Adaptive Spatio-Temporal Models for Satellite Ecological Data

Carlo GRILLENZONI

This article develops models for environmental data recorded by meteorological satellites. In general, such data are continuously available for suitable space and time units and are intrinsically nonstationary. Space-time auto-regression (STAR) is a class of models that can be used in monitoring and forecasting, but it must be adapted to nonstationary processes. A set of adaptive recursive estimators is then proposed to estimate STAR parameters that change both over space and time. An extensive application to the normalized difference vegetation index (NDVI), for a region of sub-Saharan Africa, illustrates and checks the approach.

Key Words: Recursive least squares; Space-time autoregression; Unit roots; Varying parameters.

1. INTRODUCTION

Analysis of vegetation activity is fundamental in monitoring forest and agricultural resources. The normalized difference vegetation index (NDVI) is the main tool for accomplishing this task. Since 1980 it has been recorded daily by NASA meteorological satellites that have world-wide coverage. For this reason it is extensively used by the Food and Agricultural Organization (FAO) to monitor subtropical areas. Substantial increase of desertification in sub-Saharan countries is well documented (e.g., Tucker, Dregne, and Newcomb 1991). Desertification is held responsible for many social problems, such as famines and migrations toward Europe.

Vegetation activity is a continuous space-time process and NDVI data provide a space-time *lattice* system, in the sense that observations are available over a regular grid. Spatial resolution usually ranges from 1 to 8 kilometers, while the temporal one ranges from 10 days to 1 month. In this way, more aggregate datasets can be easily constructed. It should be noted that space-time processes cannot be assimilated to 3D random fields, because

Carlo Grillenzoni is Associate Professor of Statistics, University of Venice, IUAV, Department of Planning, 30135 Venice, Italy (E-mail: carlog@iuav.it).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 9, Number 2, Pages 158–180
DOI: 10.1198/1085711043541

time provides a natural causal ordering. As a consequence time series models are the basic schemes for representing such processes.

Space-time auto-regression (STAR)—see, for example, Ali (1979), Tjøstheim (1983), or Niu, McKeague, and Elsner (2003)—is a class of linear models suitable for lattice data. In this context, each observation is placed in linear relationship with previous and surrounding values to produce forecasts. To deal with environmental phenomena, such models must be adapted to their intrinsic nonstationarity. By this we mean presence of spatial and temporal trends, existence of spatial heterogeneity, slow changes due to natural evolution, and fast changes due to satellite substitution. All of these problems create a situation of time-varying parameters which is typical of time series analysis (e.g., Grillenzoni 1998). However, the original and unexplored problem here is the presence of the spatial dimension.

This article develops a statistical methodology for space-time systems with time-varying parameters. It extends the approach developed by Grillenzoni (1994, 1998) based on the unification of adaptive recursive estimators. We aim to apply this methodology both to monitor long-term changes of vegetation activity and to check improvement in short-term forecasts. Application concerns monthly data of a sub-Saharan region of central Africa and it will be developed throughout.

There are many works concerning the analysis of NDVI data in the form of spatial matrices (images). For example, Faivre and Fischer (1996) use highest resolution images to predict (estimate) crop reflectances by means of random coefficient regression models. However, such studies usually consider NDVI and related radiance data at a given instant of time. This article deals with the whole sequence of historical data to predict future values of vegetation.

The plan of this article is as follows. Section 2 provides detailed description of the dataset. Section 3 discusses STAR models with unit-roots. Section 4 derives recursive parameter estimators and Section 5 uses the estimators for forecasting.

2. DATA DESCRIPTION

The NDVI index is generated in NASA's Goddard space flight center (GSFC), by processing data of the advanced very high-resolution radiometer (AVHRR) aboard the NOAA satellites. Specifically, AVHRR is a five-channel instrument which measures reflectances and brightness temperatures, while NOAA 7, 9, 11, 14 are polar orbiter satellites, which have flown in the periods 1981–1985, 1985–1988, 1988–1994, 1994–2000.

Radiance data from the shorter wavelength channels, namely the visible (C1) and the near-infrared (C2) one, are combined into the ratio $\text{NDVI} = (\text{C1} + \text{C2})/(\text{C2} - \text{C1})$ which gives an indication of the vegetation activity. In fact, it is positively correlated with the absorption of red light by plant chlorophyll and the reflection of infrared radiation by water-filled leaf cells.

At GSFC a consistent processing algorithm is used to recalibrate all of the measurements. In particular, radiance data are corrected for atmospheric distortions caused by rayleigh scattering and ozone absorption, to produce "at-ground" reflectance estimates.

Moreover, information from the thermal channel (C5) is used for reducing cloud masking.

However, the entire process of data generation is still affected by a number of shortcomings: namely, the difficulty in accurately determining satellite orbit, the decreasing sensitivity of the AVHRR instrument over time, the selection of an early afternoon crossing time, which generally coincides with maximum cloud development (e.g., Emery, Brown, and Nowak 1989). In order to mitigate such problems, data are built by resampling and combining the highest vegetation values obtained over 10 days of measurements. Finally, NDVI data are mapped to Hammer-Aitoff projection and aggregated into spatial units ("pixels") of size 7.638 kilometers (km), which allow suitable display on a 1024 by 1280 screen.

The sample. At the Internet address http://daac.gsfc.nasa.gov, an online ftp service is available for retrieving NDVI data in binary format. For the needs of the present study we considered a rectangular sub-Saharan region placed at the borders of Chad, Sudan, and the Central Africa Republic. To be precise, the region has latitudes (4.5, 16.5) north and longitudes (22.9, 25.9) east; it lies 1,000 km southeast of lake Chad and includes mountains such as Gebel Marra (3,088 m). In terms of 7.6 km "pixels," it corresponds to a matrix with 168 rows and 45 columns. Data are of monthly type and cover the period from September 1981 to December 1999.

The entire dataset can be organized as a three-way matrix $\mathbf{Z} = \{Z_{ijs}\}$ of size $168 \times 45 \times 220$, where i, j, s indicate latitude, longitude, and time in months. To have a synthetic view of the NDVI phenomenon, data can be averaged both in space and in time. Indicating with t time in years and noticing that for the first year (1981) only four months were available, we have

$$\begin{aligned}
\text{Spatial Averages}: \quad \bar{Z}_{\bullet\bullet s} &= \frac{1}{168 \cdot 45} \sum_{i=1}^{168} \sum_{j=1}^{45} Z_{ijs}\,, \quad s = 1, 2 \ldots 220 \\
\text{Yearly Averages}: \quad \bar{Z}_{ijt} &= \frac{1}{12} \sum_{r=5+(t-1)\cdot 12}^{(t\cdot 12)} Z_{ijr}\,, \quad t = 1, 2 \ldots 18 \\
\text{Longitudinal Averages}: \quad \bar{Z}_{i\bullet t} &= \frac{1}{45} \sum_{j=1}^{45} \bar{Z}_{ijt}\,, \qquad i = 1, 2 \ldots 168. \qquad (2.1)
\end{aligned}$$

The results are displayed in Figure 1; in particular, Figure 1(a) provides the time path of the spatial averages $\bar{Z}_{\bullet\bullet s}$, and Figure 1(b) shows the latitudinal path of the longitudinal averages $\bar{Z}_{i\bullet t}$. Finally, Figure 1(c) displays in 3D the 1999 average matrix $\bar{Z}_{ij18}$.

As may be expected, NDVI data have strong seasonality, a random walk pattern at time and longitudinal level, and a significant latitudinal trend, because it increases from north (savanna) to south (forest). At spatial level, vegetation activity is also conditioned by hills and mountains which capture the humidity, as happens at the point $(i, j) = (50, 20)$ in Figure 1(d,e), which is the site of Gebel Marra. We may also appreciate that, seasonality apart, there is *not* a marked negative time trend in such data. Also, the decline in recent years (from $s = 160$ or $t = 14$) could be attributed to the last change of satellite which in fact occurred in September 1994.

