Theor Appl Climatol
DOI 10.1007/s00704-015-1480-4

ORIGINAL PAPER

Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms

S. Salcedo-Sanz[1] · R. C. Deo[2] · L. Carro-Calvo[1] · B. Saavedra-Moreno[1]

Received: 3 February 2015 / Accepted: 17 April 2015

Abstract Long-term air temperature prediction is of major importance in a large number of applications, including climate-related studies, energy, agricultural, or medical. This paper examines the performance of two Machine Learning algorithms (Support Vector Regression (SVR) and Multi-layer Perceptron (MLP)) in a problem of monthly mean air temperature prediction, from the previous measured values in observational stations of Australia and New Zealand, and climate indices of importance in the region. The performance of the two considered algorithms is discussed in the paper and compared to alternative approaches. The results indicate that the SVR algorithm is able to obtain the best prediction performance among all the algorithms compared in the paper. Moreover, the results obtained have shown that the mean absolute error made by the two algorithms considered is significantly larger for the last 20 years than in the previous decades, in what can be interpreted as a change in the relationship among the prediction variables involved in the training of the algorithms.

1 Introduction

In climate change detection and attribution problems (D&A), there are basically two main types of models that have been considered in the literature: GCM and statistical-based approaches. The GCMs are used to understand the dynamics of the physical components of the climate system, for obtaining global temporal and spatial patterns of change (usually referred to as "fingerprints"), and to make projections based on future greenhouse gas and aerosol forcing (Cramer et al. 2013). These models are applied for seasonal to decadal climate change attribution problems. Currently, the Coupled Model Inter-comparison Project Phase 5, developed by the WCRP's Working Group on Coupled Modeling (WGCM http://cmip-pcmdi.llnl.gov/cmip5/), provides a multi-model context for carrying out coordinated climate model experiments, especially well suited for D&A problems. On the other hand, statistical-based models try to obtain evidences of externally driven climate change, minimizing the use of complex climate models: they are usually easier and have low computational burden compared to GCMs, and different studies using these models have shown coincident results with those by GCM models. Different statistical-based approaches have been recently proposed, such as some methods developed in the econometrics literature (Kaufmann and Stern 1997), so called cointegration methods (Kaufmann et al. 2011; Kaufmann and Stern 2002) that analyzed the relationships between stationary and non-stationary time series or regression-based approaches (Stone and Allen 2005; Douglass et al. 2004) to evaluate time series properties of temperature data.

One alternative to the physical model is the Machine Learning (ML) approach that is being experimented in a wide variety of climate change studies or prediction problems. For example, a study conducted for the Queensland region that compared rainfall predictions using ML approaches with the official model for climate forecasting, Predictive Ocean Atmosphere Model for Australia

✉ S. Salcedo-Sanz
sancho.salcedo@uah.es

1 Department of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain

2 School of Agricultural Computational and Environmental Sciences, University of Southern Queensland, Springfield, Australia

Published online: 10 May 2015

(POAMA) has demonstrated notable improvement in prediction capability of the ML modeling framework (Abbot and Marohasy 2012, 2014; Luk et al. 2000). In that study, the forecasted rainfall for three geographically distinct regions within Queensland using neural network models demonstrated superior performance compared to forecasts generated by the POAMA. In this article, we consider two different ML approaches, Support Vector Regressions (SVR) and Multi-layer Perceptron (MLP) as statistically-based models, in a problem of monthly mean temperature prediction in eight observational stations in Australia and two in New Zealand, with implications of detecting possible climate change signals in this region. Both algorithms are well-established methodologies in computational science, that have been previously applied to different problems in temperature prediction: In the case of the MLP, this methodology has been successfully applied to different temperature prediction scenarios and countries (Chithra et al. 2014; Dombayc and Gölcü 2009; Tasaduqq et al. 2002; Smith et al. 2007, 2009; Ustaoglu et al. 2008). The SVR has been more recently applied to a number of air temperature prediction problems (Chevalier 2008; Chevalier et al. 2011; Kadu et al. 2012; Mellit et al. 2013; Paniagua-Tineo et al. 2011; Ortiz-García et al. 2011; Radhika and Shashi 2009), and better results than the MLP have been reported in the majority of these works. In spite of these important previous works on the application of ML methods to air temperature prediction, the connection of the problem with D&A of climate change patterns and the usefulness of ML techniques in this task is still a point to be evaluated and studied. Moreover, the application of SVR for prediction of atmospheric variable in Australia is a new endeavour although a plethora of research been conducted on conventional ML approaches such as artificial neural networks (e.g., Abbot and Marohasy 2012; Abbot and Marohasy 2014).

Specifically, this paper provides a complete description of these two ML techniques in a problem of mean air temperature prediction. We have used these two models (MLP and SVR) with only six input (predictive) variables to predict the monthly temperature t_n: temperature in the previous month t_{n-1}, two dummy variables to model the year temperature cycle: $d_1 = \sin\left(\frac{2 \cdot \pi \cdot n}{12}\right)$ and $d_2 = \cos\left(\frac{2 \cdot \pi \cdot n}{12}\right)$ ($n = 0, \ldots, 11$, depending on the month of the year to be evaluated), and three climate mode indices related to temperature variation in the Pacific and Indian Ocean regions: Southern Oscillation Index (SOI), Indian Ocean Dipole (IOD), and Pacific Decadal Oscillation (PDO) that have been recently proven useful in temperature prediction problems (Daneshmand et al. 2014; Saji et al. 2005; Abbot and Marohasy 2012, 2014; Mekanik et al. 2013). The aim of this investigation is then twofold: a first objective consists of evaluating the capability of the ML techniques considered in the problem of monthly air temperature prediction from the input variables considered. An accurate long-term air temperature prediction is of undoubted interest, with different important applications in climate (Jacobs et al. 2013; You et al. 2013), energy (Bertini et al. 2010), agriculture (Smith et al. 2009), or medicine (Garske et al. 2013; Williams et al. 2012; Xu et al. 2014). We will show that the SVR is able to outperform all the compared algorithms, whereas the MLP performs approximately as a multi-linear regression algorithm. A second objective of this research consists of a deeper analysis of the results provided by the ML and alternative algorithms considered, in order to detect possible patterns of climate change or climatic variability. The possibility of carrying out an analysis of the relationship among the input variables considered in the algorithms is another important aspect of the considered methodology.

The structure of the rest of the paper is as follows: in the next section, we describe in detail the available temperature data to conduct this research. We also detail the study's methodology that follows well-established methods in ML for data partition and algorithms' tuning. In Section 3, we summarize the description of the SVR and the MLP neural network used in this study. In Section 4, we report the results obtained with both algorithms in the problem of air temperature prediction for the different measuring stations considered, and also we show a pattern of climate change spotted by both algorithms when their prediction error is depicted along the time. The conclusion of the paper, mainly described in Section 5, is that ML algorithms can be used to detect climate change patterns and analyze them by considering the relationship of input variables with the prediction error made by the algorithms in specific problems.

