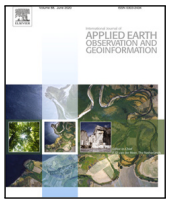
International Journal of Applied Earth Observation and Geoinformation 138 (2025) 104451



Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag





Statistical models for urban growth forecasting: With application to the Baltimore–Washington area

Carlo Grillenzoni *,[1]

Universitá IUAV di Venezia, Venice, Italy

ARTICLE INFO

Dataset link: https://svs.gsfc.nasa.gov/99, https://figshare.com/s/0f849e8d2274da5fea9e

Keywords:
Adaptive estimators
Space–time autoregression
Space–time forecasting
Unilateral contiguity
Video modeling

ABSTRACT

Monitoring and governing the development of cities are the major concerns of urban planners, since involve physical and social aspects, such as land use and population trends. Models for spatial growth have been developed both from the mathematical and empirical viewpoints, with the aim of forecasting and decision-making. Statistical models require regular space–time datasets that are provided by recent remote-sensing and geographic information systems (GIS). In this paper, we consider space–time autoregressive (STAR) models that can be applied to the timelapse video of land transformations available on Internet. The corresponding datasets are in the form of big 3D arrays and require fast algorithms of parameter estimation and forecasting. An extended application to a hybrid timelapse video over 200 years of urban growth of the Baltimore–Washington area is carried out. The video is built by combining remote sensing imagery, census data, historical cartography and data interpolation, and can be fitted and forecasted by adaptive STAR models, with robust and varying parameters.

1. Introduction

In past centuries, industrial and agricultural revolutions have pushed large masses of rural population toward urban centers. This process occurred in a complex way, with different speeds depending on the geographical areas, even within the same nation, and continues nowadays. National capital cities have reached enormous dimensions, followed by regional and provincial capitals, by absorbing the population of peripheral villages. The causal mechanism is driven by demography, which in turn depends on the economic dynamics; but cities themselves are part of the economic process, in terms of consumptions and investments. Subsequently, the large urban centers went into crisis, with deindustrialization and declining birthrates, partly compensated by international migrations. Problems of land use, degradation and the need for regeneration of abandoned areas have emerged, requiring a revision of theoretical models of constant and uniform growth.

Mathematical models for representing the physical growth of cities have followed three main approaches: gravity/transportation, cellular automata and agent-based models (Batty, 2013). The gravitational approach is inspired by the Newton universal law; it assumes that the mobility flows between two places, is proportional to their population and inversely proportional to their geographical distance. This approach is used for studying traffic patterns, population migrations (Capoani, 2023) and consequent urban scaling and growth. In particular, it enables to understand densification processes and the geometry of cities (Ribeiro and Rybski, 2023), and the graphs between the nodes of urban agglomerations (Sun et al., 2019).

Cellular automata are a class of discrete models that generate the state of a process on a regular lattice, by following interaction rules between neighbor units; their origin is connected to fractal theory in planar space. Initial states and interaction weights are established on the basis of specific laws of the investigated phenomenon (Xu et al., 2023); hence, they could also be provided by gravity models. Extensions can be developed in multidimensional (3D) spaces and irregular grids, but their application is preferable to small areas, in order to avoid model instability and divergence (Deep and Saklani, 2014; Falah et al., 2020).

Finally, in the agent-based approach, the system components are actors that make decisions and interact with others and the environment; it traces back to the Von Neumann machine, a device capable of auto-reproduction. The initial state of agents and environmental characteristics determine the interactions between neighbors; these, in turn, determine the spatial constraints (e.g. roads and infrastructures) and the distribution of resources (e.g. food and energy). On a regular spatial grid, the approach is analogous to cellular automata, but it can

* Correspondence to: St Croce, n. 1957, 30135 Venezia, Italy
E-mail address: carlog@iuav.it.
[1] Scopus: 6602822333

https://doi.org/10.1016/j.jag.2025.104451
Received 25 May 2024; Received in revised form 11 February 2025; Accepted 25 February 2025
Available online 10 March 2025
1569-8432/© 2025 The Author. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

also be applied to sparse points (such as buildings and firms), and can be reinforced by the game theory (e.g. Kaviari et al., 2019).