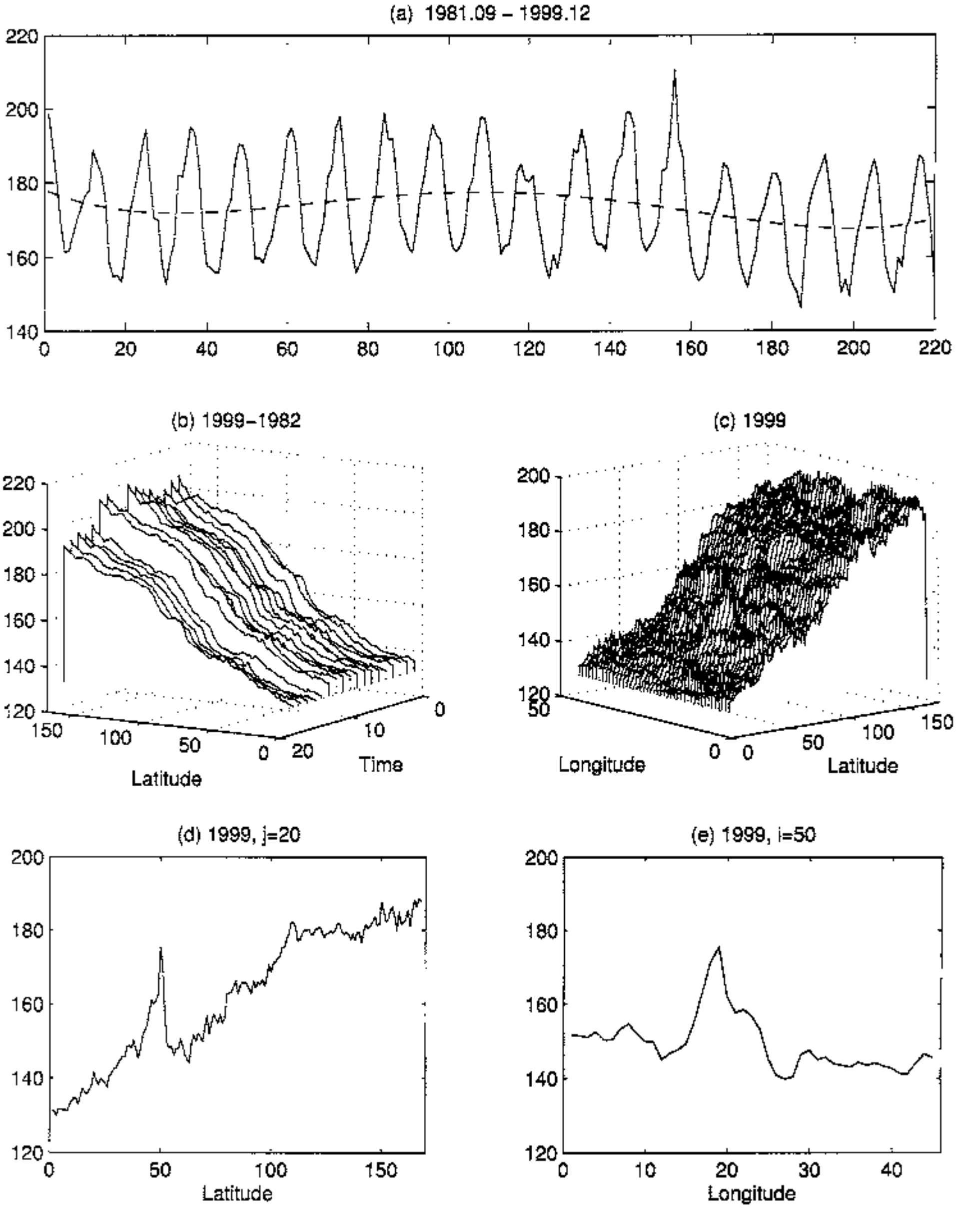

Figure 1. *Behavior of NDVI data: (a) Time path of spatial averages and their trend; (b) latitudinal path of yearly longitudinal averages; (c) 3D profile of the 1999 matrix; (d,e) plot of the series $\bar{Z}_{i20t}$ and $\bar{Z}_{50jt}$ at $t = 18$.*

Table 1. Third-Order Dependence Scheme of a Space-Time Process on a Regular Lattice. The number in parentheses indicate distances from the central cell.

↑	.	.	.	.	.	.
i	.	.	$Z_{i+2jt-1}$ (3)	.	.	.
.	.	$Z_{i+1j-1t-1}$ (2)	$Z_{i+1jt-1}$ (1)	$Z_{i+1j+1t-1}$ (2)	.	.
.	$Z_{ij-2t-1}$ (3)	$Z_{ij-1t-1}$ (1)	$\mathbf{Z}_{ijt}$ (0)	$Z_{ij+1t-1}$ (1)	$Z_{ij+2t-1}$ (3)	.
.	.	$Z_{i-1j-1t-1}$ (2)	$Z_{i-1jt-1}$ (1)	$Z_{i-1j+1t-1}$ (2)	.	.
.	.	.	$Z_{i-2jt-1}$ (3)	.	.	.
.	.	.	.	.	j	→

In the subsequent statistical analysis the dataset will be spatially subsampled and temporally aggregated. In practice, a three-way matrix of size 99 × 33 × 18 of yearly data $\{Z_{ijt}\}$ will be extracted from the original matrix **Z**. This was done both in order to reduce the amount of computations (which in certain cases are particularly demanding) and to exclude near-desert and permanent forest zones.

3. UNSTABLE STAR MODELS

Satellite NDVI data form a lattice dependent process, because the observations on the cells are expected to be correlated spatially and temporally. However, while in the time dimension there is a well-defined causal direction (from past to present), in space there is no preferred direction. This situation poses a problem in using multidimensional spatial autoregression (SAR) (e.g., Tjøstheim 1983) in modeling such processes. In fact, data matrix **Z** of the previous section does not coincide with the realization of a 3D spatial process, for example, a water or an oil field.

Following the causal ordering of time, we expect observations $\{Z_{ijt}\}$ (where i, j, t indicate latitude, longitude, and time), to depend on its past and surrounding values, as displayed in Table 1. Assuming first-order dependence in time and second order in space, the STAR representation becomes

$$\begin{aligned} Z_{ijt} &= \phi_1 Z_{ijt-1} + \phi_2 Z_{i-1j-1t-1} + \phi_3 Z_{i-1jt-1} + \cdots + \phi_9 Z_{i+1j+1t-1} + a_{ijt} \\ &= \boldsymbol{\phi}' \mathbf{x}_{ijt} + a_{ijt}, \quad a_{ijt} \sim \text{IID}(0, \sigma_a^2), \quad Z_{ij0} = a_{ij0}, \end{aligned} \tag{3.1}$$

where $\{a_{ijt}\}$ is a white noise process, $\boldsymbol{\phi}$ is the vector of parameters, and $\mathbf{x}'_{ijt} = [\, Z_{ijt-1}, Z_{i-1j-1t-1} \ldots Z_{i+1j+1t-1} \,]$ is the vector of "regressors." In practice, the process in each cell depends first on its previous value (as in standard time series) and then on the past values in the nearest neighbor cells.

An extension of the model (3.1) to general space-time lags is given by

$$Z_{ijt} = \sum_{l=1}^{p} \sum_{k=-q}^{q} \sum_{h=-q}^{q} \phi_{khl} \, Z_{i-k\,j-h\,t-l} \; + \; a_{ijt} \tag{3.2}$$

which can be denoted as STAR$(2q, p)$, where $p \geq 1$, $q \geq 0$ are the orders of the temporal and spatial parts. For simplicity, in the models (3.1) and (3.2) we have omitted instantaneous

($\phi_{kh0}\, Z_{i-k\, j-h\, t}$) and moving average ($\theta_{khl}\, a_{i-k\, j-h\, t-l}$) components, because our primary interest is the analysis of nonstationarity.

To achieve parametric parsimony in the representation (3.2), one could aggregate and average the spatial terms which are present at the same time lag, namely

$$Z_{ijt} = \sum_{l=1}^{p} \Phi_l\, \bar{Z}_{ijt-l} \,+\, a_{ijt}\,, \quad \bar{Z}_{ijt-l} = \frac{1}{(2q+1)^2} \sum_{k=-q}^{q} \sum_{h=-q}^{q} Z_{i-k\, j-h\, t-l}, \tag{3.3}$$

where $\bar{Z}_{ijt}$ is the mean process on a square. Equivalence of (3.3) and (3.2) follows in the case of homogeneous parameters at spatial level: $\phi_{khl} = \Phi_l$; a condition which may be realistic for noncentral components ($k, h \neq 0$). Equation (3.3) resembles a pure time series model and, indeed, has stability properties which are similar to an AR(p) process. We now discuss this aspect with some detail.

3.1 STATISTICAL PROPERTIES

Causality and stability are useful features for space-time models and are important conditions for the statistical properties of parameter and forecast estimates. As a general definition we adopt the following one.

Definition. A stochastic dynamical system is *causal* if the output only depends on past and present values of the input. It is *stable* if to inputs bounded in probability there correspond outputs with the same feature.

As regards space-time models, (3.1) and (3.2) are causal and stable if admit a *unilateral* and *convergent* decomposition of the type

$$Z_{ijt} = \sum_{l=1}^{\infty} \sum_{k=-lq}^{lq} \sum_{h=-lq}^{lq} \psi_{khl}\, a_{i-k\, j-h\, t-l} + a_{ijt}.$$

This representation can be obtained by solving the models with initial conditions $Z_{ij0} = a_{ij0}$; the weights ψ_{khl} depend on the parameters ϕ_{khl} in a complicated way. Absence of instantaneous components $Z_{i\pm k\, j\pm h\, t}$ in the vector of regressors $\mathbf{x}_{ijt}$ enables the above to be unilateral; its convergence depends on the roots of the transfer function $\phi(z_1, z_2, z_3) = 1 - \sum_l \sum_k \sum_h \phi_{khl}\, z_1^k\, z_2^h\, z_3^l$.

Proposition 1. Model (3.1) is stable if $\left(\sum_{k=1}^{9} |\phi_k| \right) < 1$, and (3.2) is stable if the polynomial (3.4) has roots outside the unit circle $|z| = 1$

$$\Phi(z) = \left(1 - \sum_{l=1}^{p} \Phi_l\, z^l \right), \quad \Phi_l = \sum_{k=-q}^{q} \sum_{h=-q}^{q} \phi_{khl}. \tag{3.4}$$

See Tjøstheim (1978, p. 142) and Ali (1979, p. 514) for the proof. The above recalls a well-known result of time series analysis and is motivated by the fact that STAR models are firstly temporal models. Moreover, under spatial homogeneity (isotropy) the individual processes $\{Z_{ijt}\}$ form a collection of *proxy* time series, which can be aggregated as in the model (3.3).

Although conditions in (3.4) are relatively simple, it is difficult to derive specific constraints on the coefficients ϕ_{khl}. Given a certain model, it is more convenient to check its stability by means of repeated simulation experiments. In any event, stability itself has a limited practical value: First of all, real data contain spatial and temporal trends, which can be represented only through unstable models (see simulation below). Second, as is known in time series analysis, unstability does not affect consistency of parameter estimates (e.g. Grillenzoni 1998), and this holds even in spatial processes (see Bhattacharyya, Richardson, and Franklin 1997).

Estimation of STAR models has mainly focused on Yule-Walker and maximum likelihood methods (see De Luna and Genton 2002). Owing to its flexibility, we are interested in the least squares (LS) approach. Such estimator minimizes the functional $Q_N = \sum_t \sum_i \sum_j a_{ijt}^2$, and its expression is simply given by

$$\hat{\phi}_N = \left(\sum_{t=1}^{n_t} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \mathbf{x}_{ijt}\, \mathbf{x}'_{ijt} \right)^{-1} \sum_{t=1}^{n_t} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \mathbf{x}_{ijt}\, Z_{ijt}, \tag{3.5}$$

where for model (3.2), $N = (n_t - p) \times (n_i - 2q) \times (n_j - 2q)$ is the effective sample size and $\mathbf{x}'_{ijt} = [\, Z_{ijt-1} \ldots Z_{i-q\,j-q\,t-1} \ldots Z_{i+q\,j+q\,t-p} \,]$ is the vector of regressors.

