L^* = 1$ for linear models, and analogously $r_T^* = r_T$ for STAR models in time, $r_T^* = r_T + 1$ for PLTAR in time, and $r_T^* = 1$ for stationary models, the total number of beta coefficients in (6) and (7) is $P = r_L^* r_T^* p$. The resulting model is in the state dependent form (1) with

$$\phi_j(t, X_{t-d}) = \sum_{i=1}^{r_T^*} \sum_{k=1}^{r_L^*} \beta_j(i, k) G'_{i-1}(t) G_{k-1}(X_{t-d})$$

or with S' substituted to G' and/or S substituted to G . Since the final model equation is linear in the P elementary parameters $\beta_j(i, k)$, they may be simply estimated by means of least squares.

Summarizing, a multi-regime model is identified by: its type in levels (linear, STAR or PLTAR); its type in time (stationary, STAR or PLTAR); the delay d of the driving variable; the order p ; the thresholds in levels l_k and in time t_k ; and if needed, logistic speed coefficients γ_L and γ_T . Conditional on the values of all these structural parameters, the remaining elementary parameters $\beta_j(i, k)$ which drive the evolution of the autoregressive weights, are estimated by least squares.

Ergodicity and consistency issues are addressed in Battaglia and Protopapas (2011) for the case of two regimes, and conditions on parameters are given that ensure ergodicity and consistent estimation of both the threshold and the autoregressive coefficients. The multi-regime case has not been fully investigated yet.

3.2 The optimization algorithm

The identification of a multi-regime model requires the selection of the solution that possesses the best properties, out of a large space of candidates. Such kind of problems

have been recently addressed by means of meta-heuristic methods, since these methods allow to explore the space more efficiently than simple enumeration. Among the several types of meta-heuristic methods proposed in statistical applications, the genetic algorithm is most suitable because the space of solutions is discrete, while others, e.g. differential evolution (Price et al. 2005) or particle swarm (Kennedy and Eberhart 2001) are preferred for continuous spaces. In addition to several statistical identification problems, e.g. autoregressive moving average model fitting (Gaetan 2000), outlier detection (Crawford and Wainwright 1995), variable selection in regression (Chatterjee et al. 1996), the genetic algorithm was suggested specifically for threshold models by Wu and Chang (2002), Davis et al. (2006). A review of applications of meta-heuristic methods in Statistics is in Baragona et al. (2011).

A genetic algorithm for building multi-regime models has been proposed by Battaglia and Protopapas (2012) and is employed here for the temperature series. Given a time series and an identification criterion, it selects the best model type (in time and in levels), number of regimes, order and thresholds in time and levels, and provides least squares estimates of the model coefficients.

Genetic algorithms are optimization tools which mimic the evolution of a living population towards fitness to the natural environment. The key feature is manipulation of a population of individuals, each of which is characterized by a specific *chromosome*. Chromosomes are coded as strings of characters of a given length, and each chromosome represents a feasible solution of the optimization problem. The single units of the string are called it genes. The link between the genetic algorithm and the problem is provided by the it fitness function, which establishes a mapping from the chromosomes to a set of real positive numbers. The fitness function is a monotonic transform of the objective function, and the greater the fitness, the better is the adaptation of that individual.

The algorithm is iterative and each iteration yields a new population starting from the existing one, by applying three evolutionary operators: it selection, mutation, crossover. Selection designates which member of the old population will survive in the new one, by means of a random choice where each individual has a probability of being selected proportional to its fitness. Mutation provides the chance for each gene to change its value with a given fixed probability, and allows new individuals to arise. A similar effect is produced by the crossover operator, which allows pairs of chromosomes to exchange mutually corresponding sub-strings, again with a random mechanism.

We implemented a hybrid genetic algorithm where the chromosomes encode the structural parameters (type of model, number of regimes, order, thresholds), while the autoregressive parameters $\beta_j(i, k)$ are estimated by least squares inside the fitness evaluation stage.