2 Data sources and methodology

Our study capitalized the monthly mean temperature dataset from the Australian Bureau of Meteorology (BOM), which has been made publicly available at ftp://ftp.bom.gov.au/anon/home/ncc/www/change/HQannualT (Collins 2006). These data were initially created from in-site temperature measurements. Quality control tests were conducted by Torok and Nicholls (1996) to produce a homogenised or high-quality dataset to enable reliable analyses of climate trends and variability at annual and decadal timescales. Each station record was adjusted for discontinuities caused by changes in site location and exposure, and other known data problems. Such discontinuities can be as large, or larger than, real temperature changes (e.g., the approximate 1 °C drop in maximum temperatures associated with the switch to Stevenson screen exposure) and consequently confound the true long-term trend. Generally, the high-quality records were homogenised from 1910, by which time most stations are believed to have been equipped with the current standard instrument shelter. Two hundred twenty-four temperature records were reconstructed to an acceptable standard, 181 of which were identified as being non-urban. Furthermore, the work of Della-Marta et al. (2004) updated this dataset using a suite of homogeneity techniques to assess whether records were likely to have been contaminated by urban heat effects and to use available comparison observation data in cases where a station had moved. Consequently, the original temperature dataset has significantly been improved for estimating changes and variability of Australian climate over the past century, as exemplified in a number of studies (e.g., Alexander et al. 2007; Nicholls et al. 1996).

Despite the temperature records initially available in the 1850s, the use of now-standardized Stevenson screen for the housing of thermometers cannot be assumed for pre-1900 data (Nairn and Fawcett 2013; Nicholls et al. 1996). Therefore, in this study, we focus on the prediction of monthly mean air temperature (Tmean) for a total of eight stations in Australia, three urban stations (Sydney, Melbourne and Adelaide) with data from 1900 to 2010 and five rural stations (Alice Spring, Bathurst Agricultural, Cape Otway Lighthouse, Charleville Aero, and Cobar Mo) with data from 1910 to 2010, and other two stations in New Zealand (Queenstown and Napier), with data from 1930 to 2010, see Table 1. Figure 1 shows the location of the different stations considered. Note that for the urban stations, the incidence of top-ranked heatwave events are very common at the very start and end of the data records for all three cases (Coates 1996), but these have been particularly pronounced for Melbourne and Adelaide (Nairn and Fawcett 2013). However, the correction of temperatures by Della-Marta et al. (2004) attempted to address urban heat effects on temperatures records at these cities. Additionally, the use of supplementary data for the rural stations showed little impacts of urban heating on temperature data for the capital city stations over the period of study.

We have trained the SVR and the MLP with the high-quality mean temperature dataset in each observational station using two different training and test sets. First, we consider a first training set corresponding with months from January 1900 (or the closest date available) up to December 1930 (the test set is the remaining data from 1931 up to 2010). We have called this set as 30/70 case. Second, we consider a longer training set from 1900 up to 1970, with test set from 1971 up to 2010 (70/30 case). The performance of these ML approaches is evaluated by using the

Table 1 Characteristics of the observational stations in Australia (three urban and five rural or regional) and New Zealand (two rural stations), and mean temperature (Tmean) in the complete set of data available (1900–2010 for urban and 1910–2010 for rural stations in Australia and 1930–2010 in New Zealand's stations)

Station	ID	Location	Elevation (m)	Tmean (°C)
1. Sydney (Urban)	066062	33.86 S, 151.23 E	39	18.00
2. Melbourne (Urban)	086071	37.81 S, 144.97 E	31.2	15.2
3. Adelaide (Urban)	023090	34.92 S, 138.62 E	48	16.90
4. Alice Spring (Regional)	015590	23.80 S, 133.89 E	546	20.63
5. Bathurst (Regional)	063005	33.43 S, 149.56 E	713	13.11
6. Cape Otway (Regional)	090015	38.86 S, 143.51 E	82	13.43
7. Charleville (Regional)	044021	26.41 S, 146.26 E	302	20.57
8. Cobar (Regional)	048027	31.48 S, 145.83 E	260	18.65
9. Queenstown (New Zealand,Regional)	5446	45.37 S, 168.66 E	329	10.45
10. Napier (New Zealand,Regional)	2997	39.50 S, 176.91 E	2	14.12

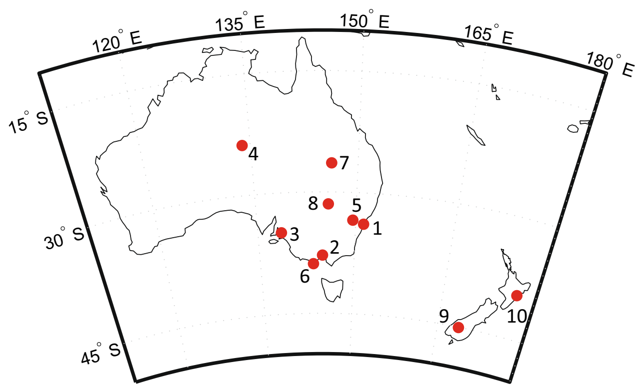


Fig. 1 Location of the considered measuring stations in Australia and New Zealand

average mean absolute error (MAE) in the test set, defined as:

$$\bar{e} = \frac{1}{N}\sum_{j=1}^{N} e_j = \frac{1}{N}\sum_{j=1}^{N} |t_j - \hat{t}_j| \tag{1}$$

where e_j stands for the MAE in sample j of the test set, t_j stands for the measured temperature at j, and $\hat{t}_j$ stands for the predicted temperature at j; N is the length (number of samples) of the test set.

In order to have a baseline to compare the results obtained by the ML approaches, we use a measure of persistence and stationarity. Persistence ($\mathcal{P}$) is defined in this work as the algorithm which assigns t_{j-1} as prediction for t_j, i.e., $t_j = t_{j-1}$, whereas stationarity ($\mathcal{S}$) is defined as the algorithm which assigns the average of the same months in the training set as prediction for t_j, i.e., $t_j = \frac{1}{N_d}\sum_{i \in \mathcal{M}_k} t_i$, where $\mathcal{M}_k$ stands for the month we are evaluating (January, February, etc.), and N_d is the number of samples in the training set corresponding to that month.

3 Materials and methods

3.1 Support vector regression

The SVR are appealing algorithms for a large variety of regression problems, since they do not only take into account the error approximation to the data but also the generalization of the model, i.e., its capability to improve the prediction of the model when a new dataset is evaluated by it. Although there are several versions of the SVR, the classical model, ϵ-SVR, described in detail in Smola and Schölkopf (2004) and used in a large number of application in Science and Engineering (Salcedo-Sanz et al. 2014), is considered in this work. The ϵ-SVR method for regression consists of, given a set of training vectors $\mathbb{T} = \{(\mathbf{x}_i, y_i), i = 1, \ldots, l\}$, training a model of the form $y(\mathbf{x}) = f(\mathbf{x}) + b = \mathbf{w}^T\phi(\mathbf{x}) + b$, to minimize a general risk function of the form

$$R[f] = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} L(y_i, f(\mathbf{x})) \tag{2}$$

where $\mathbf{w}$ controls the smoothness of the model, $\phi(\mathbf{x})$ is a function of projection of the input space to the feature space, b is a parameter of bias, $\mathbf{x}_i$ is a feature vector of the input space with dimension N, y_i is the output value to be estimated and $L(y_i, f(\mathbf{x}))$ is the loss function selected. In this paper, we use the L1-SVRr (L1 support vector regression), characterized by an ϵ-insensitive loss function (Smola and Schölkopf 2004):

