geographical analysis

Geographical Analysis ISSN 0016-7363

A Hybrid Framework for Space–Time Modeling of Environmental Data

Tao Cheng,[1] Jiaqiu Wang,[1] Xia Li[2]

[1]Department of Civil, Environmental and Geomatic Engineering, University College London, London, U.K., [2]School of Geography and Planning, Sun Yat-sen University, Guangzhou, People's Republic of China

The space–time autoregressive integrated moving average (STARIMA) model family provides useful tools for modeling space–time processes that exhibit stationarity (or near stationarity) in space and time. However, a more general method for routine use and efficient computation is needed to model the nonlinearities and nonstationarities of environmental space–time series. This article presents a hybrid framework combining machine learning and statistical methods to address this issue. It uses an artificial neural network (ANN) to extract global deterministic (nonlinear) space–time trends and a STARIMA model to extract local stochastic space–time variations in data. A four-stage procedure is proposed for analyzing and modeling space–time series. The proposed framework and procedures are applied to forecast annual average temperature at 137 national meteorological stations in China. The results demonstrate that the hybrid framework achieves better forecasting accuracy than the STARIMA model alone. This finding suggests that the combination of machine learning and statistical methods provides a very powerful tool for analyzing and modeling space–time series of environmental data that have strong spatial nonlinear and nonstationary components.

Introduction

Space–time autoregressive integrated moving average (STARIMA) models have gained widespread popularity in modeling multiple time-series data that correspond to different spatial locations, which are known as space–time series. The STARIMA model family furnishes models of different forms for space–time series analysis: space–time autoregressive (STAR), space–time moving average (STMA), space–time autoregressive moving average (STARMA), and STARIMA (Bennett 1975; Cliff and Ord 1975; Martin and Oeppen 1975; Pfeifer and Deutsch 1980). This specification has been successfully applied to model space–time processes in

Correspondence: Tao Cheng, Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, WC1E 6BT London, U.K.
e-mail: tao.cheng@ucl.ac.uk

Submitted: July 1, 2008. Revised version accepted: March 3, 2010.

many domains, such as social economics (Pace et al. 1998), transport (Kamarianakis and Prastacos 2005), and image analysis (Crespo et al. 2007).

The STARIMA models, though, are based on the assumption that space–time processes can be rendered stationary (or near stationary) in both space and time (Bennett 1975; Martin and Oeppen 1975; Pfeifer and Deutsch 1980). However, the application of the STARIMA models is not straightforward for environmental data (Kamarianakis and Prastacos 2005). They are weak nonstationary series in time because spatial structure tends to change slowly over time, as an environment evolves. Moreover, such spatial structure exhibits strong heterogeneity and nonlinearity (Brundson, Fotheringham, and Charlton 1996; Cooper et al. 1997; Haas 1998). Thus, a more general method for routine use and efficient computation is needed to tackle the nonlinearity and nonstationarity of space–time series, especially for environmental applications. The aim of our study is to extract nonlinear space–time trends (or patterns) from data before describing them with a STARIMA model. We examine this possibility by reviewing the representation and modeling of space–time series using a STARIMA model.

Space–time modeling of environmental data

In spatial (or temporal) data analysis, a spatial (or temporal) process z can be decomposed into two parts: a global deterministic spatial/temporal variation μ, and a local stochastic spatial/temporal variation ξ. The global variation refers to trends or patterns across a study area (or within a study period). Local variation refers to more localized spatial/temporal structures in data and is conceptualized as being superimposed on any geographically extensive pattern that may be present (Haining 2003; Kanevski and Maignan 2004). Therefore, we think that a space–time process (series) can be described by

$$z_i(t) = \mu_i(t) + \xi_i(t), \quad \mu_i(t) = f(i,t), \tag{1}$$

where $z_i(t)$ represents the observation of a data series at spatial location i at time t, $\mu_i(t)$ represents the space–time patterns that explain global deterministic space–time trends, and f can be expressed as a nonlinear function having space argument i and time argument t. The residual term $\xi_i(t)$ is a zero-mean space–time correlated error that includes small-scale stochastic space–time variations. Without i, equation (1) becomes a time-series analysis; without t, equation (1) becomes a spatial-series analysis.

For the conventional STARIMA model, Martin and Oeppen (1975) suggest differencing to remove a deterministic trend to satisfy the stationarity (or weak stationarity) property for a series (refer to conditions [a], [b], and [c] in equation [3]). However, differencing is unable to remove the spatial nonlinear and nonstationary trends in a space–time series of environmental data (De Luna and Genton 2005). An artificial neural network (ANN) can simulate nonlinear systems by nonlinear regression (Mitchell 2003). ANN has been successfully applied to a variety of simulation and estimation tasks and more recently to spatial data analysis (Kanevski

et al. 1996; Li, Dunham, and Xiao 2003; Cheng and Wang 2008). We think that, depending on its architecture, an ANN may also be able to capture nonlinear space–time trends at different scales, shedding light on modeling nonlinear space–time trends in a space–time series. Therefore, we use an ANN to develop a nonparametric and robust model to extract the global deterministic space–time trends (or patterns) in a series; we then use a STARIMA specification to analyze the residuals of small-scale stochastic space–time variations. Consequently, we develop a hybrid framework.

ANN to model deterministic global space–time trends

An ANN is a mathematical model that creates a nonlinear map that links a set of input variables to a set of output variables. An ANN is often called machine learning (Mitchell 2003), because it can learn from experience using numerical and, sometimes, linguistic sample data. An ANN depends on both input data and the inner structure of its model (e.g., number of neurons, hidden layers, types of connections, and information flow direction). The key elements of an ANN are its nodes, which can simulate neural synapses. In these nodes, learning algorithms combine a linear-weighted sum of input variables through a nonlinear transfer function (called activation function) to yield output variable(s). The conjunctive weights are chosen by an iterative algorithm (also called a backpropagation algorithm) that aims to minimize the difference between the calculated and the desired output. Training a network, that is, choosing the network's parameter values, or weights, resembles nonlinear regression analysis (Haykin 1994). The global space–time trend term $\mu_i(t)$ (equation [1]) can be modeled as follows:

$$\hat{\mu}_i(t) = f\left(\sum_{k=1}^{n} \beta_k f(\mathrm{i},\mathrm{t}) + \beta_0\right), \tag{2}$$

where $\hat{\mu}_i(t)$ represents the forecasting value at spatial location i at time t, which is the output of a neural network with n input nodes (i, t); function f is a nonlinear activation function; β_k is conjunctive weight; and β_0 is threshold value. This equation is trained (fitted) first by training data and then used for forecasting.

Multilayer perception (MLP) networks have the capability to undertake complex mapping between inputs and outputs that enables a network to approximate nonlinear functions. Usually, the architecture of an ANN is designed with three-layer networks (Fig. 1), consisting of an input layer, a hidden layer, and an output layer, which have been proved to be able to approximate an arbitrary nonlinear system (Hornik, Stinchombe, and White 1990). The number of output layer nodes is decided by the dimensions of an output variable. The input layer nodes are determined by the dimensions of an input variable related to the output variable. Cross-validation is usually carried out to determine the optimal number of hidden nodes because it has been proved to be the most effective approach (Mitchell 2003).