The genetic algorithm was introduced by Holland (1975). A complete treatment of theory and convergence properties is Reeves and Rowe (2003). Further details of the algorithm used here may be found in Battaglia and Protopapas (2012).

The most important step is the choice of the fitness function. In model building applications, it is customary to evaluate models by means of an identification criterion: if model M has P estimated parameters, the values

$$IC(M) = N \log \hat{\sigma}_M^2 + cP \quad (8)$$

where N is the series length and $\hat{\sigma}_M^2$ the estimated residual variance, are evaluated and the model with minimum value is selected. The constant c may be chosen according to different ideas (see e.g. [Bhansali and Downham 1977](#)), and leads to different identification criteria (for example $c = 2$ leads to the popular AIC criterion). The fitness function must be related to the identification criterion through a monotone decreasing transformation, and positiveness has to be ensured. We use a simple negative exponential transformation:

$$\text{fitness}(M) = \exp\{-IC(M)/N\} = \hat{\sigma}_M^{-2} \exp\{-cP/N\}.$$

The value used here for the constant c equals 3.0, because in this way the criterion behavior is equivalent to a test of linearity-stationarity against two-regimes alternatives proposed by [Lundbergh et al. \(2003\)](#), see [Battaglia and Protopapas \(2011\)](#).

The asymptotic properties of the criteria may be evaluated basing on results of [Sin and White \(1996\)](#), [Rivers and Vuong \(2002\)](#), [Marcellino and Rossi \(2008\)](#) and have been derived, for the case of two regimes, in Theorem 5 of [Battaglia and Protopapas \(2011\)](#). In this case, the criterion (8) is consistent if $c = o(N)$ and $c/\sqrt{N} \rightarrow \infty$ as $N \rightarrow \infty$.

The algorithm parameters we chose are as follows (more details may be found in [Battaglia and Protopapas 2012](#)): population size 50, number of generations 1,500, probability of mutation 0.025, probability of cross-over 1.0, number of regimes up to 4, order and delay up to 7. Finally, to prevent unstable estimates of the autoregressive coefficients, only regimes in time and level containing no less than 20 observations were evaluated.

In order to compare computational costs, observe that finding the exact optimal solution would require a complete enumeration of all possible models that, allowing a precision of 0.01° for the temperature thresholds, needs more than 10^8 fitness evaluations. The genetic algorithm implementation used in our case, though allowing for 1, 500 generations (an unusually large number, chosen for ensuring close convergence in any case) requires only 7.5×10^3 fitness evaluations.

4 Results

For each of the 15 annual temperature series, a genetic algorithm searched for the best nonlinear nonstationary models of the class described above, evaluating each possible model according to the criterion:

$$IC(M) = N \log \hat{\sigma}_M^2 + 3 \text{ (number of parameters)}$$

which accounts both for fitting error and parsimony. Therefore, the "optimal" found models are simply those which minimize $IC(M)$, they are not assumed at all to reproduce the unknown data generating process, but simply the best fitting solution according to the selected criterion: they are not it climatic models.