Proposition 2. Under model stability and causality, the estimator (3.5) is asymptotically equivalent to the Gaussian maximum likelihood method, and

$$\sqrt{N} \left(\hat{\phi}_N - \phi \right) \quad \xrightarrow{\text{L}} \quad \text{N} \left[\mathbf{0},\, \text{E}(\mathbf{x}_{ijt}\, \mathbf{x}'_{ijt})^{-1} \sigma_a^2 \right]. \tag{3.6}$$

For the proof see Ali (1979) or Tjøstheim (1983, p. 569). Extension of result (3.6) to unstable models has been studied only for particular cases. For the spatial model $Z_{ij} = \alpha Z_{i-1\,j} + \beta Z_{i\,j-1} - \alpha\beta Z_{i-1\,j-1} + a_{ij}$, Bhattacharyya et al. (1996, 1997) have shown that (3.6) holds even under the "unit root" condition $\alpha = \beta = 1$, by replacing $\sqrt{N}$ with $N^{3/2}$. This situation is somewhat different from time series model, where in presence of unit roots neither the distribution nor the dispersion of the estimator can be specified (e.g., Dikey and Fuller 1979).

Simulations. To check the properties of (3.5) for unstable STAR models, simulation experiments are useful. We then considered the first-order system

$$\begin{aligned} Z_{ijt} \;&=\; \pm 0.21 Z_{ijt-1} + 0.11 Z_{i+1\,j\,t-1} + 0.31 Z_{i-1\,j\,t-1} \\ &\qquad + 0.16 Z_{i\,j-1\,t-1} + 0.26 Z_{i\,j+1\,t-1} + a_{ijt} \end{aligned} \tag{3.7}$$

with $a_{ijt} \sim \text{IN}(0, 1)$ normal and sample sizes $n_i = 50$, $n_j = 100$, $n_t = 75$.

According to the previous analysis, models (3.7) are unstable because $\sum_k |\phi_k| = 1.05$. Figure 2 (a,b) confirms this behavior by plotting the final realization Z_{ij75}. Note that the

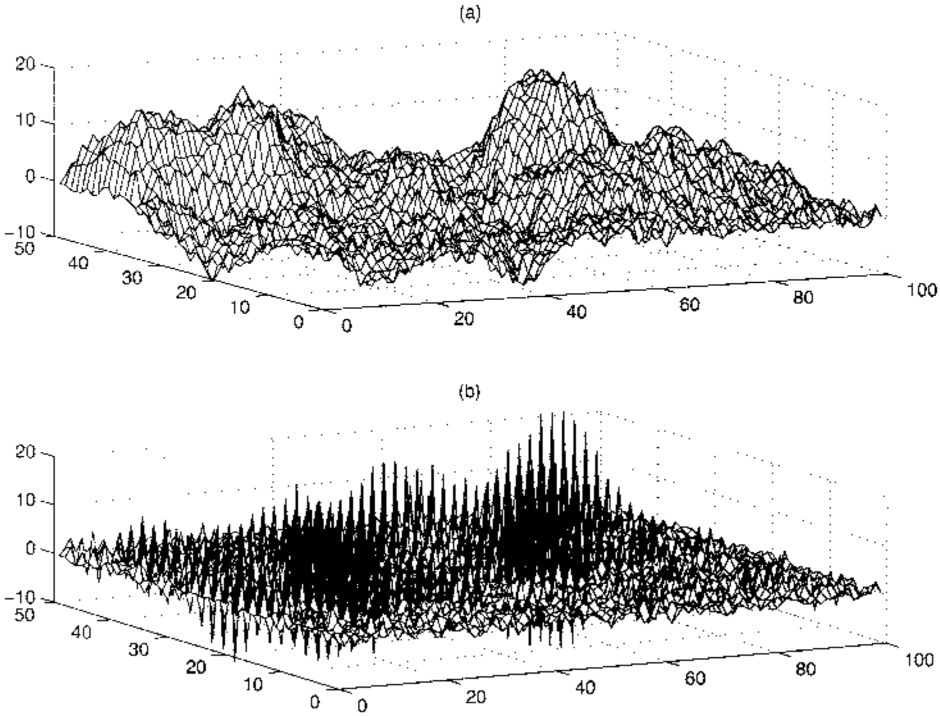

Figure 2. *Realizations of the process (3.7) with parameters: (a) $\phi_1 = +\ 0.21$; (b) $\phi_1 = -0.21$.*

model with $\phi_1 = -0.21$ is unsuitable for representing real data because the generated patterns are extremely erratic.

Estimator (3.5) was applied to 100 independent replications of the process (3.7). We also included the null term $0.0Z_{i+1\,j-1\,t-1}$ in the model, in order to check the usefulness of the estimator for identification purposes. Subsequently, we repeated the experiment on a stable version of (3.7) in which the value of parameters is reduced by 0.06 and with $\phi_1 = -0.15$. Numerical results are reported in Table 2, together with the test for normality by Bera and Jarque (1982), whose critical value is $\chi^2_{0.05}(2) = 6$. We can conclude that, even under model unstability, LS estimates retain optimal statistical properties, with the sole exception for normality when parameters have negative sign (see Table 2(b)).

Serious problems for the LS estimator may arise when the models are noncausal, that is, when instantaneous terms $Z_{i\pm k\,j\pm h\,t}$ are present in the vector of regressors $\mathbf{x}_{ijt}$. As in simultaneous equation systems, identifiability constraints must be introduced, such as the *half-plane* and the *one-quadrant* specification of the matrix $\mathbf{\Phi}_0 = \{\phi_{kh0}\}$. These solutions still preserve causality, because enable *recursive* computation of residuals and forecasts; however, they need suitable initial values on the upper-left border of the data matrix. As a consequence, Niu et al. (2003) showed that instantaneous terms improve fitting, but may significantly decrease the forecasting performance of space-time models.

Table 2. Simulation Results of Estimator (3.5) Applied to the Model (3.7). Rows: (1) mean values, (2) root mean squared errors, (3) tests for normality.

Design	ϕ_0 *(0.0)*	ϕ_1 *(0.21)*	ϕ_2 *(0.11)*	ϕ_3 *(0.31)*	ϕ_4 *(0.16)*	ϕ_5 *(0.26)*
			(a)			
Mean	0.0002	0.2100	0.1100	0.3100	0.1599	0.2599
RMSE	(0.0011)	(0.0013)	(0.0012)	(0.0012)	(0.0013)	(0.0014)
N-test	1.45	1.93	2.12	1.79	3.19	1.65
			(b)			
Design	ϕ_0 *(0.0)*	ϕ_1 *(−0.15)*	ϕ_2 *(0.05)*	ϕ_3 *(0.25)*	ϕ_4 *(0.10)*	ϕ_5 *(0.20)*
Mean	0.0001	−0.1501	0.0499	0.2501	0.0999	0.1999
RMSE	(0.0016)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0016)
N-Test	0.324	7.50	3.04	6.17	0.843	2.73

3.2 THE CASE STUDY

In this subsection we build and compare alternative STAR models for the NDVI data. Rather than selecting the best solution, we are interested in a parsimonious one, on which to apply adaptive methods. Classical methods of identification (e.g. Niu et al. 1995, 2003), first stabilize data with *differencing* $(Z_{ijt} - Z_{ijt-1})$, then analyze space-time autocorrelation functions (e.g., Pfeifer and Deutsch 1980). However, these are difficult to interpret, especially when the models have an irregular (subset) structure; moreover, differencing may be arbitrary or may be performed in one of the spatial dimensions.

Original Series. Given the consistency of the LS estimator when is applied to unstable processes, one may build dynamic models without differencing (e.g., Parzen 1982). The analysis of univariate models is a necessary starting point

$$\begin{aligned} \text{AR}(1) \quad Z_{ijt} &= 0.9987 Z_{ijt-1} + a_{ijt}, \quad \hat{\sigma}_a^2 = 16.79 \\ +\text{drift} \quad Z_{ijt} &= 5.96 + 0.963 Z_{ijt-1} + a_{ijt}, \quad \hat{\sigma}_a^2 = 16.46 \\ \text{AR}(2) \quad Z_{ijt} &= 0.878 Z_{ijt-1} + 0.120 Z_{ijt-2} + a_{ijt}, \quad \hat{\sigma}_a^2 = 16.55. \end{aligned} \tag{3.8}$$

The large number of observations ($N = 58{,}806$) involved in the study, renders model comparison difficult because it *inflates* many of the statistics.

For example, in the model (3.8) the parameter $\hat{\phi}_1 \approx 1$, but it is not a *unit root* because the studentized statistic $t_1 = (\hat{\phi}_1 - 1)/\hat{\sigma}_\phi = -12.7$ is 99% significant in the Dickey and Fuller (1979) test. Second and third model seem to indicate the usefulness of drift and AR(2) components, because F-statistics (on the reduction of the residual sum of squares with respect to the first model) are greater than 900. In both cases, however, the reduction of $\hat{\sigma}_a^2$ with respect to (3.8) is 2% only, and the magnitude of additional parameters is small.

More interesting is the analysis of the effect of spatial components on the relative size of the coefficients. Estimation of the model (3.1) provided $\hat{\sigma}_a^2 = 16.015$, $\hat{\phi}_1 = 0.404$, whereas the remaining coefficients ranged in $[0.05, 0.10]$, in such a way that $\sum_{k=2}^{9} \hat{\phi}_k = 0.595$. Also their standard errors were homogeneous because ranged in $[0.0095, 0.0125]$, so that all estimates were 99% significant. To achieve a parsimonious representation, one may aggregate the neighboring values of Z_{ijt-1} as in Equation (3.3).

Table 3. Statistics of STAR Models Estimated on Original Data. The quantities in parentheses are standard errors and the *F*-statistics are computed sequentially.