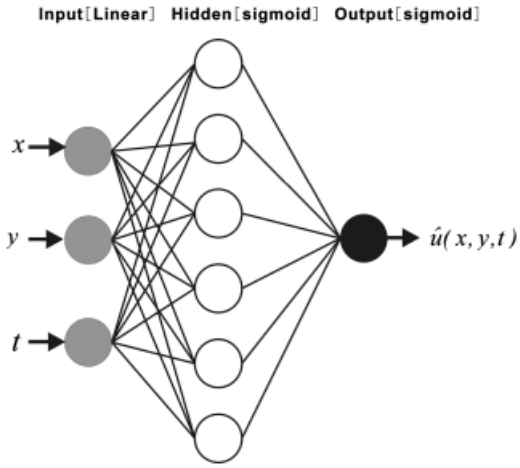

Figure 1. Structure of the implemented ANN model. In the MLP network, inputs consist of spatial coordinates x (longitude), y (latitude), and time coordinate t. Six hidden nodes are used for the hidden layer, and the output is the annual average temperature at each spatial location at time t. Training data were organized as a sample, the length of which is 137×42, and there are 137 outputs at each time t, which represent annual average temperature forecasts at the 137 stations.

STARIMA to describe local stochastic space–time processes

The principle of the STARIMA model

The STARIMA model is a space–time model that expresses each observation at time t and location i as a weighted linear combination of the previous observation at location i (a time lag) and neighboring observations lagged in space. In this study, a STARIMA model describes the local stochastic variations of the error term $\xi_i(t)$ in equation (1) as follows:

$$\xi_i(t) = \sum_{k=1}^{p} \sum_{h=0}^{m_k} \varphi_{kh} W^{(h)} \xi_i(t-k) - \sum_{l=1}^{q} \sum_{h=0}^{n_l} \theta_{lh} W^{(h)} \varepsilon_i(t-l) + \varepsilon_i(t), \qquad (3)$$

where p is the autoregressive order, q is the moving average order, m_k is the spatial order of the kth autoregressive term, n_l is the spatial order of the lth moving average term, φ_{kh} is the autoregressive parameter at temporal lag k and spatial lag h, θ_{lh} is the moving average parameter at temporal lag l and spatial lag h, $W^{(h)}$ is the $N \times N$ matrix of weights for spatial order $h(W^{(0)} = I)$, and $\varepsilon_i(t)$ is a normally distributed random error at time t and location i with the following properties:

(a) $E[\varepsilon_i(t)] = 0$,

(b) $E\left[\varepsilon_i(t)\varepsilon_j(t+s)'\right] = \begin{cases} \sigma^2 I, & i = j,\ s = 0, \\ 0, & i \neq j,\ s \neq 0, \end{cases}$

(c) $E[\xi_i(t)\varepsilon_i(t+s)'] = 0$, for $(s > 0)$.

Property (a) means that the expectation of $\varepsilon_i(t)$ is 0. Property (b) means that the $\varepsilon_i(t)$ are pairwise independent for all pairs of locations in a data set and that all locations have the same underlying variance σ^2. Property (c) means the $\varepsilon_i(t)$ are independent of the raw data set. This STARIMA model can be written as STARIMA(p, q), or STARIMA(p, $m_{k(k=1,p)}$, q, $n_{1(1=1,q)}$) as p, m_k, q, and n_1, defined in equation (3).

The Box–Jenkins three-step estimation procedure

The procedure adopted to operationalize a STARIMA model is commonly known as the Box–Jenkins method (Pfeifer and Deutsch 1980; Box, Jenkins, and Reinsel 1994), which includes three essential steps:

Model identification: The autoregressive order p and the moving average order q are chosen provisionally after an examination of space–time partial autocorrelation functions (ST-PACF) (that cut off after p lags in time, and each with m_k [$k = 1, p$] lags in space) and space–time autocorrelation functions (ST-ACF) (that cut off after q lags in time, and each with n_l [$l = 1, q$] lags in space).

Model estimation: The model parameters φ and θ are estimated. STARIMA(p, q) models with $q \neq 0$ are nonlinear in form. Accordingly, parameter estimation is performed using a variety of nonlinear optimization techniques such as nonlinear least-square estimation. Although, theoretically, maximum likelihood estimation (MLE) is best, conditional MLE is used due to lack of prior knowledge to initialize the starting values of MLE (Pfeifer and Deutsch 1980).

Diagnostic checking: The residuals of a fitted model need to be evaluated to determine whether the candidate STARIMA model adequately explains any remaining small-scale stochastic space–time variations. These residuals should be white noise, implying that the mean of the ST-ACF for these residuals essentially should be zero and that the variance should be close to $[N(T-s)]^{-1}$. In addition, the statistical significance of the estimated parameters should be checked based on asymptotic confidence intervals. Any estimated parameters that prove to be statistically nonsignificant should be removed from a candidate STARIMA model specification, and the resulting simpler model should be considered the preferable specification. This iterative procedure should be repeated until all parameters are statistically significant and the residuals display their assumed properties (Pfeifer and Deutsch 1980).

The spatial weight matrix

In conventional STARIMA analysis, a spatial weight matrix is defined based upon regular spatial hierarchical orders, and equal weights are assumed for the hth order neighbors (Martin and Oeppen 1975; Pfeifer and Deutsch 1980). Building hierarchical spatial orders for environmental data is impracticable because they have nonlinear and nonstationary spatial trends and stronger spatial correlation, and they are anisotropic and irregularly sampled. Exploratory identification of the spatial hierarchical orders and equal weights is hampered by (1) the computational costs of ST-ACF and ST-PACF becoming enormous, and (2) spatial lag order m_k or n_1 at

each temporal lag of a STARIMA model being difficult to determine. Thus, only the first-order spatial lag effect is considered for the points that lie within a prespecified distance threshold (see Table 1, the maximum range).

While an autoregressive model (such as the STARIMA) specification involves the inverse covariance matrix, a semivariogram model specification involves the covariance matrix itself (Griffith and Csillag 1993). Thus, we use a semivariogram model to specify the STARIMA model because the two models are almost identical in terms of the autocorrelation structure and variance.

In this study, experimental results show that the Gaussian semivariogram model achieves the best fit, and hence it was selected for analysis purposes. The Gaussian model is defined as (Isaaks and Srivastana 1989)

$$\begin{cases} \gamma(h) = C_0 + C[1 - e^{-3(h/D)^2}], & 0 < h \le D, \\ \gamma(h) = C_0 + C, & h > D, \\ \gamma(0) = 0, & h = 0, \end{cases} \tag{4}$$

where h is the spatial lag distance, C is the partial sill (or the spatial heterogeneity arising from spatial autocorrelation), C_0 is the nugget (or the spatial variability arising from random components such as measured error and stochastic noise, or model misspecification error), and $C+C_0$ is the sill or sample variance. C_0/sill indicates the percentage of the spatial heterogeneity caused by the stochastic factor to the total spatial heterogeneity. D is the effective range. That is, the measurement points estimated within the range are spatially autocorrelated, whereas points outside the range are considered independent.

Spatial weights are defined between two points as

$$\begin{cases} W(h) = [(C_0 + C_1) - \gamma(h)]/(C_0 + C_1), & h \le D, \\ W(h) = 0, & h = 0 \text{ or } h > D. \end{cases} \tag{5}$$

Table 1 Fitted ANN Residuals' Isotropic Semivariogram Analysis

	Range (km)	Partial sill (C)	Nugget (C_0)	Sill (C_0+C)	C_0/sill (%)
1951	1715.3	10.129	7.872	18.001	43.731
1955	1201.7	9.717	4.19	13.907	30.129
1960	1275.7	7.287	2.911	10.198	28.545
1965	1279.0	8.756	3.02	11.776	25.645
1970	1226.5	7.16	2.948	10.108	29.165
1975	1265.9	7.22	2.988	10.208	29.271
1980	1464.5	8.601	1.631	10.232	15.940
1985	1469.7	9.352	1.766	11.118	15.884
1990	1296.1	7.345	3.059	10.404	29.402
1992	1446.6	8.702	1.706	10.408	16.391

A hybrid framework for the integrated modeling of space–time series

Fig. 2 shows the hybrid framework for integrated space–time modeling and forecasting, which comprises four stages: space–time data preparation, exploratory space–time analysis, model training, and model validation.