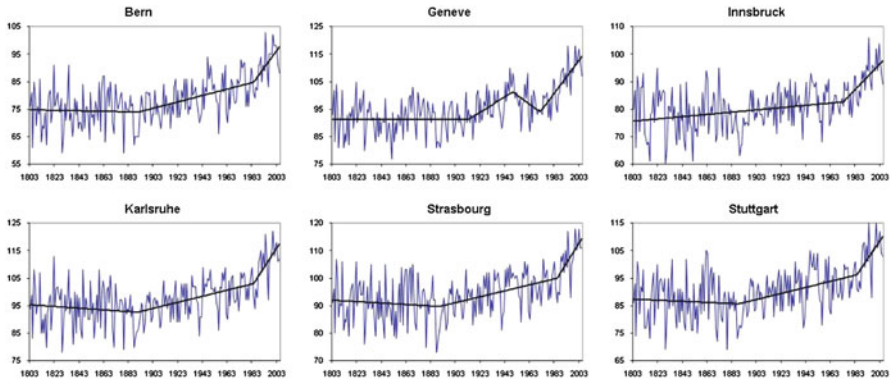


Fig. 5 Original and fitted data, stations of the North–West subregion

The first conclusion of our study is that the selected models for all stations are exactly of the same type: linear in levels, and piecewise linear threshold in time. For 13 stations the selected order was zero, and the model may be written:

$$X_t = \sum_{i=1}^{r_T} \phi_i S'_{i-1}(t) + \epsilon_t. \tag{9}$$

Only for Milan and Turin the selected model was of the same type as before, but with order 1:

$$X_t = \sum_{i=1}^{r_T} \phi_i S'_{i-1}(t) + \sum_{i=1}^{r_T} \beta_i S'_{i-1}(t) X_{t-1} + \epsilon_t. \tag{10}$$

The best model type being uniform, it is suggested again that all series exhibit fundamentally similar features, that may be reproduced by a similar structure. The series do not show a strong autoregressive nature, so the proposed model is of order zero. Though this is contradicted by the Milan and Turin results, it may be observed that, when dealing with nonlinear and nonstationary structures, model multiplicity is a serious issue, and we found that for these two stations the differences between the fitness of the zero-order and first-order models are small. The original and fitted data of the single stations are shown, separately for each subregion, in Figs. 5, 6, 7 and 8. The details of the number of regimes and thresholds may be found in Table 3. The usual diagnostic checking tools were computed, and no reason for rejecting any of the proposed model was found. An overall comparison with the linear IMA models may be based on the R^2 measures: our models ensure an increase in R^2 of about 0.1, uniformly for all subregions and series.

The fact that the genetic algorithm never selected a nonlinear structure indicates that, at least at an annual rate and on the two-century time span examined, the climate feedback processes were not able to introduce strong nonlinearities in the temperature