$$L(y_i, f(\mathbf{x})) = |y_i - f(\mathbf{x}_i)|_\epsilon \tag{3}$$

In order to train this model, it is necessary to solve the following optimization problem (Smola and Schölkopf 2004):

$$\min\left(\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right) \tag{4}$$

subject to

$$y_i - \mathbf{w}^T\phi(\mathbf{x}_i) - b \le \epsilon + \xi_i, \quad i = 1, \ldots, l \tag{5}$$

$$-y_i + \mathbf{w}^T\phi(\mathbf{x}_i) + b \le \epsilon + \xi_i^*, \quad i = 1, \ldots, l \tag{6}$$

$$\xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, l \tag{7}$$

The dual form of this optimization problem is usually obtained through the minimization of the Lagrange function, constructed from the objective function and the problem constraints. In this case, the dual form of the optimization problem is the following:

$$\max\left(-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x}_i, \mathbf{x}_j) - \epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*)\right) \tag{8}$$

subject to

$$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \tag{9}$$

$$\alpha_i, \alpha_i^* \in [0, C] \tag{10}$$

In addition to these constraints, the Karush–Kuhn–Tucker conditions must be fulfilled, and also the bias variable, b, must be obtained. The interested reader can consult (Smola and Schölkopf 2004) for reference. In the dual formulation of the problem, the function $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel matrix, which is formed by the evaluation of a kernel function, equivalent to the dot product $\langle\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle$. A usual election for this kernel function is a Gaussian function, as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = exp\left(-\gamma \cdot \|\mathbf{x}_i - \mathbf{x}_j\|^2\right). \tag{11}$$

The final form of function $f(\mathbf{x})$ depends on the Lagrange multipliers α_i, α_i^*, as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) \tag{12}$$

In this way, it is possible to obtain a SVR model by means of the training of a quadratic problem for a given hyper-parameters C, ϵ, and γ. However, obtaining these parameters is not a simple procedure, being necessary the implementation of search algorithms to obtain the optimal ones or the estimation of them (Ortiz-García et al. 2009). In this case, a grid search has been implemented in order to obtain the optimal parameters of the SVR. We have used the C implementation of the SVR described in Chang and Lin (2011), available in the Internet at http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

3.2 Multi-layer perceptrons

A MLP consists of a number of computing single elements (neurons) whose connections are adjusted by constructing an input-output mapping in a learning process, so that it is able to predict samples that it has never seen before. In this design stage, a MLP is trained and validated by using, respectively, a training set and usually a validation set to stop the training process without overfitting. The MLP is able to learn in the sense it will be able to predict samples different from those used in the aforementioned design process. This is the so-called ability to generalize and is the capability why MLPs are known to be universal approximators of a large class of functions. MLPs have been successfully applied to a huge amount of nonlinear prediction and classification problems (Haykin 1998; Bishop 1995).

Figure 2 will assist us to explain the main concepts about MLPs. It represents a single hidden layer perceptrons, the most widely used artificial neural network in prediction problems (Haykin 1998). It consists of neurons connected so that an input layer, a hidden layer, and an output layer are formed. The number of neurons in the hidden layer is a parameter to be optimized. The relationship between the input signals (x_j, for $j = 1, 2, \ldots, n$) and the output (y) of a neuron is the following:

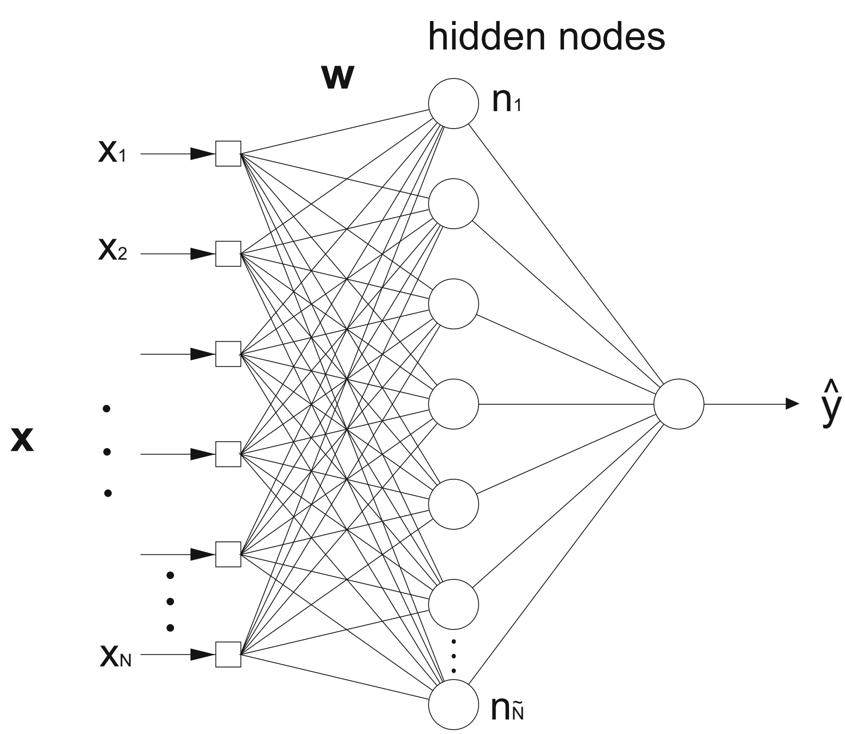
$$y = \varphi\left(\sum_{j=1}^{n} w_i x_j - \theta\right), \tag{13}$$

where φ represents a transfer function, w_j is the weight associated to the j-th input, and θ is a threshold. The transfer function φ is usually considered as the logistic function:

$$\varphi(x) = \frac{1}{1 + e^{-x}}. \tag{14}$$

Usually, the well-known Levenberg–Marquardt algorithm is used to train the MLP (Hagan and Menhaj 1994). This algorithm is based on the Jacobian matrix, which can be computed through a standard back-propagation

Fig. 2 Outline of the MLP structure used in this paper



technique, much less complex than computing the Hessian matrix (Hagan and Menhaj 1994). The Levenberg–Marquardt algorithm works by using the following Newton-like update:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left(\mathbf{J}^T\mathbf{J} + \mu\mathbf{I}\right)^{-1}\mathbf{J}^T\mathbf{e} \tag{15}$$

where $\mathbf{J}$ is the Jacobian matrix, $\mathbf{e}$ is a vector of network errors, and μ is a parameter which controls the process: when $\mu = 0$, it becomes the Newton's method, while when μ is large, it leads to a gradient descent method with small step size. The Matlab implementation of the MLP has been considered in this work.

4 Results and discussion

Table 2 shows the results obtained by the SVR and MLP algorithms in the air temperature prediction problem at the three observational sites and the two training and test periods considered. The results reported in this table consist of MAE measures in the complete test set. For comparison, purposes measures of $\mathcal{P}$ and $\mathcal{S}$ and multi-linear regression (MLR) approach (Neter et al. 1996) are also reported in this table. The results obtained show that the SVR approach is able to obtain the bes t results among all the techniques compared, in both 30/70 and 70/30 cases, and in many cases with values of $\bar{e}$ under 1°C. Note that the persistence measure $\mathcal{P}$ does not seem to be a good model for the monthly prediction of temperature considered, but stationarity $\mathcal{S}$ works quite well in this prediction problem, mainly in the 30/70 case, where the number of samples (information available) is more scarce than in 70/30 case. The MLP approach, however obtains predictions that are in general slightly worse than the MLR algorithm in the 30/70 case, and similar or better in the 70/30 set. The good performance of the MLR algorithm indicates that a multi-linear relationship is a good model to estimate Tmean from the predictive variables considered. The SVR is, however, able to extract some more information of these variables, and thus its prediction improves the multi-linear approach.