In the *space–time data preparation stage*, outliers in a data set should be detected and dealt with in some meaningful way. Anomaly detection can be done by histogram analysis and descriptive statistics analysis of the time series at each spatial location. Although transformation of data could be used to accommodate the anomaly, removing outliers from a data set is often one of the options when outliers lie far outside the range of the remaining data. Then a data set should be split into two groups: one as a sample set (usually 80% of the data) to train a model, and one as a validation set (usually 20% of the data) to test the model.

An *exploratory space–time analysis* may be used to diagnose whether data satisfy modeling conditions of STARIMA parameter estimation by examining plots of spatial surface trend and time-series ACF to check whether the data are stationary. Otherwise, an ANN model can be fitted to the data to capture nonlinear deterministic space–time trends.

In the *model training* stage, an ANN may be applied to extract global space–time trends, and then the ANN residuals (ANN-fitted values subtracted from observation values) are examined using semivariogram analysis. If they are uncorrelated, the ANN successfully describes all space–time structures represented in the raw data. Otherwise, the STARIMA model should be fitted to the residuals to account for the correlations. Then the spatial weight matrix can be defined, and the Box–Jenkins three-step procedure may be applied to train/fit the STARIMA model.

In the *model validation* stage, space–time forecasts are constructed from the sum of the ANN and STARIMA estimates (see equation [1]). The ST-ACF of the residuals of combined estimates may be calculated for further diagnostic checking, such as whether the residuals are random and normally distributed. Furthermore, to help understand the quality of results, a variogram-based analysis may be performed at each step to evaluate the spatial and temporal variability as a model is built. If no spatial correlation exists between residuals, then the (C_0/sill) ratios are close to 100%. In other words, nugget value C_0 is close to the sill variance C_0+C. The overall performance of a space–time modeling exercise is evaluated by its prediction accuracy.

A case study: Annual average temperature across China

This section presents a case study to illustrate the four-stage procedure affiliated with the hybrid model. The purpose of this case study is to test the difference between a hybrid model (ANN+STARIMA) and its ordinary STARIMA model by forecasting annual average temperature (degree/year) across China.

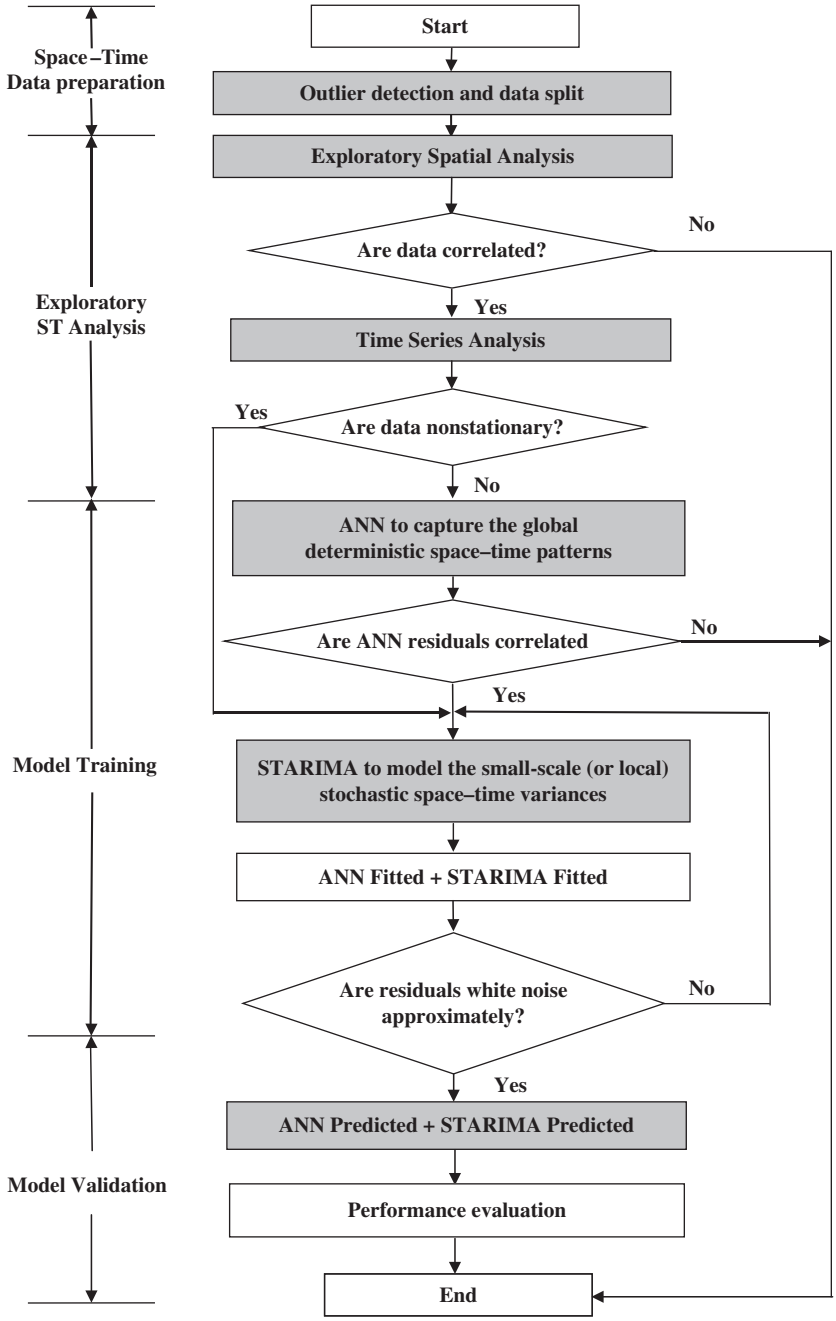

Figure 2. A hybrid framework for integrated space–time modeling of environmental data.

Data preparation

The original data are $T = 52$ year observations (1951–2002) of annual average temperature at $N = 194$ national meteorological stations provided by the National

Meteorological Centre of the People's Republic of China. Among these 194 stations, 137 observations were retained because the measurements at the remaining 57 stations may cause misspecification of the model (43 stations have only two years of observations, and another 14 stations have extreme outliers). The areas without qualified observation stations are highlighted by symbol ''X'' in Fig. 3 and in other figures. The data between 1951 and 1992 (nearly 80% of 52 years) were chosen as the training data set for the forecasting between 1993 and 2002 (nearly 20% of the 52 years). This completes the two steps of the first stage of the analysis procedure.

Exploratory space–time analysis

Time-series analysis and spatial data analysis methods were used separately to examine whether the data were correlated and stationary in time and space. Fig. 3b shows the annual temperature in Beijing with a clearly increasing trend. After first-order differencing, it becomes a stationary series in time (Fig. 3c), indicating it is a weak stationary series in time. A similar pattern can be found in the data for other stations. To discover the spatial pattern, a spatial trend analysis was conducted, and pattern maps of annual average temperature were generated using kriging (Fig. 4). Two spatial trends are detected in these maps: a decreasing spatial trend from the southeast to the middle north and an increasing spatial trend from the middle north to the northwest, indicating that the trend surface is spatially nonlinear and non-stationary. Fig. 5 presents the standard error map for the trend portrayed in Fig. 4: estimates in the southwest of the study area have larger deviations than those in southeast and northeast, due to the relative scarcity of national stations in the areas