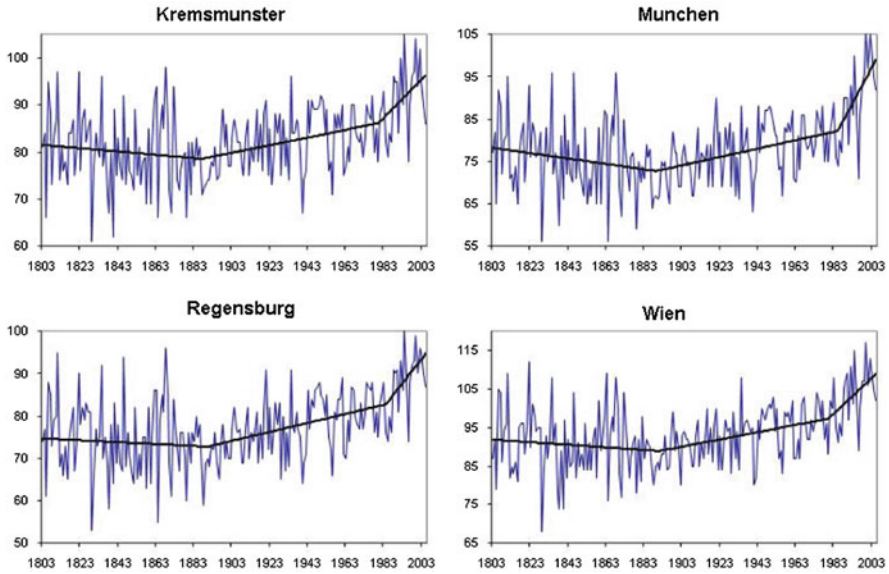


Fig. 6 Original and fitted data, stations of the North–East subregion

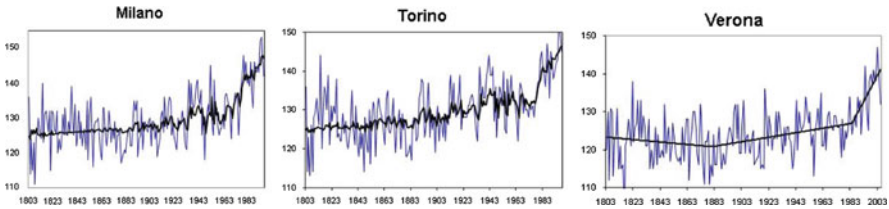


Fig. 7 Original and fitted data, stations of the South–West subregion

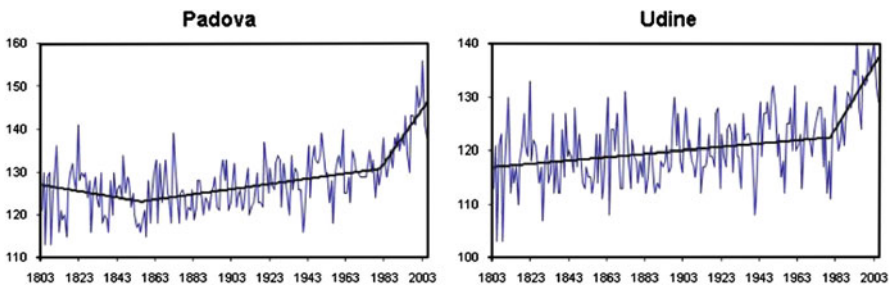


Fig. 8 Original and fitted data, stations of the South–East subregion

evolution, or the nonlinear effects of positive and negative feedbacks compensate each other.

Essentially, the models split the entire period (1803–2005) in sub-periods (the regimes) and inside each of them the behavior is described by a simple linear trend. The most frequent result (in 10 stations out of 15) is a model with three regimes,

Table 3 Fitted models

Site	Code	Subregion	Order	Regimes	Thresholds (year)	R^2
Bern	BER	NW	0	3	1891, 1984	0.44
Geneve	GEN	NW	0	4	1913, 1949, 1972	0.51
Innsbruck	INN	NW	0	2	1973	0.34
Karlsruhe	KAR	NW	0	3	1889, 1984	0.41
Kremsmunster	KRE	NE	0	3	1888, 1980	0.26
Milano	MIL	SW	1	2	1975	0.49
Munchen	MUN	NE	0	3	1889, 1984	0.33
Padova	PAD	SE	0	3	1855, 1981	0.48
Regensburg	REG	NE	0	3	1892, 1983	0.32
Strasbourg	STR	NW	0	3	1889, 1985	0.37
Stuttgart	STU	NW	0	3	1888, 1984	0.37
Torino	TOR	SW	1	2	1983	0.46
Udine	UDI	SE	0	2	1980	0.39
Verona	VER	SW	0	3	1880, 1984	0.37
Wien	WIE	NE	0	3	1891, 1980	0.29

In all cases the type is linear in levels and PLTAR in time

the first from the beginning to about 1890, the second up to 1980, and the last one involving the most recent years. For four stations (Innsbruck, Milano, Torino, Udine) only the second threshold is selected, and only two regimes are proposed, with change around 1980. Finally, the behavior of Geneve is peculiar, and the best fitted model has four regimes in time, with thresholds in the years 1913, 1949, 1972. This might be due to the fact that the Geneve data show the slight cooling phase 1950–1970, already observed for the global northern hemisphere, but not evident in the other stations. All models fit the macroscopic change of temperature evolution, which is located in time between 1980 and 1984 (only exceptions are Geneve, Innsbruck and Milano for which the threshold is located a few years before).