Eq.	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\sigma}_a^2$	*BIC*	*F*
(3.8)	0.9987 (0.0001)	.	.	.	16.791	2.821	.
(3.10)	0.384 (0.012)	0.615 (0.012)	.	.	16.067	2.778	2504
(3.11)	0.388 (0.012)	0.335 (0.012)	0.276 (0.014)	.	16.062	2.777	18.3
(3.9)	0.417 (0.013)	0.314 (0.025)	0.158 (0.023)	0.109 (0.016)	16.009	2.774	184.8

An aggregate model having a third order spatial dependence as in Table 1 is

$$\begin{aligned} Z_{ijt} \;=\;& 0.417 Z_{ijt-1} + 0.314\,(Z_{ij-1t-1} + Z_{ij+1t-1} + Z_{i-1jt-1} + Z_{i+1jt-1})/4 \\ &+0.158\,(Z_{i-1j-1t-1} + Z_{i+1j+1t-1} + Z_{i-1j+1t-1} + Z_{i+1j-1t-1})/4 \\ &+0.109\,(Z_{ij-2t-1} + Z_{ij+2t-1} + Z_{i-2jt-1} + Z_{i+2jt-1})/4 \;+\; a_{ijt}, \qquad (3.9) \end{aligned}$$

where $\hat{\sigma}_a^2 = 16.009$. The small size of the last two coefficients in (3.9) suggests simplyfing the model by retaining only the nearest neighbor terms as follows

$$Z_{ijt} \;=\; 0.384 Z_{ijt-1} + 0.615\,(Z_{ij-1t-1} + Z_{ij+1t-1} + Z_{i-1jt-1} + Z_{i+1jt-1})/4 + a_{ijt}, \qquad (3.10)$$

where $\hat{\sigma}_a^2 = 16.067$. With respect to previous estimates, model (3.10) indicates the importance (weight) of the spatial component. Such a component can be further decomposed into its latitudinal and longitudinal parts as follows

$$\begin{aligned} Z_{ijt} \;=\;& 0.388\, Z_{ijt-1} + 0.335\,(Z_{i-1jt-1} + Z_{i+1jt-1})/2 + \\ &+0.276\,(Z_{ij-1t-1} + Z_{ij+1t-1})/2 + a_{ijt}, \qquad (3.11) \end{aligned}$$

where $\hat{\sigma}_a^2 = 16.062$. As should be expected, the north-south coefficient is greater than the east-west one, and both are smaller than the central one.

Previous estimates are summarized in Table 3 and formal tests are provided. One may note that the models (3.9)–(3.11) are nearly equivalent from the fitting viewpoint (namely $\hat{\sigma}_a^2$), although their t and F statistics state the contrary. To avoid these problems one may resort to statistical information criteria, such as the consistent version of that of Akaike (see Tjøstheim 1983, p. 572)

$$\mathrm{BIC}_N(m) = \log(\hat{\sigma}_a^2) + m \log(N)/N\,, \quad m = \dim(\phi).$$

This criterion aims to establish a compromise between model fitting and parametric complexity and should posses a well-defined minimum. However, as $N \to \infty$, BIC has the same problems as t and F statistics, because $\log(N)/N \to 0$. As a consequence, in Table 3 the selected model would be (3.9), or (3.1), the less parsimonious ones.

Detrended Series. Problems of model selection can partly be solved by investigating detrended series. Owing to the behavior of NDVI data in Figure 1, such operation can be

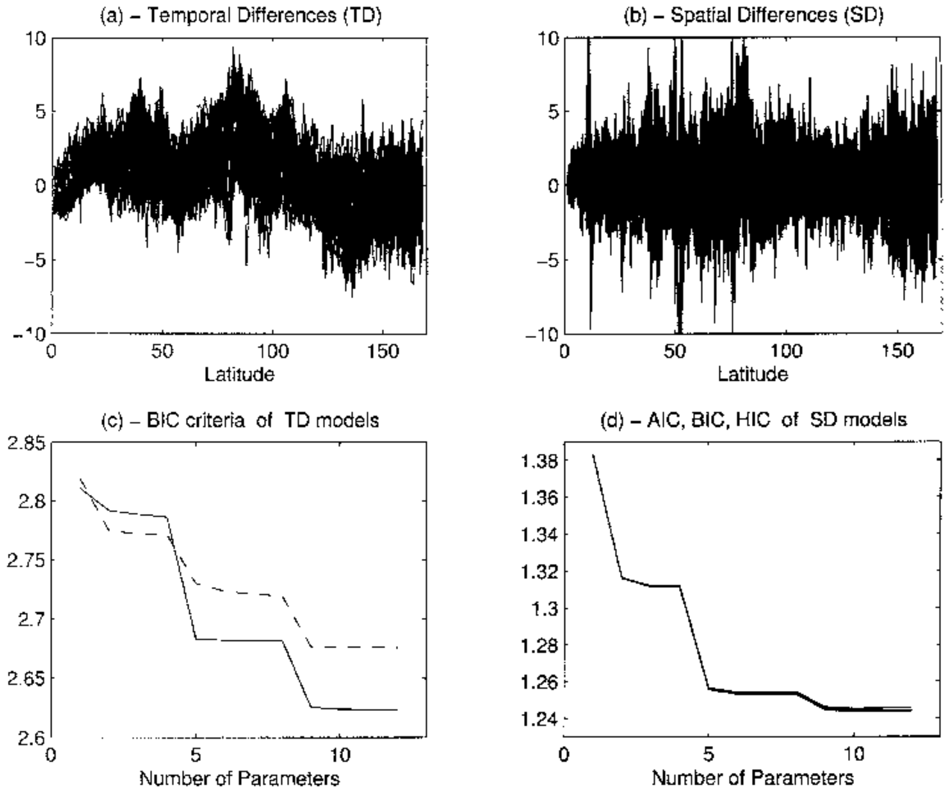

Figure 3. *(a,b) Latitudinal paths of the detrended series for the year t = 1999. (c,d) Information criteria of STAR models fitted to series x, y (solid), and Z (dashed).*

performed in two ways

$$\text{Temporal Differencing (TD)}: \; x_{ijt} \;=\; (Z_{ijt} - Z_{ijt-1})$$
$$\text{Spatial Differencing (SD)}: \; y_{ijt} \;=\; (Z_{ijt} - Z_{i-1jt}).$$

The resulting series for the year 1999 and for the entire region are displayed in Figure 3(a,b); in general, both transformations eliminate the marked north-south trend, although with different characteristics.

Figure 3(c,d) display the value of information criteria of models STAR(q, p) of *type* (3.9), fitted to the detrended and original series x, y, Z. The order of the spatial component $0 \leq q \leq 3$ was increased *conditionally* on that of the temporal one $1 \leq p \leq 3$; so that 12 schemes were analyzed: $(q, p) = (0, 1), (1, 1), (2, 1) \ldots (3, 3)$. For example, the model with orders (3,1) for Z corresponds to the Equation (3.9).

Several comments can be done on the above estimates:

1. As in Table 3, the path of information criteria is nearly proportional to that of the residual variances. As mentioned before, this is a consequence of the large sample size, which hinders finding a well-defined minimum for BIC.

Table 4. Estimates of the Parameters of a STAR(3,3) Model of Type (3.9) With a Constant Term α_0. The quantities t_{kl}^x are t-statistics of parameters ϕ_{kl} estimated on x_{ijt}; (k,l) indicate the spatial and temporal lags, as reported in Table 1.

k, l	$\hat{\phi}_{kl}^Z$	t_{kl}^Z	$\hat{\phi}_{kl}^x$	t_{kl}^x	$\hat{\phi}_{kl}^y$	t_{kl}^y
α_0	4.59	27.4	−0.391	−24.7	0.046	3.51
0, 1	0.305	24.0	−0.569	−51.2	0.299	71.5
1, 1	0.217	8.63	0.106	4.86	0.219	26.7
2, 1	0.212	9.24	0.130	6.46	0.071	8.63
3, 1	0.258	14.3	0.179	10.7	−0.019	−2.31
0, 2	0.191	14.8	−0.290	−23.4	0.199	46.0
1, 2	0.038	1.51	0.022	0.89	0.061	6.98
2, 2	−0.104	−4.48	−0.038	−1.69	−0.025	−2.92
3, 2	−0.165	−8.80	−0.059	−3.22	−0.028	−3.32
0, 3	0.095	7.49	−0.131	−11.8	0.089	20.7
1, 3	0.065	2.64	−0.008	−0.37	0.063	7.39
2, 3	−0.044	−1.94	−0.060	−3.03	0.008	0.93
3, 3	−0.097	−5.27	−0.077	−4.60	−0.021	−2.46
$\hat{\sigma}_a^2$, (BIC)	14.325	(2.665)	13.601	(2.613)	3.465	(1.245)

2. From Figure 3(c), the performance of the models fitted to series Z_{ijt}, x_{ijt} is similar, and both are significantly worse of the models for y_{ijt}. In Figure 3(d) one may note that the behavior of Akaike and Hannan criteria is similar to BIC.

3. For series y_{ijt}, biggest reductions of information criteria occur for regression terms having lags $(k, l) = (1, 1), (0, 2)$; therefore, the implied model would be

$$y_{ijt} = 0.329\, y_{ijt-1} + 0.313\, (y_{ij-1t-1} + y_{ij+1t-1} + y_{i-1jt-1} + y_{i+1jt-1})/4 + 0.238\, y_{ijt-2} + a_{ijt}.$$

Its fitting statistics are good: $\hat{\sigma}_a^2 = 3.52, t_{\hat{\phi}} \geq 50, \text{BIC} = 1.26$, and it is interesting to note that the stability constraint is satisfied, because the polynomial $\phi(z) = 1 - 0.642\, z - 0.238\, z^2$ has roots outside the unit circle.

Finally, we would like to provide detailed results of the estimation of a STAR(3,3) model of type (3.9) with the inclusion of a constant term. Table 4 shows that models fitted on Z, x have a similar performance, but the best absolute fitting is obtained on the series y. On this series, however, the statistical significance of the constant term α_0 is much lower than the other cases. The drastic reduction of the residual variance of the models for y, is a consequence of the fact that spatial differencing is similar to the inclusion of an instantaneous term. Although this usually improves fitting, it may worsen the forecasting ability, see Niu et al. (2003) and Section 5.

4. ADAPTIVE STAR MODELS

In many contexts estimation of dynamic models on real data has shown variability of coefficients (e.g., Grillenzoni 1994, 1998). It is reasonable to expect that STAR models applied to ecological data share this feature. The question is: how to model and estimate parameters that change both over space and time?

There are approaches in image processing and weather analysis where the Kalman filter is applied recursively to data organized in vector form (e.g., Daley 1995). In practice, this means choosing an a priori direction in space (e.g., the lexicographic order), and then reducing the data matrix to a long vector. In space, however, one must assume that each unit depends on the surrounding ones in every direction, and that such dependence follows unknown probabilistic laws. All of these elements lead us to use a nonparametric approach (see Fan and Gijbels 1996).