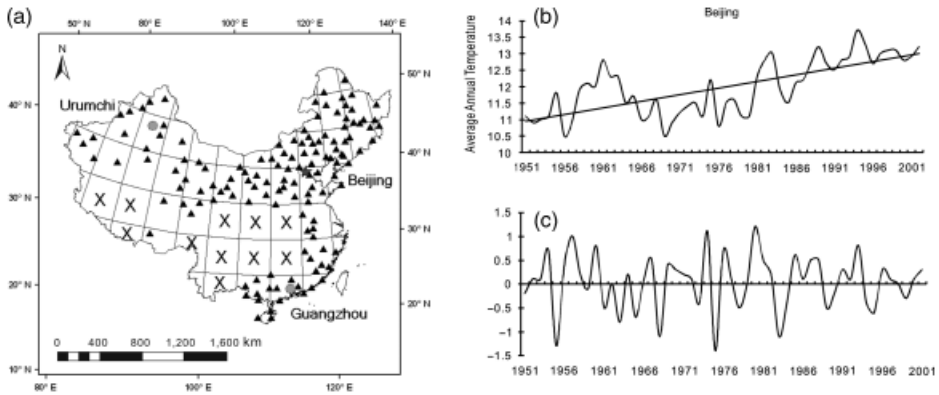

Figure 3. Meteorological stations used for the space–time series analysis in the case study: (a) spatial locations of the remaining 137 stations. The areas that are not covered by the 137 stations (such as Tibet and southwest China) are highlighted by symbol ''X'' here and in Figs. 4–6, 9, and 10; (b) time series of annual average temperature in Beijing (1951–2002), which shows an increasing trend line; (c) time series after first-order differencing of data in Fig. 3b, which does not display a trend.

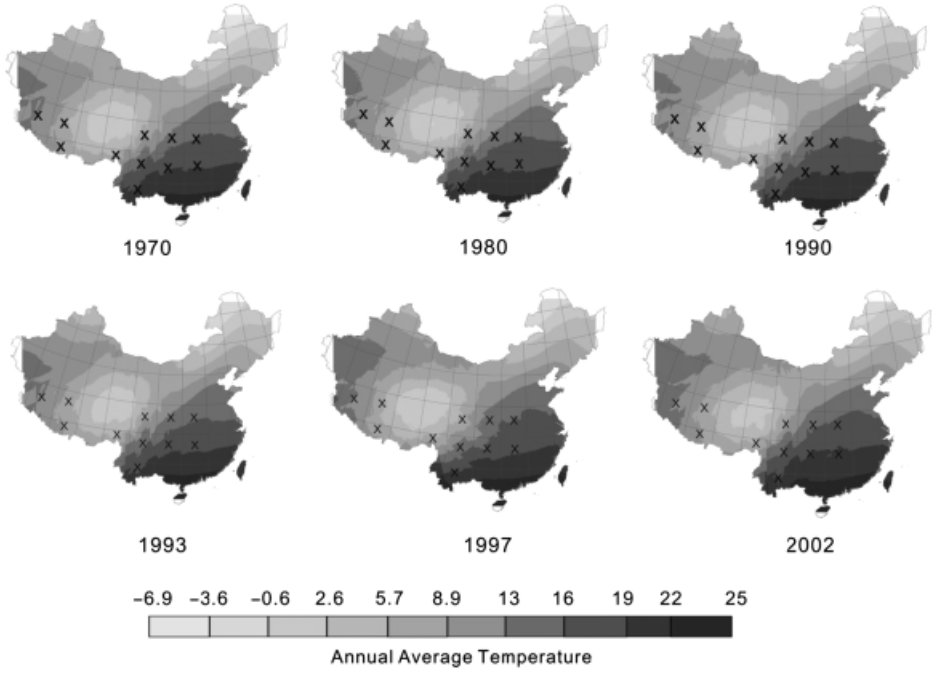

Figure 4. Spatial patterns of annual average temperature in different years. The annual average temperature is high in the southeast and low in the northwest of the study area.

highlighted by ''X.'' We included these areas purely for visualization effect and, in fact, excluded them from the analysis because we used observed rather than interpolated (kriged) values when estimating parameters of the ANN and STARIMA models. Moreover, we did not use the temperature data of provincial monitoring stations to improve the number of points, because those data are not as precise and reliable as data from the national meteorological stations. These considerations led us to use ANN for our space–time trend modeling, completing the exploratory space–time analysis task of stage 2.

An ANN model to predict deterministic space–time trends

An ANN model is implemented with an MLP network of three input nodes (coordinates of x, y, and t), one hidden layer, and one output node (see Fig. 1). Cross-validation is performed to select the optimal number of hidden nodes, and six hidden nodes were chosen based on a best-fit value. Here time t is relative time rather than absolute time (year), for computational convenience. The relative time is defined as

$$t_j = 1 + j/n \quad (j = 1, 2, \cdots, n)\,, \tag{6}$$

where n represents the total sequence length T. Input data were normalized to a specified range $[-1, 1]$. The sigmoid function was chosen as the activation function

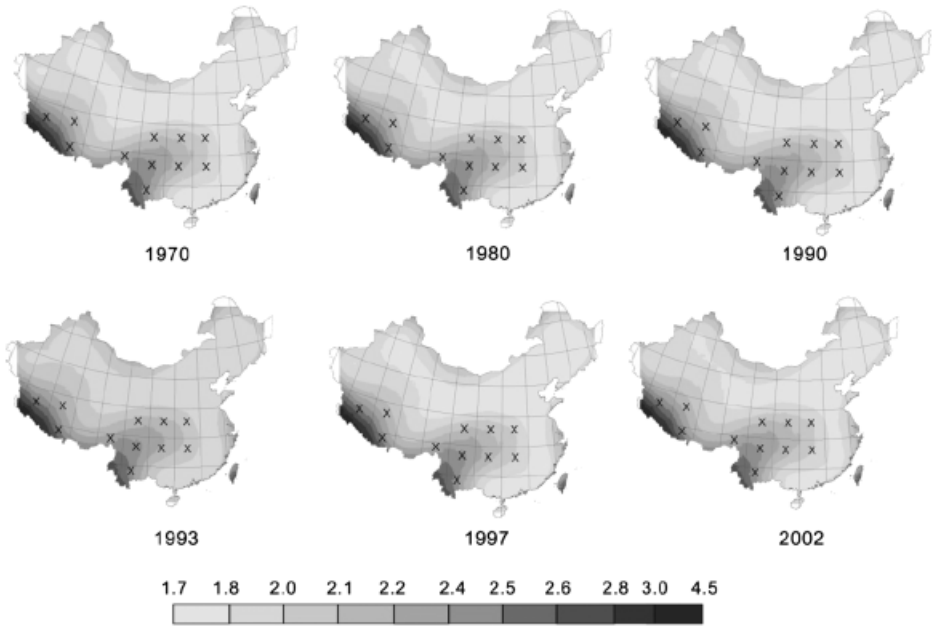

Figure 5. Kriging standard error maps for different years. The southwest part of the study area has larger deviations than the southeast and the northeast due to relatively few national stations in Tibet and in the desert areas of Sinkiang (Fig. 3a).

in the hidden layer and the output layer to implement nonlinear transformation within the output range of $[-1, 1]$.

Training data set analysis was performed on a 3.4 G MHz HP workstation and required 8 s of central processor unit time. The fitted results for 1970, 1980, and 1990 are presented in Fig. 6, which shows that the ANN model captured nonlinear deterministic space–time trends. These maps are very similar and indicate

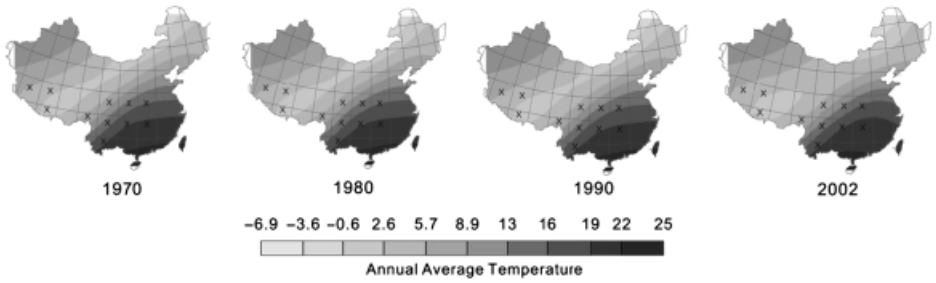

Figure 6. Nonlinear space–time trends captured by the ANN model. Fitted values in years 1970, 1980, and 1990. The three maps are very similar, indicating that the global deterministic space–time variation structures appear to be fairly stable over time. Besides, the predicted values in year 2002 displays similar patterns as in the three fitted years. These maps also show that low-temperature areas are decreasing, and high-temperature areas are increasing, indicating that the annual average temperature increased slightly over time.

that the global deterministic space–time variation structures appear to be fairly stable over time. We also find that low-temperature areas are decreasing and that high-temperature areas are increasing, which indicates that the annual average temperature in the entire study area increased slightly over time.