For the period 1800–1980 the obtained models suggest a very slow change of the average temperature, in some cases uniform (Innsbruck, Milano, Torino, Udine) and in the other cases consisting in an initial slow decrease followed by a moderate increase. To be more precise, the slopes of the fitting model for each station and regime are shown in Table 4 for the 3-regime models, and in Table 5 for the 2-regime models. Since the slopes are sums of estimated phi coefficients in (9), their significance may be evaluated approximately by means of standard least squares theory.

All the slope coefficients result significantly different from zero, at least at confidence level 0.05, with the exception of those for the regime 1800–1890, for all stations of Table 4. It indicates a substantial temperature stability in the nineteenth century, followed by a moderate increase, that becomes a fast growth after 1980. Comparatively, while the initial increase is not so different among the stations, the growth after 1980 shows different speed, and in particular the recording of the NW and southern regions have an average increase of about 0.06° per year, while the NE stations have a slower

Table 4 Slopes in each regime of the fitted models. Models with three regimes

Site	1st regime		2nd regime		3rd regime
Bern	-1.8	(1891)	23.6	(1984)	125.1
Karlsruhe	-6.3	(1889)	22.3	(1984)	141.8
Kremsmunster	-7.1	(1888)	16.5	(1980)	83.2
Munchen	-12.9	(1889)	20.1	(1984)	170.0
Padova	-15.1	(1855)	12.4	(1981)	127.3
Regensburg	-4.5	(1892)	21.5	(1983)	115.7
Strasbourg	-5.2	(1889)	21.7	(1985)	140.2
Stuttgart	-4.1	(1888)	22.2	(1984)	134.1
Verona	-6.5	(1880)	12.1	(1984)	136.7
Wien	-6.8	(1891)	19.1	(1980)	94.1

increase around 0.05° per year. Notable exceptions are the Innsbruck series, whose temperature grows a bit less (being more similar to NE rather than NW evolution) and especially Munchen, whose estimated rate of increase after 1980 is the largest one, and more than 0.08° per year. The reason for this anomaly is very likely anthropic.

The choice of the constant $c = 3$ in the evaluation criterion (8) appears a satisfying compromise between good fitting and parsimony, and is supported by theoretical reasons as mentioned in Sect. 3.2; however, we considered also a different fitness, based on the Schwarz criterion choosing $c = \log N = 5.3$, therefore more parsimonious, and repeated the entire exercise. For all series the best found models still are linear in levels and PLTAR in time, with order zero, uniformly with two regimes and thresholds between 1975 and 1980.

Finally, we have evaluated the forecasting performance of our models. Since their parameters change according to different time regimes, they are more able to adapt to a change in the evolution features, and we expect their forecast ability to improve over ARIMA models. We computed out-of-sample forecasts for all stations with both kind of models. In order to keep homogeneity, the out-of-sample forecasts were computed only for the years for which data from the Histalp file are available for all stations (2006 and 2007, but no figures for Padova). The typical behavior, obtained for all stations with moderate differences, is exemplified in Fig. 9, reporting the Stuttgart case. Together with the fitted data (solid line for linear-PLTAR, dotted line for IMA(1,1) model), the forecasts with origin 2005 and lead times up to ten years are shown; solid circles are the 2006 and 2007 out-of-sample actual temperatures. The nonstationary model has a better fit especially in the most recent years, and predicts a definitely faster long-run increase than the integrated moving average model; the latter tends to under-estimate the 2006 and 2007 data. The overall results are summarized in Table 6 where the global forecasting errors for all stations appear (Padova is excluded since the out-of-sample data were not available). Looking at both absolute and square errors, the forecasts obtained with the linear-PLTAR model are considerably more precise than those of the moving average model. The trends for the future, deriving from our models, seem to indicate an even more serious warming than in the recent past, and