A deeper analysis of the results, that shows a possible climate change pattern, can be carried out by considering how the prediction error of the different algorithms considered spreads over the time in each test set. In order to carry out this analysis, we calculate the average error in the next 10 years (decadal performance, 120 months) for each point in the test set, i.e, the following figure of merit (decadal prediction performance of algorithm $\mathcal{A}$):

$$g_k(\mathcal{A}) = \frac{1}{120}\sum_{j=k+1}^{k+120} e_j; \tag{16}$$

Table 2 Results ($\bar{e}$ in the test set in °C) obtained by the different approaches considered in this paper: SVR, MLP, MLR, $\mathcal{P}$ and $\mathcal{S}$

Method	Adelaide	Melbourne	Sydney	Alice Springs	Bathurst	Cape Otway	Charleville	Cobar	Queenstown	Napier
70/30										
SVR	0.9976	1.0183	0.8236	1.3295	0.9349	0.7318	1.2774	1.0081	1.0024*	0.9498*
MLP	1.0871	1.0995	0.8909	1.3143	0.9417	0.7718	1.3052	1.0997	1.1465*	1.0057*
MLR	1.0389	1.0298	0.8362	1.3518	0.9425	0.7731	1.3183	1.0645	1.0069*	0.9573*
$\mathcal{P}$	2.2251	1.9314	1.7731	3.0957	2.5417	1.3619	2.9197	2.9537	2.2761*	1.9555*
$\mathcal{S}$	1.0606	1.1775	0.9043	1.4201	1.0128	0.9009	1.3195	1.1615	1.0508*	1.0077*
30/70										
SVR	0.9989	0.9511	0.7335	1.2777	0.9370	0.7232	1.2039	1.0147	–	–
MLP	1.6452	1.2052	0.8665	1.5070	1.0476	0.8230	1.3595	1.1079	–	–
MLR	1.1153	0.9542	0.7714	1.3482	0.9696	0.7436	1.3080	1.0529	–	–
$\mathcal{P}$	2.2057	1.9322	1.7412	3.1097	2.4431	1.3340	2.9551	2.9513	–	–
$\mathcal{S}$	1.4541	1.0095	0.8010	1.3477	1.0545	0.7853	1.2522	1.0831	–	–

The symbol * stands for values in Queenstown and Napier stations, where data are only available from 1930, though they are depicted within 70/30 case where $\mathcal{A}$ stands for the algorithm to be evaluated (SVR or MLP in this case), e_j stands for the MAE obtained in sample j of the test set and $k = 1, \cdots, N - 121$. Note that $\mathcal{G}(\mathcal{A}) = [g_1, \cdots, g_{N-120}]$ is an $N - 120$-length vector of average MAE values, each one representing the behavior of the algorithm $\mathcal{A}$ in terms of its prediction capability in the next ten years. First of all, we can check out that the SVR improvement over MLP, $\mathcal{P}$ and $\mathcal{S}$ are consistent in terms of prediction capability ($\mathcal{G}$) of the algorithms in all the period considered). Figure 3 shows this fact for the 70/30 case,in Melbourne, Cobar, Alice Springs and Napier stations (for space reasons, we have selected four of the stations to show this point). It can be seen that the SVR algorithm outperform baseline algorithms $\mathcal{P}$ and $\mathcal{S}$ consistently and not only in the final period of the test set. Figure 4 shows this performance for the case 30/70, for Melbourne, Cape Otway, Bathurst, and Alice Springs stations.

We can now depict in a better detail the vector $\mathcal{G}$ for some selected stations and different algorithms, in order to show the possible climate change pattern obtained from the performance of the different algorithms discussed in this paper. We have chosen one urban station (Melbourne), one rural station in Australia (Alice Springs) and one rural station in New Zealand (Queenstown), where the climate change pattern from the algorithms prediction is evident. Figure 5 depicts the prediction performance vector ($\mathcal{G}$) for the SVR, MLP, and MLR approaches, respectively, in the 30/70 and 70/30 cases, in Melbourne. As can be seen in all cases, the prediction performance of the different algorithms compared (Machine Learning ones and MLR) starts getting worse approximately from 1990. This increasing of $\mathcal{G}$ in the last 20 years of the test period indicates a pattern of change in the relationship between the input (predictive) variables considered. In other words, the relationship between the six