The semivariogram analysis of the ANN model residuals exhibits weak anisotropy, which shows that the strong anisotropy in the original data (Fig. 4) is largely removed (Fig. 6) after global nonlinear space–time trends are extracted by the ANN model. Results reported in Table 1 indicate that the C_0/sill ratios for all years are $<50\%$, indicating that the ANN model residuals have strong correlations and that a STARIMA model is needed as part of specification.

The STARIMA model/description of small-scale to capture stochastic space–time variances

Here we use STARIMA to capture the small-scale stochastic space–time variances by following Box–Jenkins three-step modeling procedure.

(1) *Building the spatial weight matrix*: The means of the partial sill ($C_1 = 3.86$), nugget ($C_0 = 6.57$), and isotropic range ($D = 1467.02$) (Table 1) were used to define a 137×137 row-standardized spatial weight matrix for the STARIMA model (see equation [5]). Although misspecified distances lead to biased results and spatial heterogeneity (Tiefelsdorf 2003), here Euclidean rather than spheroid distance has been used to define the spatial weight matrix since the formulas for semivariogram and inverse distance weighting (IDW) interpolation analysis are only valid for planar surfaces in ArcGIS (the package we used for the spatial analysis: since conducting this research, IDW has become possible for spheroids using R). Furthermore, the difference between these two distance measures is $<1\%$ in the geographic coordinate system UTM_WGS_1984. We think the discrepancy between the Euclidean and great circle distance measures is acceptable within the distance range threshold (D) of 1467 km.

(2) *Training/fitting a STARIMA model*: Here we follow the Box–Jenkins three steps to fit a STARIMA model. Fig. 7 presents ST-ACF (Fig. 7a) and ST-PACF (Fig. 7b) correlograms for the ANN residuals. The ST-ACFs (the moving average part, q) cut off at the first temporal lag for spatial lag zero and one; the ST-PACFs (the autoregressive part, p) appear to tail off at the first and second temporal lags for spatial lags zero and one. We hereby select the following STARIMA(2, 1) specification as the candidate model:

$$\begin{aligned}\xi_i(t) = {} & \phi_{10}\xi_i(t-1) + \phi_{11}W^{(1)}\xi_i(t-1) + \phi_{20}\xi_i(t-2) + \phi_{21}W^{(1)}\xi_i(t-2) \\ & - \theta_{10}\varepsilon_i(t-1) - \theta_{11}W^{(1)}\varepsilon_i(t-1) + \varepsilon_i(t).\end{aligned} \tag{7}$$

The parameter estimations for equation (7) are calculated with nonlinear least squares implemented in MATLAB 7.0. Table 2 tabulates the parameter estimates and their 95% confidence intervals; $\hat{\theta}_{10}$ and $\hat{\theta}_{11}$ are not statistically significant when the confidence level is set to 0.95. This result implies that these two parameters should be removed from equation (7). After reestimating, the reduced model, STAR

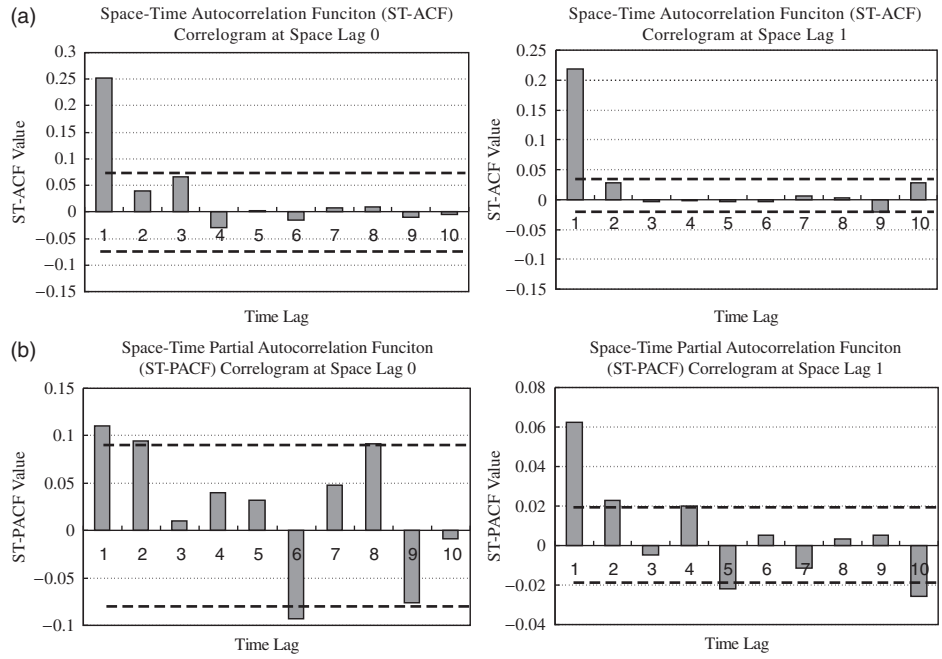

Figure 7. (a) ST-ACF and (b) ST-PACF correlograms for the hybrid model. Bars indicate the space–time autocorrelation and partial function value at time lags 1–10 and space lags zero and one, and transverse lines show the approximate 95% upper and lower confidence bounds, indicating that the space–time autocorrelation (a) and partial function (b) values are considered approximately zero.

(2, 0), was considered:

$$\begin{aligned}\xi_i(t) =& 0.7023\xi_i(t-1) + 0.2296W^{(1)}\xi_i(t-1) + 0.2624\xi_i(t-2) \\ &+ 0.0538W^{(1)}\xi_i(t-2) + \hat{\varepsilon}_i(t),\end{aligned} \tag{8}$$

where $\hat{\varepsilon}_i(t)$ is the estimated value for $\varepsilon_i(t)$.

(3) *Diagnosis*: Residuals for the STARIMA model are further assessed by evaluating its ST-ACF correlogram (see Table 3), whose mean (0.0017) and variance (0.00027) are close to the expected mean of zero and variance of $[N(T-s)]^{-1} = 0.0002$ (here $T = 42$, $N = 137$, and $s = 2$). This result indicates that the candidate STARIMA model captures the majority of the small-scale stochastic space–time variances in the ANN residuals. Through computation of the fitted results of the STARIMA model plus of the ANN fitted results (see equation [1]), we obtain the fitted results of the hybrid model from 1953 to 1992, completing stage 3 of the procedure.