A simple way to obtain time-varying and space-varying estimates is simplyfing the LS estimator (3.5) by dropping suitable summation terms, such as

$$\begin{aligned} \text{Time-varying} : \hat{\boldsymbol{\phi}}_t &= \left(\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \mathbf{x}_{ijt}\,\mathbf{x}'_{ijt} \right)^{-1} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \mathbf{x}_{ijt}\, Z_{ijt} \\ \text{Space-varying} : \hat{\boldsymbol{\phi}}_{ij} &= \left(\sum_{t=1}^{n_t} \mathbf{x}_{ijt}\,\mathbf{x}'_{ijt} \right)^{-1} \sum_{t=1}^{n_t} \mathbf{x}_{ijt}\, Z_{ijt}. \end{aligned}$$

These *sequential* methods usually provide very erratic estimates which may not be interpretable nor may be useful in forecasting. Smooth alternatives arise by considering *local* versions of (3.5); one of the most elegant solutions is that based on exponential weighting (EW).

4.1 VARIABILITY IN TIME

We start with the case where parameters only change over time. The method of EW regression concerns with the local LS estimator

$$\hat{\boldsymbol{\phi}}_t = \left(\sum_{l=1}^{t} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \lambda^{t-l} \mathbf{x}_{ijl}\,\mathbf{x}'_{ijl} \right)^{-1} \sum_{l=1}^{t} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \lambda^{t-l} \mathbf{x}_{ijl}\, Z_{ijl} \tag{4.1}$$

the coefficient $0 \leq \lambda \leq 1$ is known as "forgetting factor" and gives more weight to the recent observations. In this way, the estimator (4.1) adapts the models to the underlying technological and environmental changes. With respect to other forms of weighting, it enables to easily obtain the recursive expression of (4.1).

Simple algebraic manipulations (as in Ljung and Söderström 1983) and the inclusion of other adaptation coefficients, lead to the recursive LS estimator

$$\begin{aligned} \mathbf{R}_t &= \lambda\,\mathbf{R}_{t-1} + \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \mathbf{x}_{ijt}\,\mathbf{x}'_{ijt}, \quad \mathbf{R}_0 = \rho_0 \mathbf{I} \\ \hat{a}_{ijt} &= Z_{ijt} - \hat{\boldsymbol{\phi}}'_{t-1} \mathbf{x}_{ijt} \\ \hat{\boldsymbol{\phi}}_t &= \hat{\boldsymbol{\phi}}_{t-1} + \mu\,\mathbf{R}_t^{-1} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \mathbf{x}_{ijt} \hat{a}_{ijt}, \quad \hat{\boldsymbol{\phi}}_0 = \boldsymbol{\phi}_0, \end{aligned} \tag{4.2}$$

where $\hat{a}_{ijt}$ are prediction errors, the matrix $\mathbf{R}_t$ is "denominator" of (4.1), μ is a stepsize coefficient and $\boldsymbol{\phi}_0, \rho_0$ are initial values.

Table 5. Estimates (4.3) of the Coefficients of the Filter (4.2) Applied to the Model (3.11)

Coefficient	ϕ_{01}	ϕ_{02}	ϕ_{03}	ρ_0	λ	μ	σ_a^2
Estimate (a)	0.3051	0.349	0.349	818.2	0.117	−0.613	12.07
Estimate (b)	0.408	0.124	0.487	263.2	1.13	1*	17.66

The algorithm (4.2) considerably saves calculations and with $\mu = 1$ is asymptotically equivalent to (4.1). However, the presence of the factor $\mu \neq 1$ in (4.2), together with the initial values $\boldsymbol{\phi}_0, \rho_0 \neq 0$, improve the adaptive capabilities with respect to the estimator (4.1). By this we mean both its ability to track the trajectory of parameters and to reduce the variance of prediction errors.

Optimization. The design of the adaptation coefficients λ, μ and of the initial values $\boldsymbol{\phi}_0, \rho_0$ is fundamental for the performance of (4.2). The heuristic approaches used in the Bayesian and engineering literature (e.g., Ljung and Söderström 1983), may not be useful for forecasting. As in Grillenzoni (1994) one can determine their *optimal* values by minimizing the sum of squared prediction errors:

$$[\,\hat{\lambda},\, \hat{\mu};\, \hat{\rho}_0,\, \hat{\boldsymbol{\phi}}_0'\,] = \text{arg min} \left[\hat{Q}_N = \sum_{t=1}^{n_t} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (Z_{ijt} - \hat{\boldsymbol{\phi}}_{t-1}' \mathbf{x}_{ijt})^2 \right]. \tag{4.3}$$

This approach is consistent because it coincides with the cross-validation method used in nonparametric regression for the design of the bandwidths.

Owing to the possible nonsmoothness of the objective function $\hat{Q}_N$, the solution of the optimization problem (4.3) is carried out by search algorithms. In our experience the more sensitive parameter is μ, whereas ρ_0 does not have an important role and could be fixed a priori. A sensible benchmark for evaluating the performance of the method is represented by the residual variance $\hat{\sigma}_a^2$ of the corresponding constant parameter models. The greater the reduction of this variance, the greater the degree of nonstationarity of the models.

We tentatively applied the method (4.2)–(4.3) to parsimonious models of the NDVI series, such as (3.11). Estimates of the coefficients (4.3) are reported in Table 5, and the corresponding recursive estimates (4.2) are displayed in Figure 4. We may see that the adaptive method reduces the innovation variance by about 20% with respect to the model (3.11). On the other hand, the variance in the constrained ($\mu = 1^*$) estimation is greater than that in (3.11) because of the small values of n_t.

Monthly Data. As a further application, we consider the case of monthly NDVI data Z_{ijs}. A sensible STAR model should contain a seasonal component at lag 12 with a spatial dynamic of order one. Saving parameters as in (3.3) we have

$$\begin{aligned} Z_{ijs} \;=\;& 0.377\,(Z_{ijs-1} + Z_{ij-1s-1} + Z_{ij+1s-1} + Z_{i-1js-1} + Z_{i+1js-1})/5 \\ & +0.622\,(Z_{ijs-12} + Z_{ij-1s-12} + Z_{ij+1s-12} + Z_{i-1js-12} + Z_{i+1js-12})/5 + a_{ijs} \end{aligned} \tag{4.4}$$

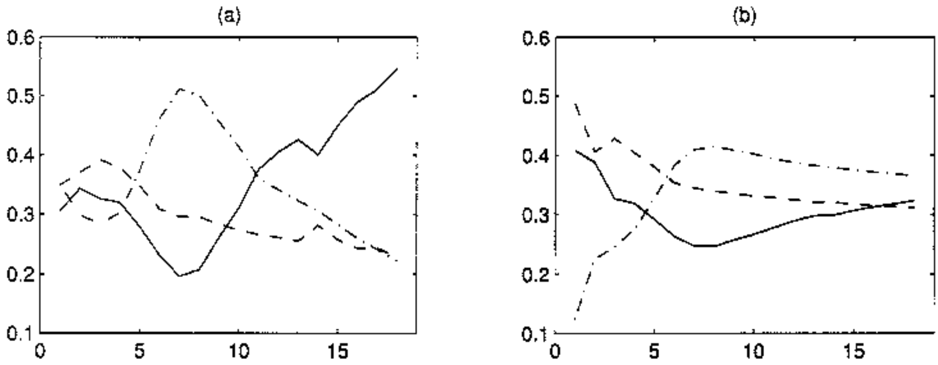

Figure 4. Recursive estimates of the model (3.11) generated with the algorithm (4.2) and the coefficients in Table 5: $\hat{\phi}_{1t}$ (solid), $\hat{\phi}_{2t}$ (dash-dot), $\hat{\phi}_{3t}$ (dashed), t = 1 . . . 18.

with $\hat{\sigma}_a^2 = 84.2$. The contribution of spatial components is significant because the univariate model alone provided a greater residual variance: $\hat{\sigma}_a^2 = 89.6$.

Owing to the huge number of data ($N \approx$ 720,000), application of the adaptive method to (4.4) was problematic. Optimization (4.3) was very slow and therefore we have assigned a priori values to the coefficients of algorithm (4.2), namely, $\phi_{01} = 0.4$, $\phi_{02} = 0.6$, $\rho_0 =$ 1,000, $\lambda = 0.93$, $\mu = 1$. The corresponding recursive estimates $\hat{\phi}_{1s}$, $\hat{\phi}_{2s}$ are displayed in Figure 5(a). As regards their variability, it may also be explained with the change of NOAA satellites, which in fact occurred in February 1985 $(s = 42)$, November 1988 $(s = 87)$ and the last in September 1994 $(s = 157)$.

Figure 5(b) provides the *sequential* parameter estimates, as obtained by separately regressing each data matrix on the previous one, namely,

$$\tilde{\boldsymbol{\phi}}_s = \left(\sum_i \sum_j \mathbf{x}_{ijs} \, \mathbf{x}'_{ijs} \right)^{-1} \sum_i \sum_j \mathbf{x}_{ijs} \, Z_{ijs}$$

as we can see, they are extremely erratic. As a consequence, their predictive ability, evaluated with the innovations $\tilde{a}_{ijs} = Z_{ijs} - \tilde{\boldsymbol{\phi}}'_{s-1}\mathbf{x}_{ijs}$, is bad because their variance $\tilde{\sigma}_a^2$=195.6 is more than twice the residual variance in Equation (4.4).