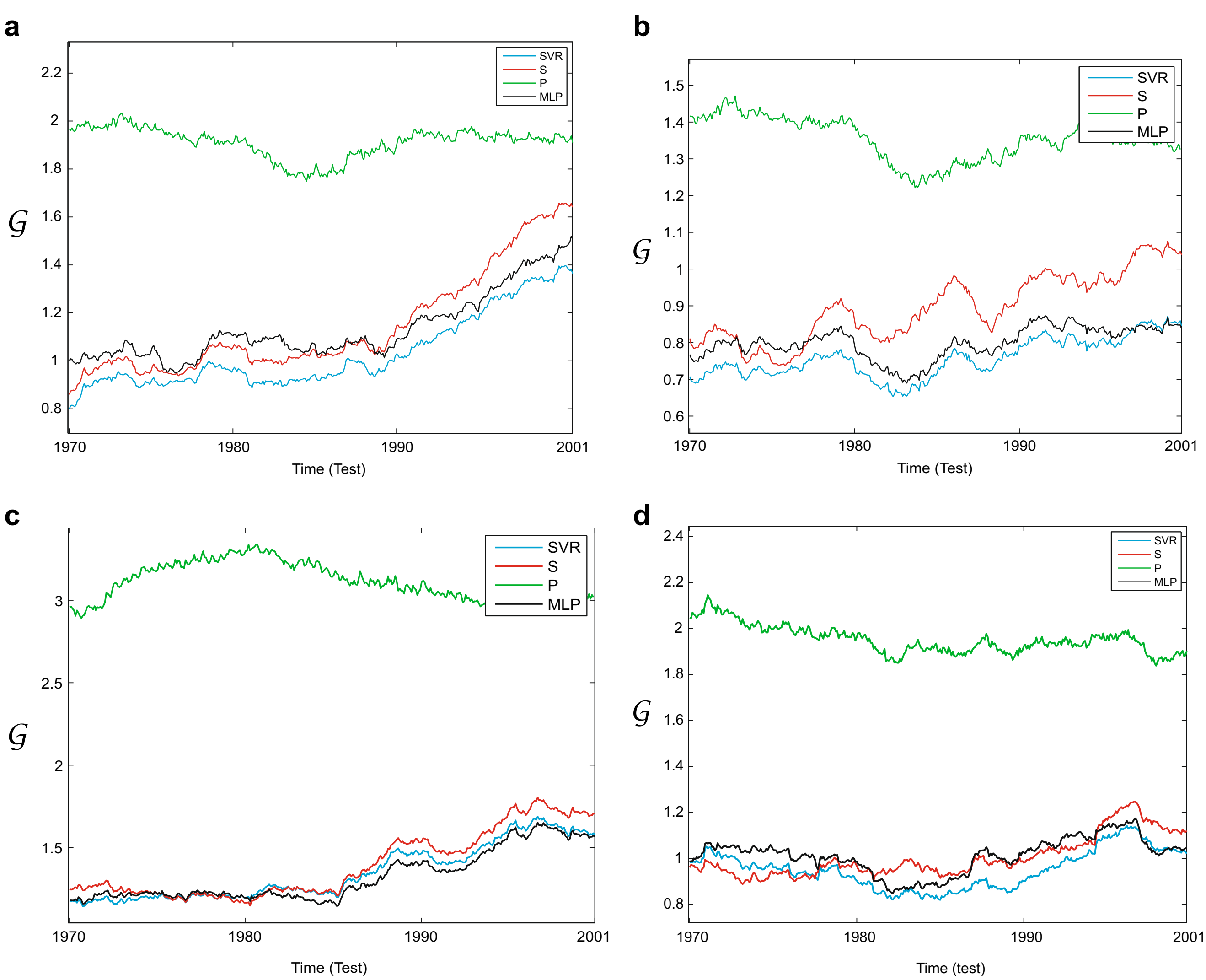


Fig. 3 Comparison of $\mathcal{G}$ for the test period 1971–2010 (70/30) in Melbourne, Cobar, Alice Springs, and Napier; **a** Melbourne; **b** Cobar; **c** Alice Springs; and **d** Napier

input variables and the objective value (monthly mean temperature) changes in the last years of the test period. Note that from 1931 up to about 1990 the prediction performance of the algorithms remains more or less uniform and only starts getting worse at the final years of the test set. This pattern of change from 1990 can be explained by a change in the relationship between the objective variable to be predicted (Tmean) and the predictive variables, which makes the models to be not so accurate anymore in the last years of the test period.

The same analysis can be carried out for a rural station, specifically Alice Springs. Figure 6 shows the results obtained in this station, for all the algorithms considered and the 30/70 and 70/30 cases, respectively. The behavior of the algorithms' prediction performance in this case is similar than the previous case, and also show the pattern of change in the algorithms prediction capability around 1990. For Queenstown station, in New Zealand, the prediction performance of the algorithms is shown in Fig. 7. Note that in this case, there is only one training set that starts in 1930 up to 1970, with the test set in the period 1971–2010. Also, in this case is evident the different performance of the SVR, MLP and MLR algorithms over the time, with a clear lower performance of the techniques in the last years of the test period, what indicates that the climate change pattern can be also detected in New Zealand.

A final discussion on these results can be done, by considering the variables involved in the temperature prediction problem tackled. First of all, it can be checked out that the climate indices just slightly modify the regressors predictions. These variables provide a final tuning of the prediction, and only slightly worse results can be obtained by using the temperature of the previous month and the two dummy variables. However, we have run several experiments by removing some of these variables core variables (previous month and dummies), and in this case,

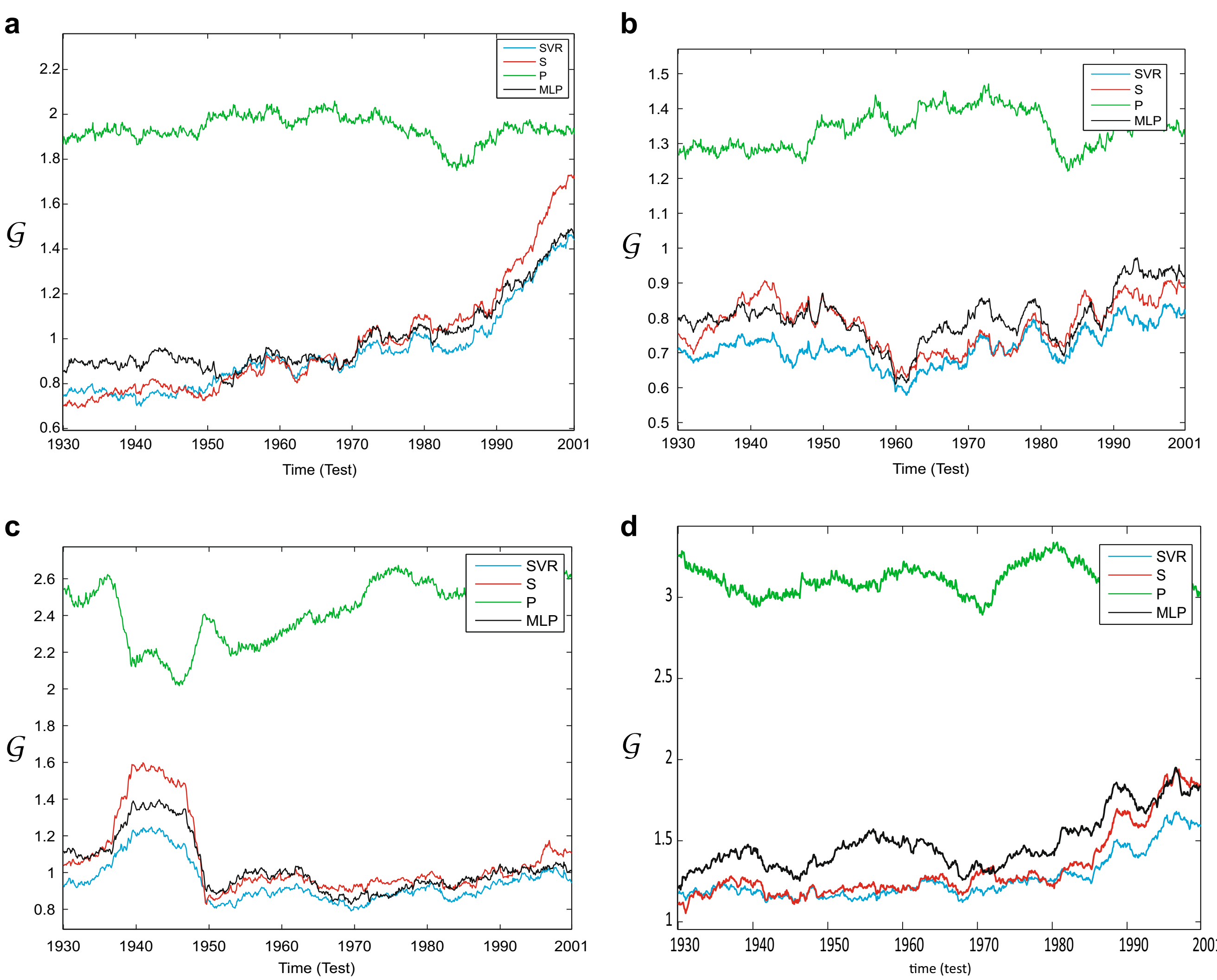


Fig. 4 Comparison of $\mathcal{G}$ for the test period 1931–2010 (30/70) in Melbourne, CapeOtway, Bathurst, and Alice Springs; **a** Melbourne; **b** Cape Otaway; **c** Bathurst; and **d** Alice Springs