Model validation

The final stage is validation of the hybrid model. Only one-step-ahead forecasting is considered. The root mean squared error (RMSE) is selected as the forecasting

Table 2 Parameter Estimates for the Hybrid Model

	Variable	φ_{10}	φ_{11}	φ_{20}	φ_{21}	θ_{10}	θ_{11}
Hybrid (candidate)	Coefficient	0.7024	0.2295	0.2622	0.0531	0.0255	− 0.0081
	t-statistic	12.871	6.228	6.409	2.372	0.338	− 0.075
	Probability	0.0000	0.0000	0.0000	0.0113	0.368	0.529
	95% confidence interval	[0.6812, 0.7251]	[0.2045, 0.2421]	[0.2352, 0.2891]	[0.0311, 0.0786]	[− 0.0182, 0.0475]	[− 0.0633, 0.0478]
Hybrid (update)	Coefficient	0.7023	0.2296	0.2624	0.0538	—	—
	t-statistic	12.87	6.2283	6.412	2.376		
	Probability	0.0000	0.0000	0.0000	0.018		
	95% confidence interval	[0.6809, 0.7252]	[0.2043, 0.2425]	[0.2351, 0.2903]	[0.0308, 0.0792]		

Table 3 Space–Time Autocorrelations (ST-ACFs) for the Hybrid Model Residuals

Space lag (h) Time lag (k)	0	1
1	0.02925	0.02138
2	0.01424	− 0.02198
3	− 0.01111	− 0.01897
4	0.00969	− 0.01759
5	− 0.00786	0.01651
6	− 0.00259	0.01138
7	− 0.00183	0.00941
8	0.00152	0.00752
9	− 0.00128	− 0.00911
10	0.00079	0.00388

accuracy measure. Fig. 6 shows nonlinear space–time trend maps for forecasting in the year 2002. Compared directly with the results from other years shown in Fig. 6, we find that the areas that have higher annual average temperatures continue to increase.

Although direct comparisons are not possible between hybrid STARIMA and ordinary STARIMA, we use an ordinary STARIMA model as a benchmark to demonstrate the effectiveness of the formulated hybrid model. We used the data after first-order differencing (which removes the trend in time, as shown in Fig. 3c) to be fitted with an ordinary STARIMA model. We did not implement higher-order differencing or multipolynomial STARIMA models because calibration of the parameters is far too complicated and no literature exists about such STARIMA models.

The ST-ACF and ST-PACF correlograms for the ordinary STARIMA model are portrayed in Fig. 8, which indicates that the ST-ACFs (Fig. 8a) cut off at the second temporal lag for spatial lags zero and one, implying a moving average order $q = 2$, and that the ST-PACFs (Fig. 8b) tail off at the first temporal order for spatial lags zero and one, implying an autoregressive parameter $p = 1$. Thus, the following STARIMA(1, 2) specification is regarded as the candidate model:

$$\begin{aligned}\xi_i(t) =& \phi_{10}\xi_i(t-1) + \phi_{11}W^{(1)}\xi_i(t-1) - \theta_{10}\varepsilon_i(t-1) - \theta_{11}W^{(1)}\varepsilon_i(t-1) \\ & - \theta_{20}\varepsilon_i(t-2) - \phi_{21}W^{(1)}\varepsilon_i(t-2) + \varepsilon_i(t).\end{aligned} \tag{9}$$

Results summarized in Table 4 reveal that $\hat{\varphi}_{11}$ and $\hat{\theta}_{11}$ are statistically nonsignificant. Therefore, the ordinary STARMA model is revised to the following form:

$$\begin{aligned}\xi_i(t) =& -0.4096\xi_i(t-1) + 0.1982\hat{\varepsilon}_i(t-1) - 0.042\hat{\varepsilon}_i(t-2) \\ & - 0.0594W^{(1)}\hat{\varepsilon}_i(t-2) + \hat{\varepsilon}_i(t).\end{aligned} \tag{10}$$

The residuals of the predictions from the ordinary STARIMA model closely mimic a normal distribution (Table 5), indicating that the model efficiently describes the small-scale space–time variances.

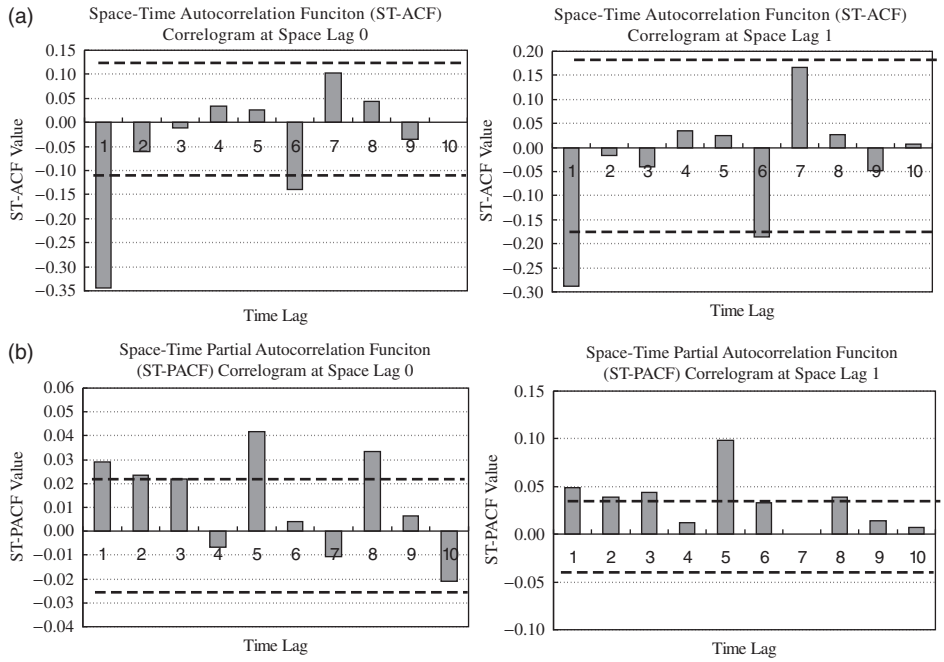

Figure 8. ST-ACF (a) and ST-PACF (b) correlograms for the ordinary STARIMA model.

Fig. 9 presents the prediction maps for the hybrid model and the ordinary STARIMA model, and Fig. 10 compares the forecast results from these two models. Table 6 summarizes their accuracy measures using RMSE. Alternatively, we may also use resampling techniques to estimate prediction variance of the hybrid model.

Results reported here show that the hybrid framework largely improves the fitted and slightly improves the forecasting accuracy obtained with the ordinary STARIMA model. In addition, the RMSE errors of both the hybrid and the STARIMA models become increasingly large over time. As portrayed in Fig. 10, in 1993 the forecasting maps of both models are close to the actual map (Fig. 4), but in 1997 and 2002, they deviated noticeably from the actual maps. This indicates that their performance in the short term is better than that of metaphase and long-term forecasting for the two models.

To compare further the modeling ability of the hybrid model with the ordinary STARIMA model vis-à-vis spatial heterogeneity or nonstationarity, both models' residuals were analyzed using the Gaussian semivariogram function. We can see that the sill variances of both models' residuals for all years sharply decrease, from a maximum of 18.001 in Table 1 to a minimum of 0.0665 in Table 7, which shows that both models explain most of the variance in the space–time series. But the sill variances of the hybrid model residuals (Table 7) are smaller than those of the STARIMA model residuals. In addition, the C_0/sill ratios of the hybrid model are close to 50%, indicating that their residuals have weaker spatial correlation than

Table 4 Parameter Estimates for the Ordinary STARIMA Model

Variable	φ_{10}	φ_{11}	θ_{10}	θ_{11}	θ_{20}	θ_{21}
Coefficient	−0.3375	0.0417	−0.1253	0.0405	−0.2278	0.0697
t-statistic	−6.892	2.231	−4.824	2.218	−5.647	2.914
Probability	0.0000	0.2038	0.0000	0.2001	0.0113	0.0321
95% confidence interval	[−0.3619, −0.3131]	[−0.0226, 0.1061]	[−0.1512, −0.0994]	[−0.0215, 0.1026]	[−0.2525, −0.2031]	[−0.0059, 0.1336]
Coefficient	−0.4096	—	−0.1982	—	−0.042	0.0594
t-statistic	−7.143		−5.021		−2.326	2.847
Probability	0.0000		0.0000		0.0218	0.0364
95% confidence interval	[−0.427, −0.3814]		[−0.2134, −0.1725]		[−0.0671, −0.0215]	[0.0872, 0.0306]