4.2 VARIABILITY IN SPACE AND TIME

The most general nonstationary situation is where parameters of STAR models change both in space and in time. To deal with this problem, we extend the weighted estimator (4.1) as follows

$$\begin{aligned} \hat{\boldsymbol{\phi}}_{ijt} \quad &= \quad \left(\sum_{l=1}^{t} \lambda_1^{t-l} \sum_{k=1}^{n_i} \sum_{h=1}^{n_j} \lambda_2^{|\,i-k\,|} \, \lambda_3^{|\,j-h\,|} \, \mathbf{x}_{khl} \, \mathbf{x}'_{khl} \right)^{-1} \\ &\times \sum_{l=1}^{t} \lambda_1^{t-l} \sum_{k=1}^{n_i} \sum_{h=1}^{n_j} \lambda_2^{|\,i-k\,|} \, \lambda_3^{|\,j-h\,|} \, \mathbf{x}_{khl} \, Z_{khl} \,, \quad 0 < \lambda_{1,2,3} < 1. \qquad (4.5) \end{aligned}$$

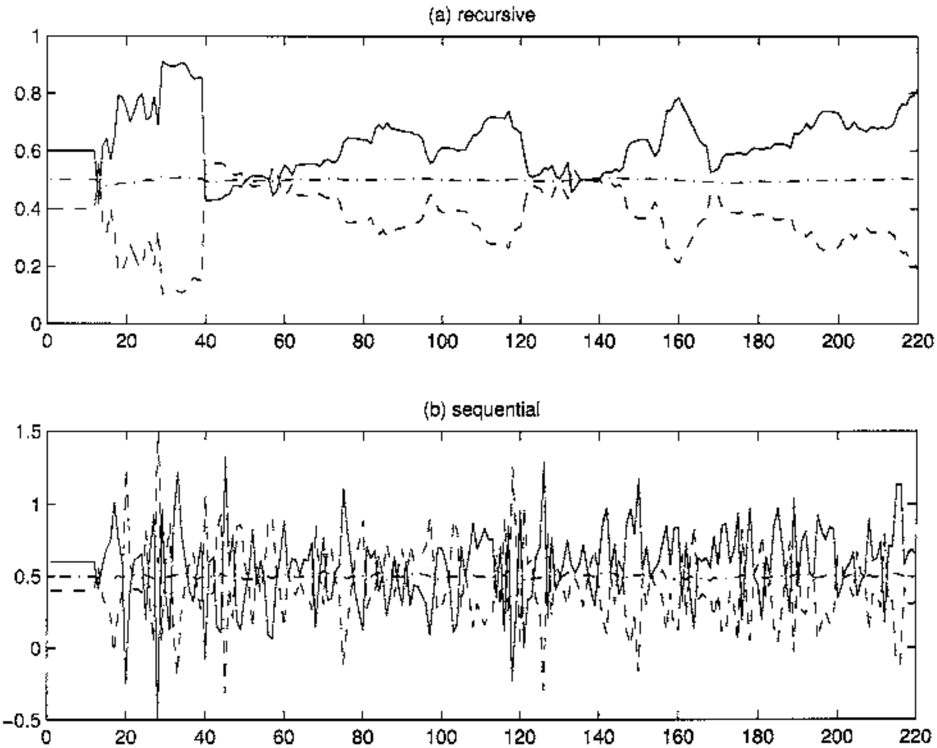

Figure 5. Recursive and sequential estimates of the parameters of model (4.4): $\hat{\phi}_{1s}$ (dashed), $\hat{\phi}_{2s}$ (solid), $(\hat{\phi}_{1s} + \hat{\phi}_{2s})/2$ (dash-dot), $s = 1, 2 \ldots 220$.

With this solution, parameter estimates depend on local data in every direction of space, and not only along the latitudinal or longitudinal directions.

Notice, that following the classical nonparametric regression, the exponential weights $\lambda_2^{|\,i-k\,|}$ could be replaced by the kernel weights $K[(i-k)/h_2]/h_2$, where $0 < h_2 < \infty$ is the so-called bandwidth. However, in the time dimension it is preferable to keep the exponential weighting because it has a one-sided profile and enables to obtain the recursive implementation of estimators.

The recursive (in t) version of (4.5) can be derived as in the case of (4.2)

$$\begin{aligned} \hat{a}_{kht} &= (\, Z_{kht} - \hat{\boldsymbol{\phi}}'_{ijt-1}\mathbf{x}_{kht}\,)\,, \\ \mathbf{R}_{ijt} &= \lambda_1\,\mathbf{R}_{ijt-1} + \sum_{k=1}^{n_i}\sum_{h=1}^{n_j} \lambda_2^{|\,i-k\,|}\,\lambda_3^{|\,j-h\,|}\,\mathbf{x}_{kht}\,\mathbf{x}'_{kht}\,, \quad \mathbf{R}_{ij0} = \rho_0\mathbf{I} \\ \hat{\boldsymbol{\phi}}_{ijt} &= \hat{\boldsymbol{\phi}}_{ijt-1} + \mu\,\mathbf{R}_{ijt}^{-1}\sum_{k=1}^{n_i}\sum_{h=1}^{n_j} \lambda_2^{|\,i-k\,|}\,\lambda_3^{|\,j-h\,|}\,\mathbf{x}_{kht}\,\hat{a}_{kht}\,, \quad \hat{\boldsymbol{\phi}}_{ij0} = \boldsymbol{\phi}_0. \qquad (4.6) \end{aligned}$$

This implementation reduces the amount of computation, and the inclusion of the stepsize μ improves the adaptive capability with respect to (4.5).

The design of the coefficients $\lambda_1,\ \lambda_2,\ \lambda_3,\ \rho_0,\ \phi_0$ can be carried out with the criterion

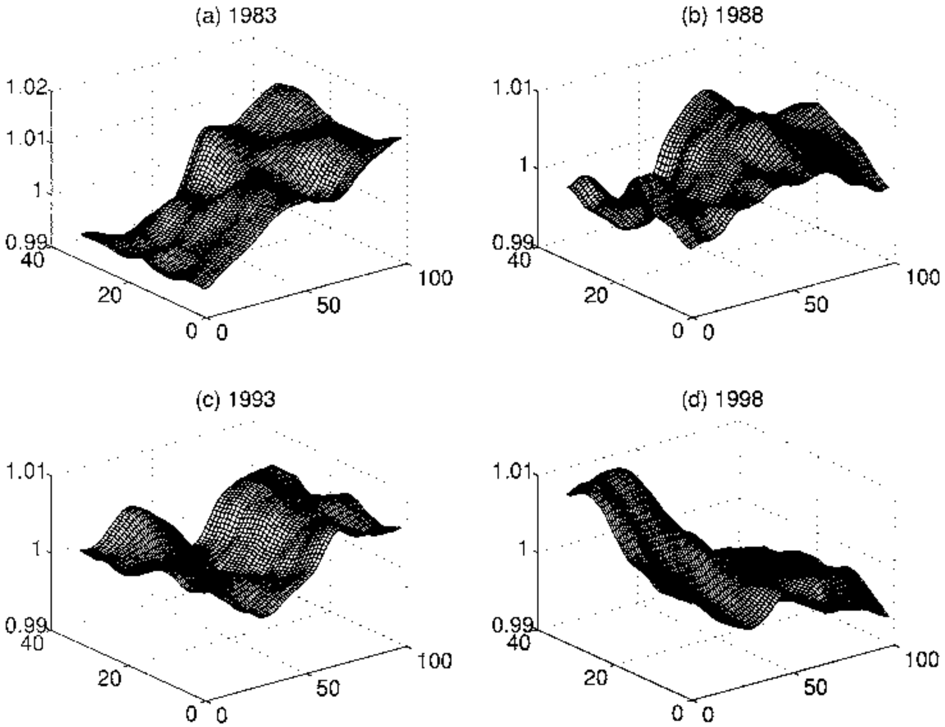

Figure 6. Adaptive estimates of the parameter of model (4.7) obtained with the filter (4.6) and subjective values for the smoothing coefficients. Estimates $\hat{\phi}_{ijt}$ are displayed for the years t = 1983, 1988, 1993, 1998 (x-axis = latitude).

(4.3); however, computations result very demanding because the algorithm (4.6) performs $n_i \times n_j$ times the operations of (4.2). For this reason, we applied the method to the simplest STAR model:

$$Z_{ijt} = 0.9987\,(Z_{ijt-1} + Z_{ij-1t-1} + Z_{ij+1t-1} + Z_{i-1jt-1} + Z_{i+1jt-1})/5 + a_{ijt}, \quad (4.7)$$

where $\hat{\sigma}_a^2 = 16.093$. With respect to (3.10) the parameter ϕ of (4.7) can be interpreted as a *space-time unit-root*, namely a unit root both in space and in time.

Since computation remained demanding, we estimated (4.7) recursively with the algorithm (4.6) implemented with the a priori choices $\phi_0 = 1, \rho_0 = 100, \lambda_1 = 0.5, \lambda_2 = 0.9, \lambda_3 = 0.7, \mu = -0.2$, which yield $\hat{\sigma}_a^2 = 14.86$. This variance is greater than that of previous adaptive estimations, but it is smaller than that of (4.7), with constant parameters. Recursive estimates are displayed in Figure 6, every five years. The actual variability seems moderate because $\hat{\phi}_{ijt}$ lie in the range $[0.99, 1.01]$. However, it should be noted that even small fluctuations of a root on the unit circle may determine complex nonstationary patterns in the process.

Computational problems of algorithm (4.6) are due to the fact that it is recursive only with respect to the time index. Recursivity in space can be developed only in a particular direction, along the latitude or the longitude. In the Appendix a *doubly recursive* estimator is developed and tested on NDVI data.

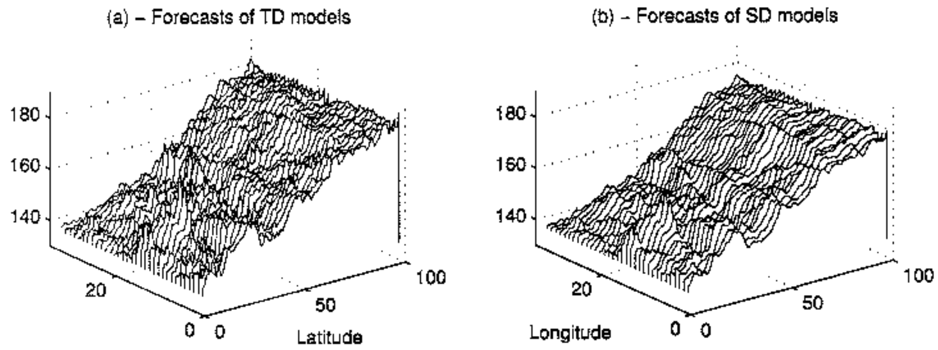

Figure 7. *Out-of-sample forecasts of Z_{ijt} for the year $t = 1999$, obtained with the formulas (5.1) and (5.2).*

5. FORECASTING

As a final exercise we evaluate the forecasting ability of some systems we have presented so far. In particular, we are interested in comparing models of original and detrended series, and models with constant and varying parameters. Returning to Section 2, the one-step-ahead predictors of the detrended models are

$$\text{TD models}: \quad \hat{Z}_{ijt+1} = Z_{ijt} + \hat{x}_{ijt+1}, \quad \hat{x}_{ijt+1} = \boldsymbol{\phi}'\mathbf{x}_{ijt} \tag{5.1}$$

$$\text{SD models}: \quad \hat{Z}_{ijt+1} = \hat{Z}_{i-1jt+1} + \hat{y}_{ijt+1}, \quad \hat{y}_{ijt+1} = \boldsymbol{\phi}'\mathbf{y}_{ijt} \tag{5.2}$$

where the vectors **x, y** contain present and past values of the series x, y.