Mixed models, that integrate the various techniques, have been provided, together with the application of recent numerical techniques, such as neural networks and fuzzy logic (see Shafizadeh-Moghadam, 2019). However, all of these approaches involve strong assumptions about the mathematical models, the causal variables and the links between the spatial units (contiguity and response). Statistical models do not involve great a-priori assumptions but need sampling data; fortunately, these are nowadays available from remote sensing technologies, with regular temporal cadence and wide spatial coverage. The resulting datasets are in the form of 4D arrays, indexed by longitude, latitude, time and spectral band and can be visualized in the form of timelapse videos (Google Earth, 2023). Modeling is a challenging issue because of the big data involved.

In this paper, we follow the statistical approach of space–time autoregressive (STAR) models, that put each pixel in linear relationships with the contiguous ones (previous and surrounding). The approach is similar to cellular automata, but the rules of interaction are estimated on data; given the massive arrays involved, efficient estimation algorithms are required. The basic framework only needs univariate data (e.g. gray color) and does not require covariates; its fundamental goal is short-term forecasting for surveillance and decision-making. The accuracy of forecasts depends on the speed of evolution of the phenomenon, the absence of sudden changes (wars, floods, etc.) and the frequency and resolution of data.

In general, other statistical approaches to the study of urban growth manage separately the temporal and spatial features of the data. Static spatial AR models with exogenous variables (namely SARx) have been applied to medium-sized cities (e.g. Jaoude et al., 2022); however, they are unable to forecast their evolution, unless they are combined with external Markov chains, as in Qiu et al. (2022). On the other hand, pure time series models (ARx), which put the building construction activity in relation to explanatory variables (such as economic activity, birth and migration rates, etc.), do not represent the spatial dimension of the growth, see Dentinho and Reid (2021). The underlying spatial diffusion can be estimated by interpolating the historical maps at city level (Acevedo and Masuoka, 1997), or by statistical smoothing the building data at regional level.

On the contrary, STAR models represent in detail the urban areas and their forecasts may provide warning signals about the urban sprawl, soil occupation and other environmental indicators. Their forecast functions may also be employed to restore parts of images that are blurred by clouds (to be used in cartography editing), and classification filters may extract the built-up areas and detect the presence of illegal constructions. Finally, when exogenous variables are available (e.g. on population mobility, business establishments, climatic conditions, etc.), then STARx models may define planning strategies on specific topics and areas. The major problem is the "rasterization" of data that are not detected by remote sensing; however, gridding microdata of buildings, firms and households has growing treatment in geomatics and GIS sciences.

Unlike the STAR models for *point* data (e.g. Mukhaiyar et al., 2022; which can be cast in vector AR form), the literature on raster STAR systems is scarce and concerned with environmental data, e.g. Niu et al. (2003) and Grillenzoni (2004). Recent methodological advances regard the application of robust estimation (Bayer et al., 2021), logistic regression (Zhang et al., 2022) and integration with classification techniques (Guerra et al., 2023). These advances can be applied to urban growth imagery detected, in the visible range, by Landsat and Copernicus probes. In this paper, we focus on the timelapse video of Foresman and Masuoka (1996), about the annual growth of the Baltimore–Washington area from 1792 to 1992. It is realized with mixed historical maps, census information, satellite images and stochastic interpolation (Foresman et al., 1997), using image processing and GIS software. However, it can be transformed into a sequence of realistic (non-redundant) remote-sensing imagery as the recent videos of Google Earth (2023).

We deal with various model specifications and we compare their performance in the out-of-sample forecasting (i.e. leaving the ending frames out of estimation). Given its flexibility and computational simplicity, we rely on the least squares (LS) algorithm with weighted observations. On the other hand, the consistency of LS (in a statistical sense) requires STAR models with a *unilateral* structure of the simultaneous dependence matrix; this solution is also suitable for the chain rule of forecasting. – The paper is organized as follows: Section 2 presents parsimonious STAR models, their LS estimators and performs simulation experiments. Section 3 begins the application to the video data of Chesapeake and Amazon areas and evaluates the forecasting performance. Section 4 discusses adaptive and robust algorithms and applies them to the data of Section 3.