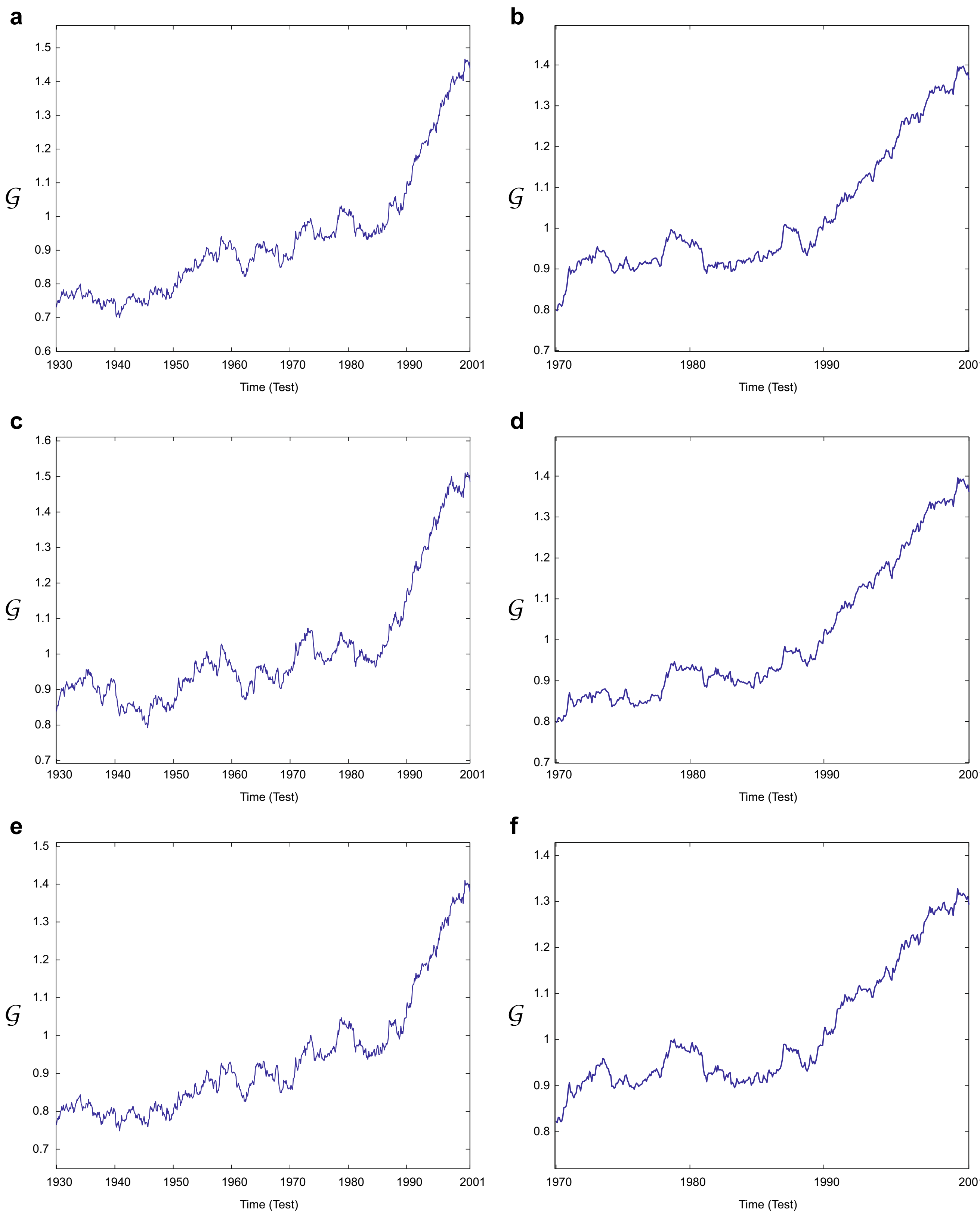


Fig. 5 Vector $\mathcal{G}$ obtained by SVR, MLP, and MLR algorithms in the different training and test cases, for Melbourne station; **a** SVR 30/70; **b** SVR 70/30; **c** MLP 30/70; **d** MLP 70/30; **e** MLR 30/70; and **f** MLR 70/30

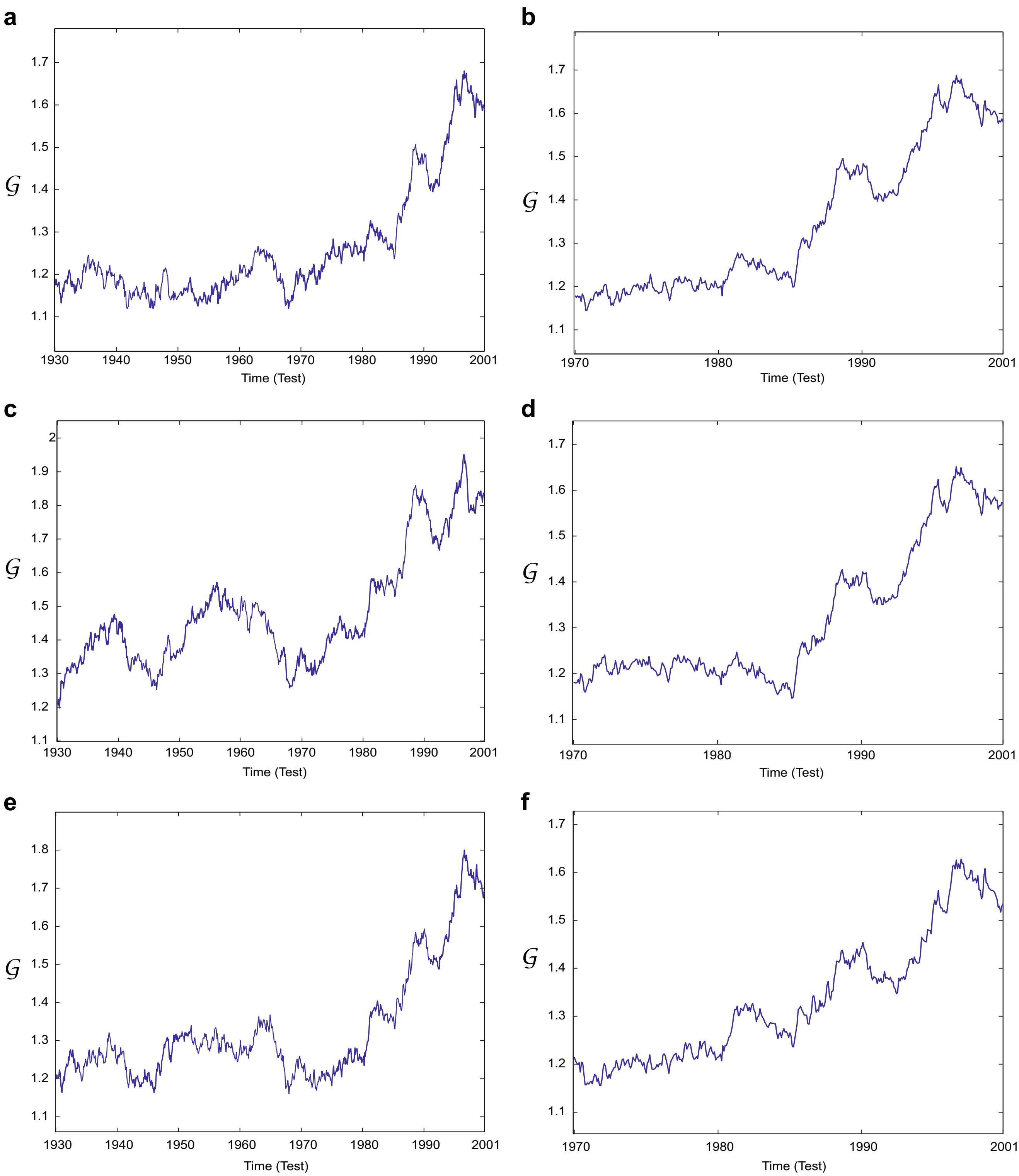


Fig. 6 Vector $\mathcal{G}$ obtained by SVR, MLP, and MLR algorithms in the different training and test cases, for Alice Springs station; **a** SVR 30/70; **b** SVR 70/30; **c** MLP 30/70; **d** MLP 70/30; **e** MLR 30/70; and **f** MLR 70/30