From (5.2) it is clear that predictions of SD models strongly depend on the initial values $\hat{Z}_{0jt+1}$, $j = 1, 2 \ldots n_j$. These can be obtained with univariate forecasting of the border series, or simply setting $\hat{Z}_{0jt+1} = Z_{0jt}$, which is the random walk prediction. In any event, these values influence all of the generated forecasts, because the predictor can be expressed as $\hat{Z}_{ijt+1} = \hat{Z}_{0jt+1} + \sum_{k=1}^{i} \hat{y}_{kjt+1}$. For this reason one may expect that performance of (5.2) is worse than that of (5.1).

Figure 7 shows the pattern of (5.1) and (5.2) for the year $t + 1 = 1999$. The underlying model for series x, y was a STAR(3,2), and (5.2) was initialized with $\hat{Z}_{0jt+1} = Z_{0jt+1}$, the real value. Despite this choice, predictions in Figure 7(b) are more smooth and lower than those in Figure 7(a). Statistical evaluation of forecasts can be obtained with the mean squared forecast errors (MSFE)

$$\text{MSFE}_{t+1} = \frac{1}{n_i\, n_j} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \left(\hat{Z}_{ijt+1} - Z_{ijt+1} \right)^2. \tag{5.3}$$

We computed (5.3) for models estimated on original and detrended series, with orders $(q, p) = (3, 1)$, $(3, 2)$ and with the inclusion of deterministic components (this solution was suggested by a referee). Algorithm (5.2) was initialized with the real value Z_{0jt+1} and results are reported in Table 6.

Table 6. MSFE Statistics of STAR Models of Order (3,1), (3,2), Fitted to Original and Differenced Series. In the last four columns, a constant term α_0 and a latitudinal trend $(\alpha_0 + \alpha_1\, i)$ were included into the models.

Data	Model	(3,1)	(3,2)	(3.1) + c	(3.2) + c	(3.1) + i	(3.2) + i
Original	$Z_{ijt} = \phi_1 Z_{ij\,t-1} + \dots$	7.44	7.81	6.53	6.86	6.06	6.09
TD Series	$Z_{ijt} = Z_{ij\,t-1} + x_{ijt}$	7.56	7.52	7.43	9.07	7.33	8.97
SD Series	$Z_{ijt} = Z_{i-1\,jt} + y_{ijt}$	8.30	8.21	10.16	12.60	9.75	12.13
Original	$Z_{ijt} = \phi_1 Z_{i-1\,jt} + \dots$	12.66	12.88	11.58	11.12	10.32	10.12

As we see, the first two models (that do not contain instantaneous components), have the best forecasting performance. The other two significantly worsen in passing from the real initial value to the random-walk prediction Z_{0jt}. Performance of the models fitted to original series significantly improves (up to 20%) by including deterministic spatial components, such as the latitudinal trend $(\alpha_0 + \alpha_1\, i)$. This improvement, however, stops by using space-time polynomials or spatial polynomial of higher order. On the other hand, inclusion of deterministic components (as in the simple form of a constant term), has a negative impact on the forecasts of the models of differenced series. This is particularly true for STAR models of higher order, as (3, 2). In summary, the best forecasting performance is provided by the first model (on original series and with latitudinal trend), whereas the worst one by the fourth model (on original series and with instantaneous component).

Adaptive Forecasts. As a final exercise we compare the forecasting capability of constant and varying parameters models. In this context, we considered simpler schemes, such as (3.10) and (4.7). Space-time varying estimates of the parameters of model (3.10) for the year $t = 1998$ are reported in Figure 8. Smoothing coefficients of the algorithm (4.5) were selected a priori as $\lambda_1 = 0.5$, $\lambda_2 = 0.9$, $\lambda_3 = 0.7$. Owing to the decomposition of the unit-root in auto and cross spatial components, they show a greater variability than estimates in Figure 6.

With such estimates the adaptive one-step-ahead predictor $\hat{Z}_{ijt+1} = \hat{\phi}'_{ijt}\mathbf{x}_{ijt}$ is im-

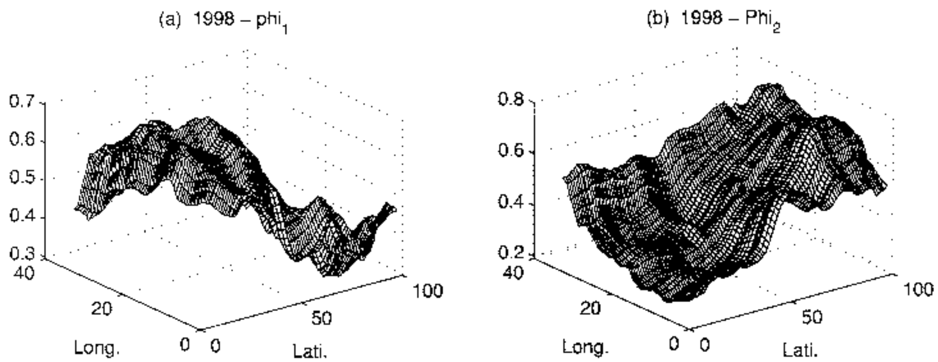

Figure 8. Adaptive estimates of the parameters of model (3.10) for the year $t = 1998$, obtained with the filter (4.5) and subjective choices for λ_k, $k = 1, 2, 3$.

Table 7. MSFE Statistics of Models (3.10) and (4.7), Estimated With Least Squares and Adaptive Methods.

Estimator	*(3.5)*	*(4.2)*	*(4.6)*
Model (4.7)	7.62	6.84	6.66
	↔	(−10%)	(−13%)
Model (3.10)	7.41	6.51	6.27
	↔	(−12%)	(−15%)
	↕	(−5%)	(−6%)

plemented, and its MSFE statistics for the year $t + 1 = 1999$ are reported in Table 7. As one can see, model (3.10) performs better than (4.7); moreover, adaptive estimation has a positive impact on forecasting, even compared with the results in Table 6. It should be noted that results in Table 7 are significantly different in percentage, and much more in statistical terms if one compares them with a standard F-test.

6. CONCLUSIONS

Satellite environmental data are nonstationary both in the spatial and in the temporal dimension. Such data can be represented with space-time unit-root auto-regressive models, namely STAR models which have parameters on the border of the stability region. Tables 6 and 7 have shown that parsimonious models, estimated on original data, may outperform in forecasting more complex models (with instantaneous components) estimated on detrended data.

This performance can be further improved by allowing coefficients to vary both in space and in time. The greatest variability occurs in time, because of the change in satellites and natural evolution. Algorithms for estimating time-varying parameters are simple and can be optimized even in the presence of many observations. They provide the best in-sample and out-of-sample predictive performance.

It would be interesting to compare the forecasting results of Tables 6 and 7 with those of numerical techniques of environmental sciences. However, in that context NDVI data are mainly used for climate classification and trend extraction (see the FAO's site at http://www.fao.gov). Contrary to certain catastrophic perspectives, this study has assessed that vegetation activity measured by satellites is relatively stable, or at least random walk. In any case, we realize that final conclusions can obtained only with at-ground investigations.

APPENDIX: VARIABILITY IN TIME AND LATITUDE

This Appendix considers the case where parameters of STAR models change both in time and with respect to the latitude, namely ϕ_{it}. This situation is of practical interest in the NDVI case study for the presence of a marked north-south trend. To extend the methods

Table A.1. Estimates (4.3) of the Coefficients of the Algorithm (A.5), Applied to the Model (4.7)

Coefficient	ϕ_0	ρ_0	λ_1	λ_2	μ	$\hat{\sigma}_a^2$
Estimate (a)	0.99967	2213.3	0.10520	0.98237	−0.43064	13.61
Estimate (b)	0.99865	987.8	0.3506	1.0976	1*	18.85

(4.1) and (4.5), we consider the *doubly weighted* estimator

$$\hat{\boldsymbol{\phi}}_{it} = \left(\sum_{l=1}^{t} \lambda_1^{t-l} \sum_{k=1}^{i} \lambda_2^{i-k} \sum_{j=1}^{n_j} \mathbf{x}_{kjl}\mathbf{x}_{kjl}' \right)^{-1} \sum_{l=1}^{t} \lambda_1^{t-l} \sum_{k=1}^{i} \lambda_2^{i-k} \sum_{j=1}^{n_j} \mathbf{x}_{kjl} Z_{kjl} \qquad \text{(A.1)}$$

it is easy to show that this minimizes the local objective function

$$P_{it}(\boldsymbol{\phi}) = \sum_{l=1}^{t} \sum_{k=1}^{i} \sum_{j=1}^{n_j} \lambda_1^{t-l}\, \lambda_2^{i-k}\, (Z_{kjl} - \boldsymbol{\phi}'\mathbf{x}_{kjl})^2, \quad 0 \leq \lambda_1, \lambda_2 \leq 1.$$

The recursive version of (A.1), can be obtained by treating its sums sequentially and then by expressing each of them in recursive form. If $\mathbf{y}_{ijt} = \sum_{h=1}^{j} \mathbf{x}_{iht} Z_{iht}$ is the longitudinal sum, then the "numerator" $\mathbf{S}_t$ of (A.1) can be written as

$$\begin{aligned} \mathbf{y}_{ijt} &= \mathbf{y}_{ij-1t} + \mathbf{x}_{ijt} Z_{ijt}, & \mathbf{y}_{i0t} &= \mathbf{0} \\ \mathbf{p}_{it} &= \lambda_2\, \mathbf{p}_{i-1t} + \mathbf{y}_{in_jt}, & \mathbf{p}_{0t} &= \mathbf{0} \\ \mathbf{V}_t &= [\, \mathbf{p}_{1t},\, \mathbf{p}_{2t}\, \ldots\, \mathbf{p}_{n_it}\,], & & \\ \mathbf{S}_t &= \lambda_1\, \mathbf{S}_{t-1} + \mathbf{V}_t, & \mathbf{S}_0 &= \mathbf{O}. \end{aligned} \qquad \text{(A.2)}$$

This formulation implies that the latitude index i runs *before* the time index *t*, and subsequently *all* resulting values are filtered in the temporal dimension.