2. Space–time autoregressive models

Regular lattice data are mostly present in remote sensing and digital images. These data are in the form of multi-dimensional arrays of the type $\mathbf{Y} = \{Y_{ijlt}\}$, where $i = 1,2\dots n$, $j = 1,2\dots m$ are the position indices of the pixels (which may be transformed into latitude and longitude), $l = 1,2\dots k$ are the spectral bands ($k = 3$ for RGB colors) and $t = 1,2\dots T$ is the time index (daily, monthly, annual). The main dimensional reduction consists of transforming the spectral bands into a single indicator, such as the grayscale in the visible band and NDVI (normalized difference vegetation index) in near-infrared. The resulting values of Y_{ijt} are usually autocorrelated (ACR, in space and time) and one of the main goals is to filter the array $\mathbf{Y}$ with its values, to obtain interpolations and forecasts.

2.1. Model representations

To represent the space–time ACR, the STAR modeling puts each cell in relation with the contiguous ones, such as $Y_{ijt} = g\left(Y_{i\pm h,j\pm k,t-l}\right) + e_{ijt}$, where $h,k,l = 1,2\dots p$ are spatial and temporal lags and e_{ijt} is an unpredictable sequence. Under linearity of $g(\cdot)$, the model representation can be denoted as STAR(p)

$$Y_{ijt} = \alpha_0 + \sum_{h=0}^{p}\sum_{k=0}^{p}\theta_{hk}\,Y_{i-h,j-k,t} + \sum_{l=1}^{p}\sum_{h=-p}^{p}\sum_{k=-p}^{p}\phi_{hkl}\,Y_{i-h,j-k,t-l} + e_{ijt}, \qquad (1)$$

where $\theta_{00} = 0$ and $e_{ijt} \sim \mathrm{IN}(0,\sigma_e^2)$ is an independent and normal (IN) sequence. The first part of the model (1), with parameters θ_{hk}, has a peculiar nature; it deals with the *simultaneous* relationships between the cells, and it is well studied in the (static) spatial AR literature (e.g. Bustos et al., 2009; Mojiri et al., 2018). For computational reasons, we assume that it has a one-quadrant (triangular) structure as in Eq. (1); in this case, data processing may proceed from the upper-left corner of each frame $\boldsymbol{Y}_t$, to the lower-right one, in a sequential way.

In general, the model (1) is over-parametrized, since $p = 1, 2$ yields $M_p = 34, 59$ coefficients respectively; to achieve parametric parsimony one may aggregate the pixels i, j as in the Chess schemes of Fig. 1: 1/4 *queen* (W), *rook* (X), *bishop* (Z). Hence, by defining the vector index $\boldsymbol{s} = [i, j]$, one has the local averages

$$W_{\boldsymbol{s}-1,t} = \left(Y_{i-1,j,t} + Y_{i,j-1,t} + Y_{i-1,j-1,t}\right)/3, \qquad (2a)$$

$$X_{\boldsymbol{s}-1,t-1} = \left(Y_{i-1,j,t-1} + Y_{i+1,j,t-1} + Y_{i,j-1,t-1} + Y_{i,j-1,t-1}\right)/4, \qquad (2b)$$

$$Z_{\boldsymbol{s}-1,t-1} = \left(Y_{i-1,j-1,t-1} + Y_{i+1,j-1,t-1} + Y_{i-1,j+1,t-1} + Y_{i+1,j+1,t-1}\right)/4, \qquad (2c)$$