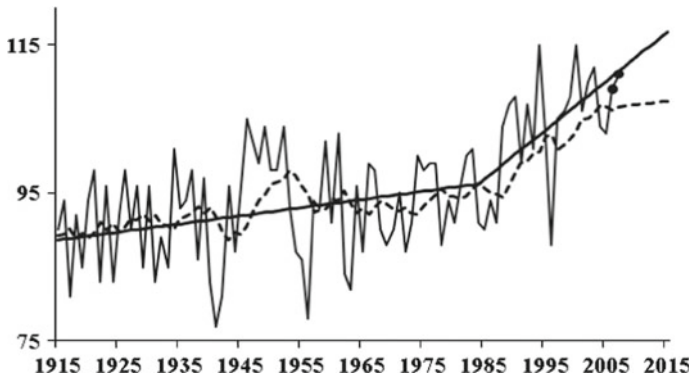


Fig. 9 Temperature at Stuttgart. Models fitted through data up to 2005: linear-PLTAR (solid line), IMA(1,1) (dotted line). Forecasts with origin 2005 and lead times 1–10 years. Solid circles are 2006 and 2007 actual temperatures

Table 5 Slopes in each regime of the fitted models. Models with two regimes

Site	Slope 1st regime		Slope 2nd regime
Innsbruck	8.3	(1973)	84.7
Milano ^a	7.8	(1975)	117.1
Torino ^a	10.0	(1983)	135.1
Udine	6.2	(1980)	120.3

^a Model with order 0 was fitted for this calculation

Table 6 Out-of-sample forecast errors for 2006 and 2007, average on all stations

Year Model	2006		2007		2006 and 2007	
	IMA	PLTAR	IMA	PLTAR	IMA	PLTAR
Mean error	2.5	-1.3	6.4	1.9	4.5	0.3
Mean absolute error	2.7	1.9	6.4	3.1	4.5	2.5
Mean square error	11.1	5.5	56.5	15.1	33.8	10.3

suggest that the EU policy target of an increase limited to 2° over the pre-industrial average, will be very hard to meet without an effective action against greenhouse gas, and seems already missed for the GAR. Indeed, the difference between the mean temperature 2000–2005 and the average 1850–1899 (as representative of the pre-industrial era) is already larger than 2° for many of the Histalp stations (see Table 7).

5 Conclusions

Based on the results of our analysis we are now in a position to give some indication about the three questions we posed initially.

Table 7 Increase of the average 2000–2005 temperature over the pre-industrial average

North–West		North–East		South–West		South–East	
BER	20.9	KRE	16.3	MIL	21.5	PAD	22.2
GEN	21.9	MUN	25.9	TOR	19.8	UDI	17.6
INN	20.0	REG	19.8	VER	18.5		
KAR	22.2	WIE	18.7				
STR	23.4						
STU	22.4						

When did the temperature increase start? Was the rate of increase uniform, or significantly varying? There is evidence that a very moderate temperature increase started at the beginning of the twentieth century, nearly uniformly for all stations, at a rate of about 0.01° per year. But around 1980 the temperature growth speeded up dramatically, exceeding the rate of 0.05° per year, and there is no recent evidence of any slowing down. It is well known that Europe warmed more than the rest of the northern hemisphere, and in the GAR the temperature increase was even more serious. This region appears already beyond the target of no more than 2° increase over the pre-industrial era, stated by the United Nations Framework Convention on Climate Change in the 2009 Copenhagen Accord.

Are there relevant differences in the temperature evolution among the different stations? The Histaip project findings (based on several climatic indicators) suggested evidence of four different subregions, NW, NE, SW and SE. Stations in the southern regions show a larger mean and a smaller variability in time. Concerning the warming, the fundamental behavior is similar for all stations, but some differences arise. A slight decrease of the temperature during the nineteenth century is recognized by our models for all cases except Innsbruck, Milano, Torino and Udine, though such a decrease cannot be judged statistically significant. The change from the slight decrease to the moderate increase regime is located in our models at nearly simultaneous dates, between 1889 and 1891, with few exceptions. And the start of the fast growth is also synchronous about 1980, but with some differences: earlier for Milano and Innsbruck (1975 and 1973 respectively), in 1980 or 1981 for Kremsmunster, Padova and Wien, and between 1983 and 1985 for the other stations, without a clear-cut reason relating to subregions or other climate features. The temperature growth rates in the last regime are also slightly different, being slower in the NE stations, and larger in the South and NW subregions. A remarkably inhomogeneous behavior is found for Munchen, where the temperature estimated rate of increase in the last thirty years was considerably larger than all other stations.