Similarly, if $\mathbf{X}_{ijt} = \sum_{h=1}^{j} \mathbf{x}_{iht}\mathbf{x}_{iht}'$ then the "denominator" $\mathbf{R}_t$ of (A.1) becomes

$$\begin{aligned} \mathbf{X}_{ijt} &= \mathbf{X}_{ij-1t} + \mathbf{x}_{ijt}\mathbf{x}_{ijt}', & \mathbf{X}_{i0t} &= \mathbf{O} \\ \mathbf{P}_{it} &= \lambda_2\, \mathbf{P}_{i-1t} + \mathbf{X}_{in_jt}, & \mathbf{P}_{0t} &= \mathbf{O} \\ \mathbf{W}_t &= [\, \mathbf{P}_{1t},\, \mathbf{P}_{2t}\, \ldots\, \mathbf{P}_{n_it}\,], & & \\ \mathbf{R}_t &= \lambda_1\, \mathbf{R}_{t-1} + \mathbf{W}_t, & \mathbf{R}_0 &= \mathbf{O}. \end{aligned} \qquad \text{(A.3)}$$

Finally, denoting with $\mathbf{R}_t^{(i)}$, $\mathbf{S}_t^{(i)}$ the ith blocks of the matrices in (A.2) and (A.3), the estimator (A.1) can be rewritten as $\hat{\boldsymbol{\phi}}_{it} = [\,\mathbf{R}_t^{(i)}\,]^{-1}\mathbf{S}_t^{(i)}$.

At this point, by defining the prediction errors $\hat{a}_{ijt} = (Z_{ijt} - \hat{\boldsymbol{\phi}}_{i-1t-1}'\mathbf{x}_{ijt})$, a recursive version of (A.1) can be obtained as in (4.2)

$$\hat{\boldsymbol{\phi}}_{it} = \hat{\boldsymbol{\phi}}_{i-1t-1} + \mu\, [\,\mathbf{R}_t^{(i)}\,]^{-1} \sum_{j=1}^{n_j} \mathbf{x}_{ijt}\, \hat{a}_{ijt}, \quad \hat{\boldsymbol{\phi}}_{00} = \boldsymbol{\phi}_0. \qquad \text{(A.4)}$$

Empirical applications have, however, shown that a better tracking capability is yielded by smoothing the sums $\sum_j \mathbf{x}_{ijt}\, \hat{a}_{ijt}$ with the coefficient λ_2 :

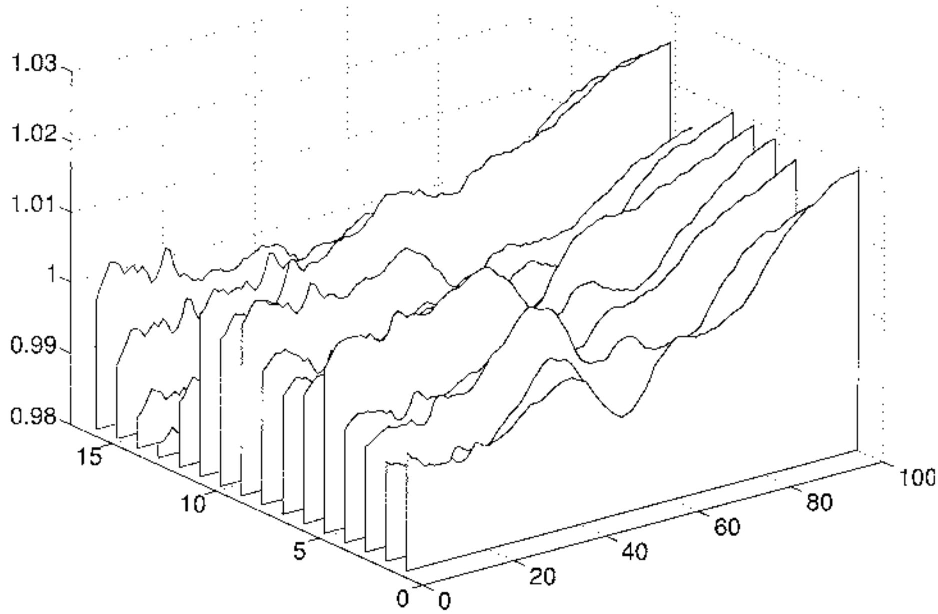

Figure A.1. Recursive estimates of the parameter of model (4.7) obtained with the filter (A.5) and the coefficients in Table 8(a): $\hat{\phi}_{it}$, i = 1, 2 ... 99 , t = 3 , 4 ... 18.

$$\begin{aligned}
\hat{a}_{ijt} &= (Z_{ijt} - \hat{\boldsymbol{\phi}}'_{i-1t-1}\mathbf{x}_{ijt}), \\
\mathbf{q}_{it} &= \lambda_2\,\mathbf{q}_{i-1t} + \sum_{j=1}^{n_j} \mathbf{x}_{ijt}\,\hat{a}_{ijt}, \qquad \mathbf{q}_{0t} = \mathbf{0} \\
\hat{\boldsymbol{\phi}}_{it} &= \hat{\boldsymbol{\phi}}_{i-1t-1} + \mu\,[\,\mathbf{R}_t^{(i)}\,]^{-1}\,\mathbf{q}_{it}, \quad \hat{\boldsymbol{\phi}}_{00} = \boldsymbol{\phi}_0.
\end{aligned} \tag{A.5}$$

Also in this case, optimal design of the coefficients $\lambda_1, \lambda_2, \mu$ can be achieved with the approach (4.3). Computation, however, becomes more demanding.

We applied the adaptive method to the model (4.7); coefficients estimated with (4.3) are displayed in Table A.1. As in Table 5, the algorithm (A.5) with the constraint $\mu = 1$, was unable to reduce the innovation variance of the constant parameter model. This was possible through the unconstrained estimates in the first row of Table A.1. We may note that a fundamental role is played by the time weighting rate λ_1, rather than by the latitudinal one λ_2 which, in fact, approaches 1 in both cases.

Figure A.1 provides the path of recursive estimates obtained with the algorithm (A.5) and the coefficients in the first row of Table A.1. The interesting feature is that they grow from north to south as the spatial trend of the vegetation index. This is a typical feature of other time-varying unit-roots series (see Grillenzoni 1998).

ACKNOWLEDGMENTS

The author is grateful to the editors and the referees for the very constructive review process. In particular, he thanks the reviewer who has suggested the inclusion of a deterministic spatial trend into the models for original series.

[Received January 2002. Revised September 2003.]

REFERENCES

Ali, M. M. (1979), "Analysis of Stationary Spatial-Temporal Processes: Estimation and Prediction," *Biometrika*, 66, 513–518.

Bera, A. K., and Jarque, C. M. (1982), "Model Specification Tests: A Simultaneous Approach," *Journal of Econometrics*, 20, 59–82.

Bhattacharyya, B. B., Khalil, T. M., and Richardson, G. D. (1996), "Gauss-Newton Estimation of Parameters for a Spatial Autoregressive Model," *Statistics and Probability Letters*, 25, 1709–1724.

Bhattacharyya, B. B., Richardson, G. D., and Franklin, L. A. (1997), "Asymptotic Inference for Near Unit Roots in Spatial Autoregression," *The Annals of Statistics*, 28, 173–179.

Daley, R. (1995), "Estimating the Wind Field from Chemical Constituent Observations with the Kalman Filter," *Monthly Weather Review*, 123, 181–198.

De Luna, X., and Genton, M. G. (2002), "Simulation-Based Inference for Simultaneous Processes on Regular Lattices," *Statistics and Computing*, 12, 125–134.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Emery, W. J., Brown, J., and Nowak, Z. P. (1989), "AVHRR Image Navigation: A Review," *Photogrammetric Engineering and Remote Sensing*, 55, 1175–1183.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and its Applications*, New York: Chapman and Hall.

Faivre, R., and Fischer, A. (1996), "Predicting Crop Reflectances Using Satellite Data Observing Mixed Pixels," *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 87–107.

Grillenzoni, C. (1994), "Optimal Recursive Estimators of Dynamic Models," *Journal of the American Statistical Association*, 89, 777–787.

——— (1998), "Forecasting Unstable and Nonstationary Time Series," *International Journal of Forecasting*, 14, 469–482.

Ljung, L., and Söderström, T. (1983), *Theory and Practice of Recursive Identification*, Cambridge, MA: The MIT Press.

Niu, X. F., and Tiao, G. (1995), "Modeling Satellite Ozone Data," *Journal of the American Statistical Association*, 90, 969–983.

Niu, X. F., McKeague, I. W., and Elsner, J. B. (2003), "Seasonal Space-Time Models for Climate Systems," *Statistical Inference for Stochastic Processes*, 6, 111–133.

Parzen, E. (1982), "ARARMA Models for Time Series Analysis and Forecasting," *Journal of Forecasting*, 1, 67–82.

Pfeifer, P. E., and Deutsch, S. J. (1980), "Identification and Interpretation of First Order Space-Time ARMA Models," *Technometrics*, 22, 397–408.

Tjøstheim, D. (1978), "Statistical Spatial Series Modeling," *Advances in Applied Probability*, 10, 130–154.

——— (1983), "Statistical Spatial Series Modeling II: Some Further Results on Unilateral Lattice Processes," *Advances in Applied Probability*, 15, 562–584.

Tucker, C. J., Dregne, H. E., and Newcomb, W. W. (1991), "Expansion and Contraction of the Saharan Desert from 1980 to 1990," *Science*, 253, 299–301.