(where $\boldsymbol{s}-1$ means $i-1$ and/or $j-1$), and the second-order terms

$$W_{\boldsymbol{s}-2,t} = \left(Y_{i,j-2,t} + Y_{i-1,j-2,t} + Y_{i-2,j-2,t} + Y_{i-2,j-1,t} + Y_{i-2,j,t}\right)/5, \qquad (3a)$$

$$X_{\boldsymbol{s}-2,t-1} = \left(Y_{i,j-2,t-1} + Y_{i,j+2,t-1} + Y_{i-2,j,t-1} + Y_{i+2,j,t-1}\right)/4, \quad \dots \qquad (3b)$$

$$Z_{\boldsymbol{s}-1,t-2} = \left(Y_{i-1,j-1,t-2} + Y_{i+1,j-1,t-2} + Y_{i-1,j+1,t-2} + Y_{i+1,j+1,t-2}\right)/4. \qquad (3c)$$

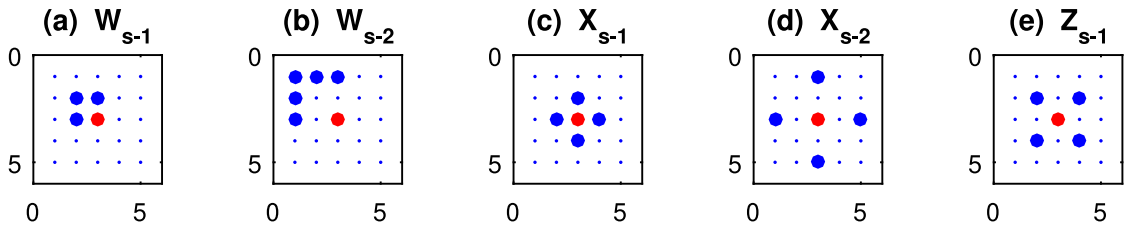


Fig. 1. Examples of Chess neighboring schemes: (a, b) W: 1/4 *Queen* (triangular), of orders 1,2; (c, d) X: *Rook* (cross) of orders 1,2; (e) Z: *Bishop* (diagonal) of order 1.

Thus, the model (1) with order $p = 2$ can parsimoniously be rewritten as

$$\begin{aligned} Y_{s,t} &= \alpha_0 + \phi_1 Y_{s,t-1} + \phi_2 Y_{s,t-1} + \theta_1 W_{s-1,t} + \theta_2 W_{s-2,t} \\ &+ \beta_{11} X_{s-1,t-1} + \beta_{12} X_{s-1,t-2} + \beta_{21} X_{s-2,t-1} + \beta_{22} X_{s-2,t-2} \\ &+ \gamma_1 Z_{s-1,t-1} + \gamma_2 Z_{s-1,t-2} + e_{s,t}, \end{aligned} \tag{4}$$

this contains just 11 parameters, which are reduced to 5 when $p = 1$. In the following, we will compare the fitting and forecasting performance of the model (4) with the presence or absence of the simultaneous terms W.

In time series the ordering of data is unidirectional as follows the past-present flow of events. In pure spatial data Y_{ij}, the unidirectional ordering may be achieved by following the *lexicographic* way of reading/writing a text; i.e. by processing the data matrix $\boldsymbol{Y}$ from the upper-left corner. Spatial schemes which satisfy this feature are called *unilateral*; examples are the row-wise $Y_{i,j-k}$, $k > 0$, the half-plane $Y_{i-h,j\pm k}$, $h > 0$ and the one-quadrant (or triangular) $Y_{i-h,j-k}$, $(h + k) > 0$ (see Bustos et al., 2009; Mojiri et al., 2018). The triangular solution in Fig. 1a, b is suitable for dynamic STAR models, as it allows the sequential computation of the terms $W_{s-k,t}$ from the upper-left corner; this is a necessary condition for the unbiased estimation of parameters and forecasts (see next Table 1).