the performance of the ML approaches gets indeed much worse. Thus, we can consider these three variables as key parameters to obtain a good quality monthly air temperature prediction. Considering the previous month and the two dummies as the core variables for the temperature prediction, the analysis of the pattern of change in the algorithms

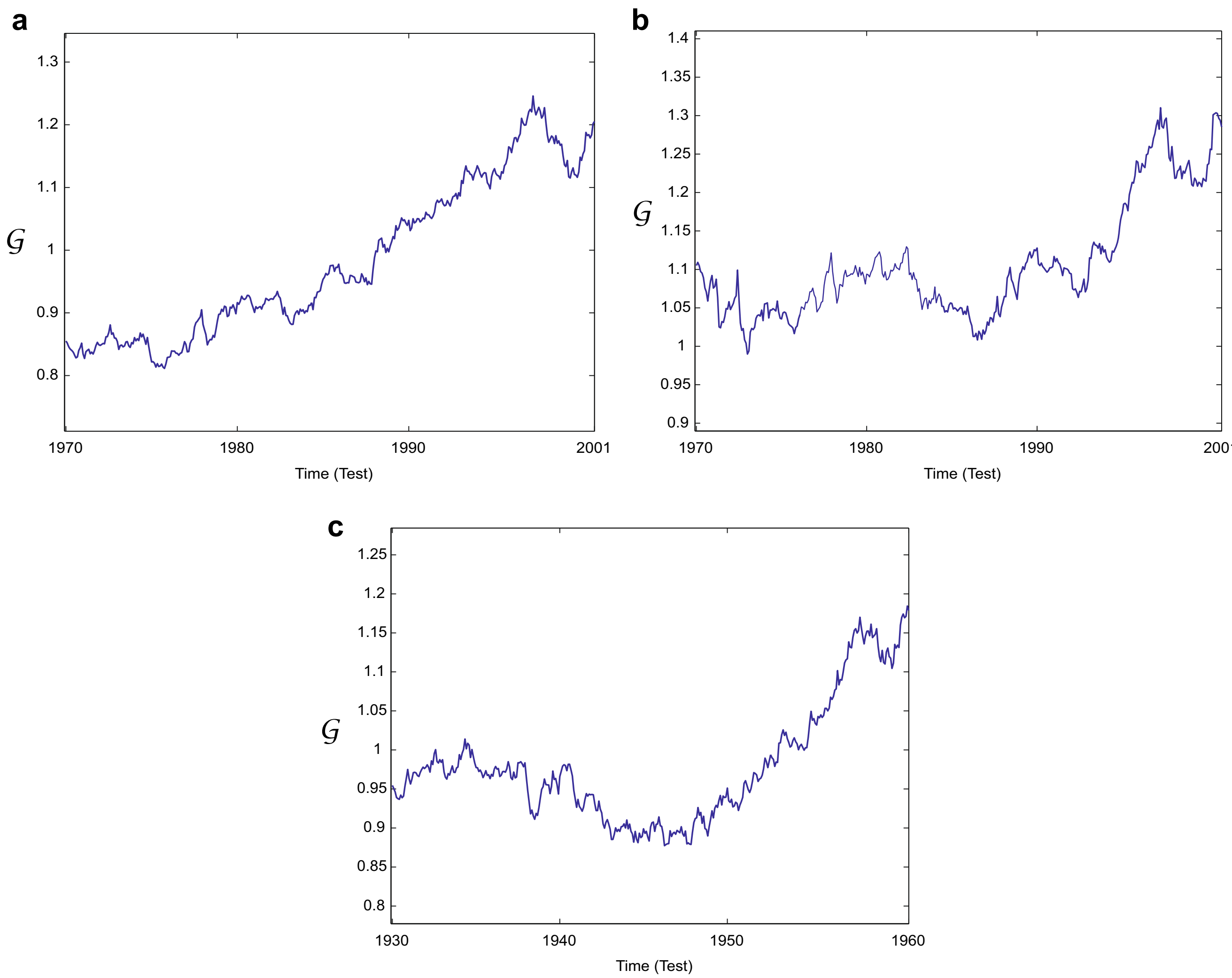


Fig. 7 Vector $\mathcal{G}$ obtained by SVR, MLP, and MLR algorithms for Queenstown station; **a** SVR 70/30; **b** MLP 70/30; and **c** MLR 70/30

prediction performance around 1990 indicates that the relationship between the temperature of the previous month and the objective variable (the temperature of the month be predicted) has been affected in the last 20/25 years. Also, the two dummy variables act as deseasonalizing variables, since each month of the year is characterized by a pair of values from these variables. The results obtained seem indicate that these variables do not help to the prediction performance of the different algorithms compared in the last years of the test set (or at least they do not help as in the previous years). Summarizing, the obtained results indicate that the relationship between the mean temperature of different months in Australia and New Zealand is getting somehow affected due to external forcings, and it has changed in the last 20 years, what affects to the prediction performance of the studied algorithms.

5 Conclusions

In this paper, we have carried out a comparative study of the performance of two ML algorithms in a problem of monthly air temperature prediction in Australia and New Zealand stations. We have shown that the SVR algorithm is the best approach among the ones that have been considered in this paper, including neural networks, multi-linear regression, and baseline persistence and stationarity approaches. We have then shown that the performance of the different algorithms (in terms of their prediction error over the tested period) can be used to spot a climate change pattern in the stations analyzed, since their performance is much worse in the last 20 years than that in the previous decades. This behavior indicates a possible change within the time in the relationship between the different variables involved in this problem of monthly temperature prediction.

Acknowledgments The high quality mean temperature datasets and climate mode indices were obtained from the Australian Bureau of Meteorology.