Table 5 Space–Time Autocorrelations (ST-ACFs) for the Ordinary STARIMA Model Residuals

Space lag (h) Time lag (k)	0	1
1	0.0302	−0.0790
2	−0.2877	−0.1752
3	−0.0223	−0.0503
4	0.0625	0.0280
5	−0.0154	−0.0240
6	−0.1179	−0.1544
7	0.0909	0.1665
8	0.0794	0.0817
9	−0.0461	−0.0745
10	−0.0681	−0.0710

those for the ordinary STARIMA model, for which the C_0/sill ratios were between 32.97% and 39.44%. This finding also implies that the hybrid model has a better explanatory capacity than the STARIMA model for spatial heterogeneity and non-stationarity. The reason for this may be that the nonlinear space–time trends in environmental space–time series data are better modeled by an ANN. This finding also explains why the hybrid model improves prediction accuracy only to a certain extent, and, in this case, not to a great extent.

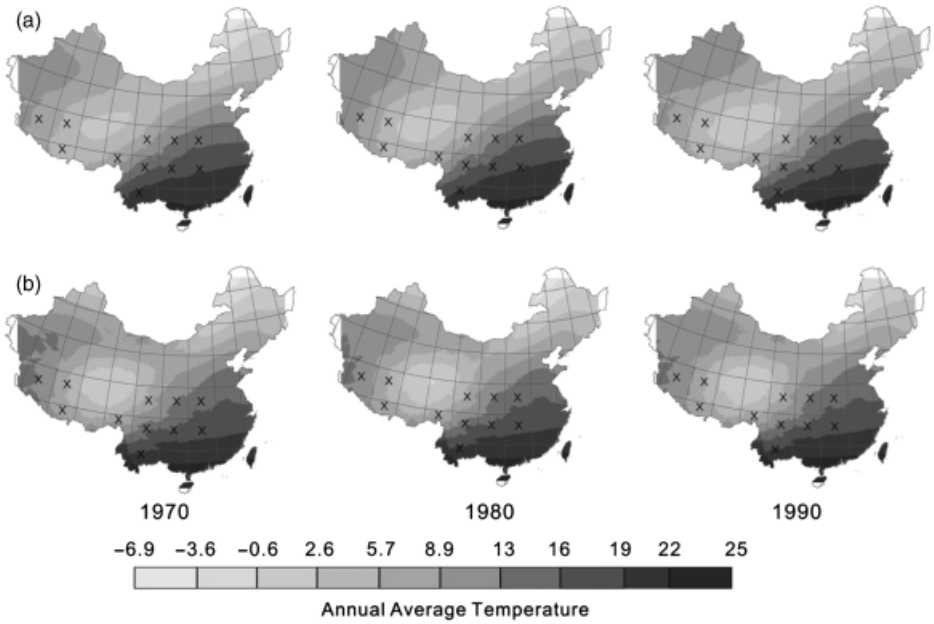

Figure 9. Comparison of the fitted results produced by (a) the ordinary STARIMA model and (b) the hybrid model.

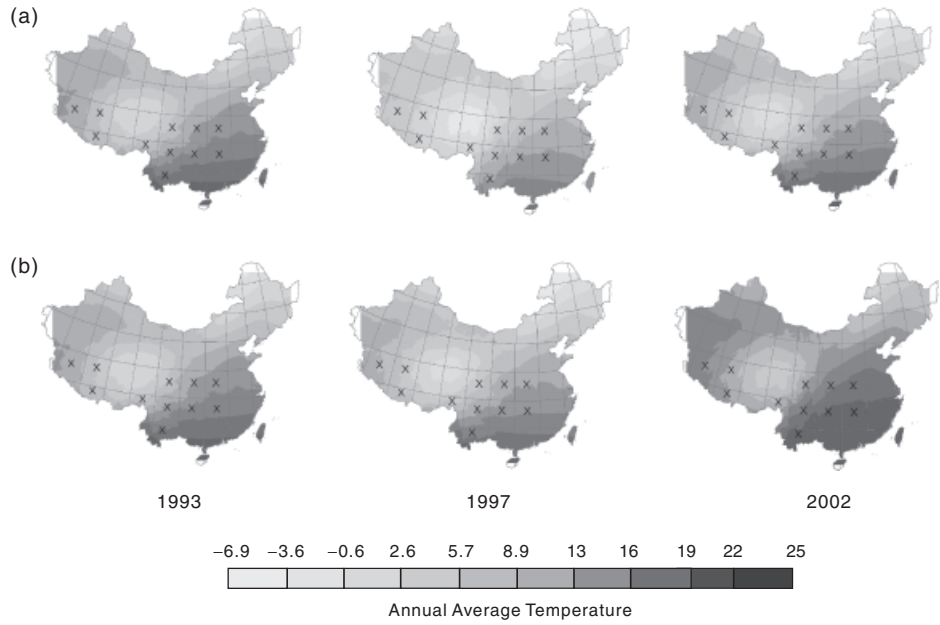

Figure 10. Comparison of the forecasting results produced by (a) the ordinary STARIMA model and (b) the hybrid model.

Conclusions and discussion

A hybrid model is developed in this article to describe a space–time series of environmental data exhibiting strong spatial and/or temporal nonlinearity and nonstationarity. This model can be considered part of a semiparametric method, because it combines a data-driven ANN model, used to extract nonlinear deterministic space–time trends, with a STARIMA model, to account for stochastic

Table 6 Accuracy Measures for the Fitted and Forecasting Results

	Hybrid RMSE	STARIMA RMSE	Accuracy improved %
Fitted			
1970	0.385	1.762	78.13
1980	0.489	1.894	74.17
1990	0.677	1.97	65.55
Forecasting			
1993	0.615	0.649	5.24
1997	3.086	3.377	8.62
2002	6.703	7.728	13.3

Table 7 Comparison of the Residuals Via Isotropic Semivariogram Analysis

	Range (km)	Partial sill (C)	Nugget (C_0)	Sill (C_0+C)	C_0/sill (%)
Hybrid					
1970	1052.3	0.0541	0.0544	0.1085	50.12
1980	1013.4	0.0919	0.0903	0.1822	49.57
1990	1125.2	0.1301	0.1203	0.2504	48.05
1993	1160.3	0.0382	0.0283	0.0665	42.56
1997	1314.1	0.1443	0.1311	0.2754	47.60
2002	1234.2	0.2189	0.1698	0.3887	43.68
STARIMA					
1970	1119.7	2.9511	1.9219	4.8730	39.44
1980	1121.9	2.7705	1.6728	4.4433	37.65
1990	1369.7	2.6279	1.4913	4.1192	36.20
1993	1093.2	1.9458	1.1971	3.1429	38.09
1997	1511.5	2.8066	1.4771	4.2837	34.48
2002	1160.8	2.6081	1.2826	3.8907	32.97

space–time variations. Although this is an ad hoc approach, the significance of such a framework is fourfold:

(1) First, unlike most frameworks, which are either spatially or temporally dominated, or first separated and then combined linearly (Li, Dunham, and Xiao 2003; Cheng and Wang 2008), this hybrid model furnishes an integrated specification that handles both space and time simultaneously and seamlessly.

(2) Second, a four-stage procedure is proposed for analyzing and modeling space–time series of environmental data that includes data preparation, exploratory space–time analysis, model training, and model validation. In particular, the exploratory space–time analysis is discussed systematically, as are the methods used to explore space–time stationarity, the spatial and temporal correlations, and the validity of formulated models, all of which can be used as a guideline and as the first step in all space–time analysis.