Stability conditions are useful features for dynamic systems as allow properties of stationarity and convergence. For STAR models, parametric constraints have been established only for simple schemes; e.g. the model (1) with $\theta_{hk} = 0$ (i.e. absence of contemporaneous components) is stable if the composite polynomial

$$\Phi(z) = \Big(1 - \sum_{l=1}^{p} \Phi_l z^l\Big), \qquad \Phi_l = \sum_{h=-p}^{p} \sum_{k=-p}^{p} \phi_{hkl}, \tag{5}$$

has roots outside the unit circle $|z| = 1$, see Ali (1979, p. 514). This condition is similar to that of standard AR(p) models and follows from the fact that, under spatial stationarity (isotropy), the individual processes $Y_{ij,t}$ are similar. But it becomes complicated and demanding in the presence of the instantaneous components $\theta_{hk} \neq 0$, even for $p \leq 2$ (see Mojiri et al., 2018). However, many territorial phenomena are non-stationary and divergent, and this does not hinder the statistical properties of estimates and predictors. On the contrary, they become *super-consistent* (converge at a faster rate) when the roots of the polynomial (5) are inside the unit circle $|z| = 1$, i.e. when the signal-to-noise ratio σ_Y^2/σ_e^2 increases (see Bhattacharyya et al., 1997; Grillenzoni, 2004).

Forecasting. The main goal of STAR models is forecasting future values $Y_{ij,T+\ell}$, $\ell = 1 \dots L$, based on the available data. To derive their prediction function, we rewrite the model (4) in regression form as

$$Y_{s,t} = \boldsymbol{\delta}' \boldsymbol{\xi}_{s,t} + e_{s,t}, \tag{6a}$$

$$\boldsymbol{\delta}' = [\alpha_0, \phi_1 \dots \beta_1 \dots \gamma_2], \tag{6b}$$

$$\boldsymbol{\xi}'_{s,t} = [1, Y_{s,t-1} \dots X_{s-1,t-1} \dots Z_{s-1,t-2}], \tag{6c}$$

where the regressors W, X, Z are generated from the data $\{Y_{ijt}\}$ with the formulas (2)–(3). Thus, as in standard time series models, the predictor is obtained by updating the vector $\boldsymbol{\xi}_{ij,t}$ with the past forecasts as

$$\hat{Y}_{ij,T+\ell} = \mathrm{E}\big(Y_{ij,T+\ell} \big| \boldsymbol{Y}_{T-k}, k \geq 0\big) = \boldsymbol{\delta}' \hat{\boldsymbol{\xi}}_{ij,T+\ell-1}, \tag{7a}$$

$$\hat{\boldsymbol{\xi}}_{ij,T+\ell-1} = \big[1, \hat{Y}_{ij,T+\ell-1} \dots \hat{Z}_{ij,T+\ell-2}\big]', \quad \ell = 1 \dots L, \tag{7b}$$

where $\hat{Z}_{ij,T+\ell-2} = \big(\hat{Y}_{i-1,j-1,T+\ell-2} + \hat{Y}_{i+1,j+1,T+\ell-2} + \hat{Y}_{i-1,j+1,T+\ell-2} + \hat{Y}_{i+1,j-1,T+\ell-2}\big)/4$ are the bishop averages for $\ell > 2$, and so on for the other components.

A significant computational issue in the formula (7) arises from the contemporaneous terms $\widehat{W}_{ij,T+\ell}$, which depend on the forecasts $\hat{Y}_{i-h,j-k,T+\ell}$. Although these elements satisfy the unilateral constraint (and can be computed recursively from the upper-left corner), they require the border values $\hat{Y}_{ij,T+\ell}$, $i, j = 1 \dots p$ to start. These values can be obtained as the forecasts of simple AR models applied to the edge pixels, or by simply setting $\hat{Y}_{ij,T+\ell} = Y_{ij,T}$, as in random-walk models. By the concatenation of forecasts in Eq. (7), the edge values of W will influence the entire frames $\hat{\boldsymbol{Y}}_{T+\ell}$, and so could worsen the forecasts only based on the lagged components Y, X, Z.

The statistical distribution of the forecasts (7) is difficult to derive; approximate confidence intervals for each $\hat{Y}_{ij,T+\ell}$ can be constructed, under the condition of Normality, with the sigma rule $\pm 2\,\sigma_e$. Alternatively, one may follow the bootstrap technique, by resampling the estimated residuals and generating the empirical forecasts $\tilde{Y}^*_{ij,T+\ell} = \hat{Y}_{ij,T+\ell} + \hat{e}^*_{ijT}$. In both cases, one obtains the confidence frames $\tilde{\boldsymbol{Y}}^*_{T+\ell}$, which can be used as measures of uncertainty for (7).