References

Abbot J, Marohasy J (2012) Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Adv Atmos Sci 29:717–730

Abbot J, Marohasy J (2014) Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. Atmos Res 138:166–178

Alexander L, Hope D, Collins B et al. (2007) Trends in Australia's climate means and extremes: A global context. Aust Meteorol Mag 56:1–18

Bertini I, Ceravolo F, Citterio M et al. (2010) Ambient temperature modelling with soft computing techniques. Sol Energy 84:1264–1272

Bishop C-M (1995) Neural networks for pattern recognition. Oxford University Press

Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(27):1–27

Chevalier R-F (2008) Air Temperature Prediction Using Support Vector Regression and GENIE: the Georgia Extreme-weather Neural-network Informed Expert. PhD dissertation, University of Georgia, 2008

Chevalier R-F, Hoogenboom G, McClendon R-W, Paz J-A (2011) Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks. Neural Comput & Applic 20(1):151–159

Chithra N-R, Thampi S-G, Surapaneni S, Nannapaneni R, Kumar Reddy AA, Kumar J-D (2014) Prediction of the likely impact of climate change on monthly mean maximum and minimum temperature in the Chaliyar river basin, India, using ANN based models. Theoretical and Applied Climatology, in press

Cramer W-G, Yohe et al. (2013) Detection and Attribution of observed impacts. IPCC5 Work Group 2, 5th Assessment Report, Chapter 18:1–94

Coates L (1996) An Overview of fatalities from some natural hazards in Australia. In: Heathcoote, Cuttler, Koetz (eds) Natural Disaster Reduction (NDR96): conference proceedings, Institute of Engineers Australia

Collins D (2006) High quality Australian annual temperature dataset, Statement on Bureau of Meteorology High Quality Dataset (ftp://ftp.bom.gov.au/anon/home/ncc/www/change/HQannualT), National Climate Centre, February

Daneshmand H, Tavousi T, Khosravi M, Tavakoli S (2014) Modeling minimum temperature using adaptive neuro-fuzzy inference system based on spectral analysis of climate indices: a case study in Iran. J Saudi Soc Agric Sci. doi:10.1016/j.jssas.2013.06.001. in press, 2014

Della-Marta P, Collins D, Braganza K (2004) Updating Australia's high-quality annual temperature dataset. Aust Meteorol Mag 53:75–93

Dombayc O-A, Gölcü MM (2009) Daily means ambient temperature prediction using artificial neural networks method: a case study of Turkey. Renew Energy 34:1158–1161

Douglass D-H, Blackman E-G, Knox R-S. (2004) Temperature response of Earth to the annual solar irradiance cycle. Phys Lett A 325:315–322

Garske T, Ferguson N-M, Ghani A (2013) Estimating air temperature and its influence on malaria transmission across Africa. PLoS ONE 8(2)

Hagan M-T, Menhaj M-B (1994) Training feed forward network with the Marquardt algorithm. IEEE Trans Neural Netw 5(6):989–993

Haykin S (1998) Neural networks: a comprenhensive foundation. Prentice Hall

Jacobs S-J, Pezza A-B, Barras V, Bye J, Vihma T (2013) An analysis of the meteorological variables leading to apparent temperature in Australia: Present climate, trends, and global warming simulations. Glob Planet Chang 107:145–156

Kadu P, Wagh K, Chatur P (2012) Analysis and Prediction of Temperature using Statistical Artificial Neural Network. IJCSMS Int J Comput Sci Manag Stud 12(2):2231–5268

Kaufmann R-K, Stern D-I (1997) Evidence for human influence on climate from hemispheric temperature relations. Nature 388:39–44

Kaufmann R-K, Stern D-I (2002) Cointegration analysis of hemispheric temperature relations. J Geophys Res: Atmos 107:4012

Kaufmann R-K., Kauppi H, Mann M-L, Stock J-H (2011) Reconciling anthropogenic climate change with observed temperature 1998-2008. PNAS 108:11790–11793

Luk K, Ball J, Sharma A (2000) A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. J Hydrol 227:56–65

Mekanik F, Imteaz M, Gato-Trinidad S, Elmahdi A (2013) Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. J Hydrol 503:11–21

Mellit A, Massi Pavan A, Benghanem MM (2013) Least squares support vector machine for short-term prediction of meteorological time series. Theor Appl Climatol 111:297–307

Nairn J, Fawcett R (2013) Defining heatwaves: heatwave defined as a heat impact event servicing all community and business sectors in Australia. The Centre for Australian Weather and Climate Research, CAWCR Technical Report No. 060

Neter J, Kutner M-H, Nachtsheim C-J, Wasserman W (1996) Applied Linear Statistical Models, IRWIN. The McGraw-Hill Companies, Inc.

Nicholls N, Lavery B, Frederiksen C, Drosdowsky W, Torok S (1996) Recent changes in relationships between the El Nio Southern Oscillation and Australian rainfall and temperature. Geophys Res Lett 23:3357–60

Ortiz-García E-G, Salcedo-Sanz S, Pérez-Bellido A-M, Portilla-Figueras J-A (2009) Improving the training time of support vector regression algorithms through novel hyper-parameters search space reductions. Neurocomputing 72:3683–3691

Ortiz-García E-G, Salcedo-Sanz S, Casanova-Mateo C, Paniagua-Tineo A, Portilla-Figueras A (2011) Accurate local very short-term temperature prediction based on synoptic situation support vector regression banks. Atmos Res 107:1–8

Paniagua-Tineo A, Salcedo-Sanz S, Casanova-Mateo CC, Ortiz-García E-G, Cony M-A, Hernández-Martín E (2011) Prediction of Daily Maximum Temperature using a Support Vector Regression Algorithm. Renew Energy 3(11):3054–3060

Radhika Y, Shashi M (2009) Atmospheric temperature prediction using support vector machines. Int J Comput Theory Eng 1(1):1793–8201

Saji N-H, Ambrizzi T, Ferraz S-E (2005) Indian Ocean Dipole mode events and austral surface air temperature anomalies. Dyn Atmos Oceans 39:87–101

Salcedo-Sanz S, Rojo J-L, Martínez-Ramón M, Camps-Valls G (2014) Support vector machines in engineering: an overview. WIREs Data Mining and Knowledge Discovery, in press, 2014

Smith B-A, Hoogenboom G, McClendon R-W (2007) Improving air temperature prediction with artificial neural networks. Int J Comput Intell 3(3):179–186

Smith B-A, Hoogenboom G, McClendon R-W (2009) Artificial neural networks for automated year-round temperature prediction. Comput Electron Agric 68:52–61

Smola A-J, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222

Stone D-A, Allen M-R (2005) Attribution of global surface warming without dynamical models. Geophys Res Lett 32: L18711

Tasaduqq I, Rehman S, Bubshait K (2002) Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. Renew Energy 25:545–554

Torok S-J, Nicholls N (1996) A historical annual temperature dataset for Australia. Aust Meteorol Mag 45:251–260

Ustaoglu B, Cigizoglu H-K, Karaca M (2008) Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorol Appl 15:431–445

Williams S, Nitschke M, Sullivan T, Tucker G-R, Weinstein P, Pisaniello D-L et al. (2012) Heat and health in Adelaide, South Australia: Assessment of heat thresholds and temperature relationships. Sci Total Environ 414:126–133

WGCM Coupled Model Inter-comparison Project Phase 5, last access April 27th 2014: http://cmip-pcmdi.llnl.gov/cmip5/

http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Xu Z, Liu Y, Ma Z, Li S, Hua W, Tong S (2014) Impact of temperature on childhood pneumonia estimated from satellite remote sensing. Environ Res 132:34–341

You Q, Fraedrich, Min J, Kang S, Zhu X, Ren G, Meng X (2013) Can temperature extremes in China be calculated from reanalysis? Glob Planet Chang 111:268–279