(3) Third, one artificial network model, the MLP network, was applied to model global deterministic space–time structures, solving the problem of spatial nonlinearity and nonstationarity for conventional statistical methods, while also extending the ANN application from spatial analysis to space–time modeling.

(4) Fourth, using a semivariogram to define a spatial weight matrix largely enables the STARIMA model to: (i) analyze space–time series that are continuous in space; (ii) calibrate spatial heterogeneity and local nonstationarity in environmental data; and (iii) accommodate spatial autocorrelation in different directions.

However, the designed ANN model cannot capture all deterministic space–time structures, and the candidate STARIMA model does not necessarily fully

explain all stochastic space–time variability. Other factors may exist (such as altitude, humidity, wind speed, or distance from the sea) that affect degrees of temperature and should be included in a model specification. For this reason, a logical extension of the methodology would be to incorporate explanatory variables (altitude or humidity) into both an ANN and a STARIMA model. Furthermore, impacts of spatial and temporal autocorrelation on an ANN is not discussed here but is addressed in Cheng and Wang (2009), who report that considering spatial associations in the calibration of an ANN improves computational efficiency and goodness of fit.

With the advance in space–time series data collection, space–time data analysis theory has became increasingly popular, especially with respect to the development of integrated space–time dynamic models that take into account the special features of geographic space–time data, such as space–time autocorrelation, variability, and heterogeneity (Hepple 1978; Baltagi 2005; Heuvelink and Griffith 2010). The development of integrated space–time data analysis and modeling methodology, rather than the separate processing of space and time, is indispensable. We think that the *hybrid* framework developed here represents a useful attempt in this aspect. As Openshaw (1999) notes with the first sentence of the text on his web page, ''the immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS, emphasises the importance of developing data driven inductive approaches to geographical analysis and modeling, to facilitate the creation of new knowledge and aid the processes of scientific discovery.'' With the advances of data-driven inductive approaches, geographical analysis and modeling should be able to promote the creation of new space–time data analysis and modeling methodology. Thus, a space–time artificial neural network (STANN) should be developed that will be able to overcome the limits of the MLP network and the STARIMA model specification by tackling nonlinearity and nonstationarity in space–time series. This will be the direction of our own future research.

Acknowledgements

The research is supported by 973 Program (2006CB701305) and 863 Program (2009AA12Z206), China, and EPSRC (EP/G023212/1), U.K. Very detailed and constructive comments from the editor, as well as those from three anonymous referees are highly appreciated, which greatly improved the readability of the article. However, the authors are solely responsible for any mistakes that remain.

References

Baltagi, B. (2005). *Econometric Analysis of Panel Data*. Hoboken, NJ: Wiley.

Bennett, R. J. (1975). ''The Representation and Identification of Spatio-Temporal Systems: An Example of Population Diffusion in North-West England.'' *Transactions of the Institute of British Geographers* 66, 73–94.

Box, G. E. P., G. M. Jenkins, and G. Reinsel. (1994). *Time Series Analysis, Forecasting and Control*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

Brundson, C. F., A. S. Fotheringham, and M. Charlton. (1996). ''Geographically Weighted Regression—A Method for Exploring Spatial Non-Stationarity.'' *Geographical Analysis* 28, 281–98.

Cheng, T., and J. Wang. (2008). ''Integrated Spatio-Temporal Data Mining for Forest Fire Prediction.'' *Transactions in GIS* 12, 591–611.

Cheng, T., and J. Wang. (2009). ''Accommodating Spatial Associations in DRNN for Space–Time Analysis.'' *Computers, Environment and Urban Systems* 33, 409–18.

Cliff, A. D., and J. K. Ord. (1975). ''Space–Time Modelling with an Application to Regional Forecasting.'' *Transactions of the Institute of British Geographers* 64, 119–28.

Cooper, S. D., L. Barmuta, O. Sarnelle, K. Kratz, and S. Diehl. (1997). ''Quantifying Spatial Heterogeneity in Streams.'' *Journal of the North American Benthological Society* 16, 174–88.

Crespo, J. L., M. Zorrilla, P. Bernardos, and E. Mora. (2007). ''A New Image Prediction Model Based on Spatio-Temporal Techniques.'' *The Visual Computer* 23, 419–31.

De Luna, X., and M. G. Genton. (2005). ''Predictive Spatio-Temporal Models for Spatially Sparse Environmental Data.'' *Statistica Sinica* 15, 547–68.

Griffith, D., and F. Csillag. (1993). ''Exploring Relationships between Semi-Variogram and Spatial Autoregressive Models.'' *Papers in Regional Science* 72, 283–95.

Haas, T. (1998). ''Multivariate Spatial Prediction in the Presence of Non-Linear Trend and Covariance Non-Stationarity.'' *Environmetrics* 7, 145–65.

Haining, R. P. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. New York: MacMillan.

Hepple, L. (1978). ''The Econometric Specification and Estimation of Spatial-Temporal Models.'' In *Timing Space and Spacing Time. Vol. 3, Time and Regional Dynamics*, 66–81, edited by T. Carlstein, D. Parkes, and N. Thrift. London: Edward Arnold.

Heuvelink, G., and D. Griffith. (2010). ''Space–Time Geostatistics for Geography: A Case Study f Radiation Monitoring across Parts of Germany.'' *Geographical Analysis* 42, 161–79.

Hornik, K., M. Stinchombe, and H. White. (1990). ''Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks.'' *Neural Networks* 3, 551–60.

Isaaks, E. H., and R. M. Srivastana. (1989). *An Introduction to Applied Geostatistics*. Oxford: Oxford University Press.

Kamarianakis, Y., and P. Prastacos. (2005). ''Space–Time Modeling of Traffic Flow.'' *Computers and Geosciences* 31, 119–33.

Kanevski, M., R. Arutyunyan, L. Bolshov, V. Demianov, and M. Maignan. (1996). ''Artificial Neural Networks and Spatial Estimation of Chernobyl Fallout.'' *GeoInformatics* 7, 5–11.

Kanevski, M., and M. Maignan. (2004). *Analysis and Modelling of Spatial Environmental Data*. Lausanne: EPFL Press.

Li, Z., M. H. Dunham, and Y. Xiao. (2003). ''STIFF: A Forecasting Framework for Spatio-Temporal Data.'' In *Mining Multimedia and Complex Data, Lecture Notes in Computer Science*, Vol. 2797: 183–98, edited by O. R. Zaïane, S. J. Simoff, and C. Djeraba. Berlin: Springer-Verlag.

Martin, R. L., and J. E. Oeppen. (1975). ''The Identification of Regional Forecasting Models Using Space–Time Correlation Functions.'' *Transactions of the Institute of British Geographers* 66, 95–118.

Mitchell, T. M. (2003). *Machine Learning*. New York: McGraw-Hill.

Openshaw, S. (1999). ''Geographical Data Mining: Key Design Issues.'' In *The Proceedings of the Fourth International Conference on GeoComputation*, Mary Washington College, Fredericksburg, VA, 25–28. Available at http://www.geovista.psu.edu/sites/geocomp99/gc99/051/gc_051.htm (accessed on January 6, 2011).

Pace, R. K., R. Barry, J. Clapp, and M. Rodriguez. (1998). ''Spatio-Temporal Autoregressive Models of Neighborhood Effects.'' *Journal of Real Estate Finance and Economics* 17, 15–33.

Pfeifer, P. E., and S. J. Deutsch. (1980). ''A Three-Stage Iterative Procedure for Space–Time Modelling.'' *Technometrics* 22, 35–47.

Tiefelsdorf, M. (2003). ''Misspecifications in Interaction Model Distance Decay Relations: A Spatial Structure Effect.'' *Journal of Geographical Systems* 5, 25–50.