Although statistically sound, the prediction ability of STAR models must be empirically checked with the out-of-sample forecasting, i.e. on L frames $\boldsymbol{Y}_{T-k}$ excluded from parameter estimation. The typical statistic used for evaluation is the mean squared forecast errors (MSFE) and its normalized R^2 index

$$\mathrm{MSFE}_{T'+\ell} = \frac{1}{(n-2p)(m-2p)} \sum_{i=p+1}^{n-p} \sum_{j=p+1}^{m-p} \Big(Y_{ij,T'+\ell} - \hat{Y}_{ij,T'+\ell}\Big)^2, \tag{8a}$$

$$R^2_\ell = 1 - \mathrm{MSFE}_{T'+\ell} \Big/ \hat{\sigma}^2_{\boldsymbol{Y}_{T'+\ell}}, \qquad \ell = 1 \dots L, \tag{8b}$$

where $T' = T - L$. The statistics (8) allow to compare all subset models of Eq. (4), with $p = 1, 2$ and the presence/absence of the components W.

2.2. Parameter estimation

A powerful parameter estimator for STAR models is the maximum likelihood (ML). It requires vectorization of the data matrix and the application of iterative algorithms (see Niu et al., 2003); in general, it can be applied to array of small dimensions n, m. Instead, the least squares (LS) method can be directly applied to the regression representation (6), to provide the compact estimator

$$\hat{\boldsymbol{\delta}}_N = \Big(\sum_{t=p+1}^{T} \sum_{i=p+1}^{n-p} \sum_{j=p+1}^{m-p} \boldsymbol{\xi}_{ijt} \boldsymbol{\xi}'_{ijt}\Big)^{-1} \sum_{t=p+1}^{T} \sum_{i=p+1}^{n-p} \sum_{j=p+1}^{m-p} \boldsymbol{\xi}_{ijt} Y_{ijt}, \tag{9a}$$

$$= \boldsymbol{\delta} + \Big(\sum_{t=p+1}^{T} \sum_{i=p+1}^{n-p} \sum_{j=p+1}^{m-p} \boldsymbol{\xi}_{ijt} \boldsymbol{\xi}'_{ijt}\Big)^{-1} \sum_{t=p+1}^{T} \sum_{i=p+1}^{n-p} \sum_{j=p+1}^{m-p} \boldsymbol{\xi}_{ijt} e_{ijt}, \tag{9b}$$

Table 1
Performance of the LS estimator (9) applied to the model (4) with order $p = 1$, with triangular and rook contemporaneous component W, and stable (0.5) and unstable (1.0) coefficients; Mean, RMSE, P-value are over 500 replications.

Param.	Value	W-triang.			W-rook		
		Mean	RMSE	N-test	Mean	RMSE	N-test
α_0	0.5	0.504	0.042	>0.5	0.190	0.312	0.110
ϕ_1	−0.5	−0.497	0.025	0.411	−0.416	0.089	0.287
θ_1	0.5	0.493	0.035	0.343	0.775	0.280	>0.5
β_1	−0.5	−0.502	0.048	0.153	−0.310	0.200	0.352
γ_1	0.5	0.498	0.044	>0.5	0.594	0.104	>0.5
Abs. Ave.	0.5	**0.499**	**0.039**	0.381	0.457	0.197	0.350
α_0	1	1.001	0.030	>0.5	−0.950	1.969	0.002
ϕ_1	−1	−1.000	0.000	0.001	−1.831	0.847	0.012
θ_1	1	1.000	0.000	0.001	1.408	0.424	0.001
β_1	−1	−1.000	0.000	0.001	2.478	3.502	0.073
γ_1	1	1.000	0.000	0.001	0.120	0.899	0.005
Abs. Ave.	1	**1.000**	**0.006**	0.101	1.357	1.528	0.019

where (9b) is obtained from (6a) and $N = (n-2p)(m-2p)(T-p)$ is the effective sample size. Notice that Eq. (9a) only involves the inversion of a matrix of moderate size M_p, regardless of the dimension $n \times m$ of the area.