Finally, our analysis does not support the presence in the data of any relevant non-linear component due to feedback processes.

Concluding, we have seen that the use of relatively simple threshold models, allowing both for nonlinearity and nonstationarity, may help to describe more clearly data with non uniform and changing dynamic features.

Acknowledgments This research was supported by the EU Marie Curie Research and Training Network it COMISEF Computational Methods in Statistics Econometrics and Finance. The authors thank Reinhard Böhm and Manfred Ganekind of Zentralanstalt für Meteorologie und Geodynamik, Wien for kindly providing Swiss data, and the anonymous referees for useful comments.

References

- Auer I, Böhm R, Jurkovic A, Lipa W, Orlik A, Potzmann R, Schöner W, Ungersböck M, Matulla C, Briffa K, Jones PD, Efthymiadis D, Brunetti M, Nanni T, Maugeri M, Mercalli L, Mestre O, Moisselin J-M, Begert M, Müller-Westermeier G, Kveton V, Bochnicek O, Stastny P, Lapin M, Szalai S, Szentimrey T, Cegnar T, Dolinar M, Gajic-Capka M, Zaninovic K, Majstorovic Z, Nieplova E (2007) HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region. *Int J Climatol* 27:17–46
- Bai J, Perron P (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66:47–78
- Baragona R, Battaglia F, Cucina D (2004) Fitting piecewise linear threshold autoregressive models by means of genetic algorithms. *Comput Stat Data Anal* 47:277–295
- Baragona R, Battaglia F, Poli I (2011) Evolutionary statistical procedures. Springer, Berlin
- Battaglia, F, Protopapas MK (2011) Time-varying multi-regime models fitting by genetic algorithms. *J Time Ser Anal* 32:237–252
- Battaglia, F, Protopapas MK (2012) Multi-regime models for nonlinear nonstationary time series. *Comput Stat* 27:319–341
- Bhansali RJ, Downham DY (1977) Some properties of the order of an autoregressive model selected by a generalization of Akaike's EPF criterion. *Biometrika* 64:547–551
- Böhm R, Auer I, Brunetti M, Maugeri M, Nanni T, Schöner W (2001) Regional temperature variability in the European Alps: 1760–1998 from homogenized instrumental time series. *Int J Climatol* 21:1779–1801
- Böhm R, Jones PD, Hiebl J, Frank D, Brunetti M, Maugeri M (2010) The early instrumental warm-bias: a solution for long central European temperature series 1760–2007. *Clim Change* 101:41–67
- Cai Z, Fan J, Yao Q (2000) Functional coefficient regression models for nonlinear time series. *J Am Stat Assoc* 95:941–955
- Chan KS, Tong H (1986) On estimating thresholds in autoregressive models. *J Time Ser Anal* 7:179–190
- Chan KS, Tong H (1990) On the likelihood ratio tests for threshold autoregression. *J R Stat Soc B* 52:469–476
- Chatterjee S, Laudato M, Lynch LA (1996) Genetic algorithms and their statistical applications an introduction. *Comput Stat Data Anal* 22:633–651
- Chen C, Liu LM (1993) Joint estimation of model parameters and outlier effects in time series. *J Am Stat Assoc* 88:284–297
- Chen R, Tsay R (1993) Functional coefficient autoregressive models. *J Am Stat Assoc* 88:298–308
- Crawford KD, Wainwright RL (1995) Applying genetic algorithms to outlier detection. In: Eshelman LJ (ed) Proceedings of the sixth international conference on genetic algorithms. Morgan Kaufmann, San Mateo, CA, pp 546–550
- Davis R, Lee T, Rodriguez-Yam G (2006) Structural break estimation for nonstationary time series models. *J Am Stat Assoc* 101:223–239
- Dueker MJ, Sola M, Spagnolo F (2007) Contemporaneous threshold autoregressive models: estimation, testing and forecasting. *J Econom* 141:517–547
- European Environment Agency (2008) Impacts of Europe's changing climate—2008 indicator-based assessment. EEA report no. 4/2008. <http://www.eea.europa.eu>. ISBN 978-92-9167-372-8
- Gaetan C (2000) Subset ARMA model identification using genetic algorithms. *J Time Ser Anal* 21:559–570
- Hamilton J (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384
- Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor
- IPCC Core Writing Team, Pachauri RK, Reisinger A (eds) (2007) Climate change 2007: synthesis report. Contribution of working groups I, II and III to the fourth assessment report of the intergovernmental panel on climate change. IPCC, Geneva. ISBN 92-9169-122-4
- Kamgaing JT, Ombao H, Davis RA (2009) Autoregressive processes with data-driven regime switching. *J Time Ser Anal* 30:505–533