The expression (9b) shows the consistency of the LS estimator when the contemporaneous terms W_{ijt} have unidirectional structure; in fact, this allows the sequential calculation of residuals and their independence of regressors: $\mathrm{E}(\boldsymbol{\xi}_{ijt} e_{ijt}) = \mathbf{0}$. Under this constraint, and the conditions of stationarity and isotropy, one can show the classical convergence property (see Kilian and Lütkepohl, 2017 p.31)

$$\sqrt{N}(\hat{\boldsymbol{\delta}}_N - \boldsymbol{\delta}) \to \mathrm{N}\Big[\mathbf{0}, \mathrm{E}(\boldsymbol{\xi}_{ijt}\boldsymbol{\xi}'_{ijt})^{-1}\sigma_e^2\Big], \quad \text{as} \quad N \to \infty. \tag{10}$$

It follows that the dispersion matrix of the estimator (9) is given by

$$\hat{\boldsymbol{\Sigma}}_{\hat{\delta}} = \boldsymbol{R}_N^{-1}\,\hat{\sigma}_e^2, \qquad \boldsymbol{R}_N = \sum_{t=p+1}^{T}\sum_{i=p+1}^{n-p}\sum_{j=p+1}^{m-p}\boldsymbol{\xi}_{ijt}\boldsymbol{\xi}'_{ijt}, \tag{11a}$$

$$\hat{\sigma}_e^2 = (N - M_p)^{-1}\sum_{t=p+1}^{T}\sum_{i=p+1}^{n-p}\sum_{j=p+1}^{m-p}\big(Y_{ijt} - \hat{\boldsymbol{\delta}}'_N\boldsymbol{\xi}_{ijt}\big)^2, \tag{11b}$$

where M_p is the length of $\boldsymbol{\delta}$, for example $M_p = 5, 11$ for $p = 1, 2$.

When the value of N is large, the Student-type statistics $\hat{\boldsymbol{\delta}}_N/\mathrm{diag}\big(\hat{\boldsymbol{\Sigma}}_{\hat{\delta}}\big)^{1/2}$ are inflated and all coefficients turn out significant. Parsimonious model selection is difficult, even with information criteria (IC) which aim to balance data fitting and model parsimony. The Bayesian criterion is given by

$$\mathrm{BIC}_N(p) = \log(\hat{\sigma}_e^2) + M_p\,\log(N)/N, \tag{12}$$

for N large, the term $\log(N)/N$ is negligible and so is the role of M_p; it follows that an objective model selection must be based on the MAFE statistics (8).

2.3. Simulation experiments

To check the statistical properties of the LS estimator (9), we perform simulation experiments on the model (4) with stationary/non-stationary parameters and unilateral/multilateral contemporaneous components. In particular, the multilateral specification is designed as the rook scheme $W_{s-1,t} = X_{s-1,t}$; in this case, the LS method should be biased for all parameters. We perform 500 replications of the system (4) with $p = 1$, $n = 13$, $m = 12$, $T = 11$, $|\delta_i| = 0.5, 1.0$, for $i = 1 \ldots 5$ and $\sigma_e = 1$; notice that the actual sample size $N = (n-2)(m-2)(T-1) = 1100$ is large enough. Simple means, root mean squared errors (RMSE = [Var+Bias^2]$^{1/2}$) and P-value of the Normality test are reported in Table 1.

Since the parameters have the same size, their statistics are directly comparable and can be averaged to provide global indicators. As a result, one may note that the worst performance occurs with multilateral W and non-stationary coefficients; whereas the models with unilateral W benefit from the greater signal-to-noise ratio σ_Y/σ_e induced by $|\delta_i| = 1$. This property is known as *super consistency*, and its side-effect is the non-Normal distribution of the estimates. This complicates the statistical inference, requiring non-standard distributions, as in the tests for unit-roots in time series (e.g. Kilian and Lütkepohl, 2017). Unit-roots in STAR models may be defined in the Eq. (5), but they are rare and do not affect forecasting.