- Kennedy J, Eberhart R (2001) *Swarm intelligence*. Morgan Kaufman, San Mateo
- Lin C, Teräsvirta T (1994) Testing the constancy of regression parameters against continuous structural change. *J Econom* 62:211–228
- Lundbergh S, Teräsvirta T, van Dijk D (2003) Time-varying smooth transition autoregressive models. *J Bus Econ Stat* 21:104–121
- Marcellino M, Rossi B (2008) Model selection for nested and overlapping nonlinear, dynamic and possibly mis-specified models. *Oxford Bull Econ Stat* 70(Suppl):867–893
- Price KV, Storn R, Lampinen J (2005) *Differential evolution, a practical approach to global optimization*. Springer, Berlin
- Priestley MB (1988) *Non-linear and non-stationary time series analysis*. Academic Press, London
- Reeves CR, Rowe JE (2003) *Genetic algorithms—principles and perspective a guide to GA theory*. Kluwer, London
- Rissanen J (2007) *Information and complexity in statistical models*. Springer, Berlin
- Rivers D, Vuong Q (2002) Model selection tests for nonlinear dynamic models. *Econom J* 5:1–39
- Sin C-Y, White H (1996) Information criteria for selecting possibly misspecified parametric models. *J Econom* 71:207–225
- Teräsvirta T (1994) Specification estimation and evaluation of smooth transition autoregressive models. *J Am Stat Assoc* 89:208–218
- Teräsvirta T (1998) Modeling economic relationships with smooth transition regression. In: Ullah A, Giles DEA (eds) *Handbook of applied economic statistics*. Marcel Dekker, New York pp 507–552
- Tong H (1990) *Non linear time series: a dynamical system approach*. Oxford University Press, Oxford
- Tong H, Lim K (1980) Threshold autoregression, limit cycles and ciclical data. *J R Stat Soc Ser B* 42: 245–292
- Tsay RS (1988) Outliers, level shifts and variance changes. *J Forecast* 7:1–20
- Wu B, Chang CL (2002) Using genetic algorithms to parameters (d,r) estimation for threshold autoregressive models. *Comput Stat Data Anal* 38:315–330
- Wu S, Chen R (2007) Threshold variable determination and threshold variable driven switching autoregressive models. *Stat Sin* 17:241–264
- Zebisch M, Zimmermann P, Böhm R, Schöner W (2009) Climate change in the Alps. In: *Regional climate change and adaptation—the Alps facing the challenge of changing water resources*. European Environment Agency report no. 8/2009. <http://www.eea.europa.eu>. ISBN 978-92-9213-006-0