3. Empirical applications

3.1. Application to the Chesapeake area

We carry out an extended empirical application to illustrate and check the methods presented in Section 2. The case study deals with the time-lapse video of Foresman and Masuoka (1996) about the annual urban growth of the Washington–Baltimore area from 1792 to 1992. The video is realized by combining historical maps, census information, aerial photography, LandSat 1–4 images and stochastic interpolation; see Foresman et al. (1997) and Acevedo and Masuoka (1997) for details on the data source and the editing process. The authors considered it as a space–time framework that describes the land use and human impact on the Chesapeake region over two centuries; however, it can be managed as a sequence of real remote-sensing imagery and fitted with STAR models.

The video can be downloaded from the Internet in MP4 format and imported into the MatLab software; it is a 4D-array $\widetilde{\mathbf{Y}} = \{\widetilde{Y}_{ijlt}\}$ of size $n = 480$ rows, $m = 720$ columns, $k = 3$ color bands and $T = 1055$ frames. Deleting the borders and 50 redundant frames, its actual size becomes $410 \times 370 \times 3 \times 1005$, which means 5 repeated frames per year; the last one is shown in Fig. 2a. The pixels have `Uint8` format, i.e. $0 \le \widetilde{Y}_{ijlt} \le 255$; the spatial averages of the bands $\overline{\widetilde{Y}}_{lt}$, and of the gray-transformation, are displayed in Fig. 2d. One may note that only the red band has a positive trend and may represent the urban growth.

We restrict the analysis to the Baltimore–Washington metropolitan area of size 240×220 and we select the central frames (out of the 5 per year), a total $T = 1005/5 = 1992–1792 = 201$ images. To detect the built-up area, the underlying digital terrain model (DTM) is removed as $\boldsymbol{Y}_t = (\widetilde{\boldsymbol{Y}}_t - \widetilde{\boldsymbol{Y}}_0)$ (where $\widetilde{\boldsymbol{Y}}_0$ is the initial frame). Fig. 2b displays the resulting $\boldsymbol{Y}_T$ and Fig. 2c shows its first (red) layer in pseudocolor. Finally, Fig. 2e plots the spatial averages of the red layer $\overline{Y}_{1t}$; given the underlying trend, we start the analysis at $t = 1881$. The final dataset $\mathbf{Y}$ becomes 3D with size $[n, m, T] = [240, 220, 111]$, a total $N = 5{,}860{,}800$ data points Y_{ijt}. This is processed with the Matlab software on a standard laptop computer, with reasonable computing times.

Given the marked time-trend in Fig. 2e, an analysis of the differenced series is in order, as in classical time-series analysis (e.g. Kilian and Lütkepohl, 2017)

$$y_{ijt} = \Big(Y_{ijt} - Y_{ij,t-1}\Big), \tag{13a}$$

$$\hat{Y}_{ij,t+1} = Y_{ijt} + \hat{y}_{ij,t+1}, \tag{13b}$$

$$\hat{Y}_{ij,t+2} = \hat{Y}_{ij,t+1} + \hat{y}_{ij,t+2}, \tag{13c}$$

where $\{y_{ijt}\}$ is modeled as the STAR model (4) and forecasted with Eqs. (13b), (13c). A data transformation which useful in the presence of heteroskedasticity (time-varying σ_e^2, which yields inefficient estimates), is the logarithm

$$\mathcal{Y}_{ijt} = \log(Y_{ijt} + 1), \tag{14a}$$

$$\hat{\hat{Y}}_{ij,t+1} = \exp(\hat{\mathcal{Y}}_{ij,t+1}) - 1, \tag{14b}$$

An inverse transformation is necessary to go back to the original data in forecasting; however, owing to the Jensen inequality: $\mathrm{E}[g(x)] \ne g[\mathrm{E}(x)]$, it may affect the properties of the linear predictor in Eq. (7), because $\hat{\hat{Y}}_{ij,t+1} \ne \hat{Y}_{ij,t+1}$.

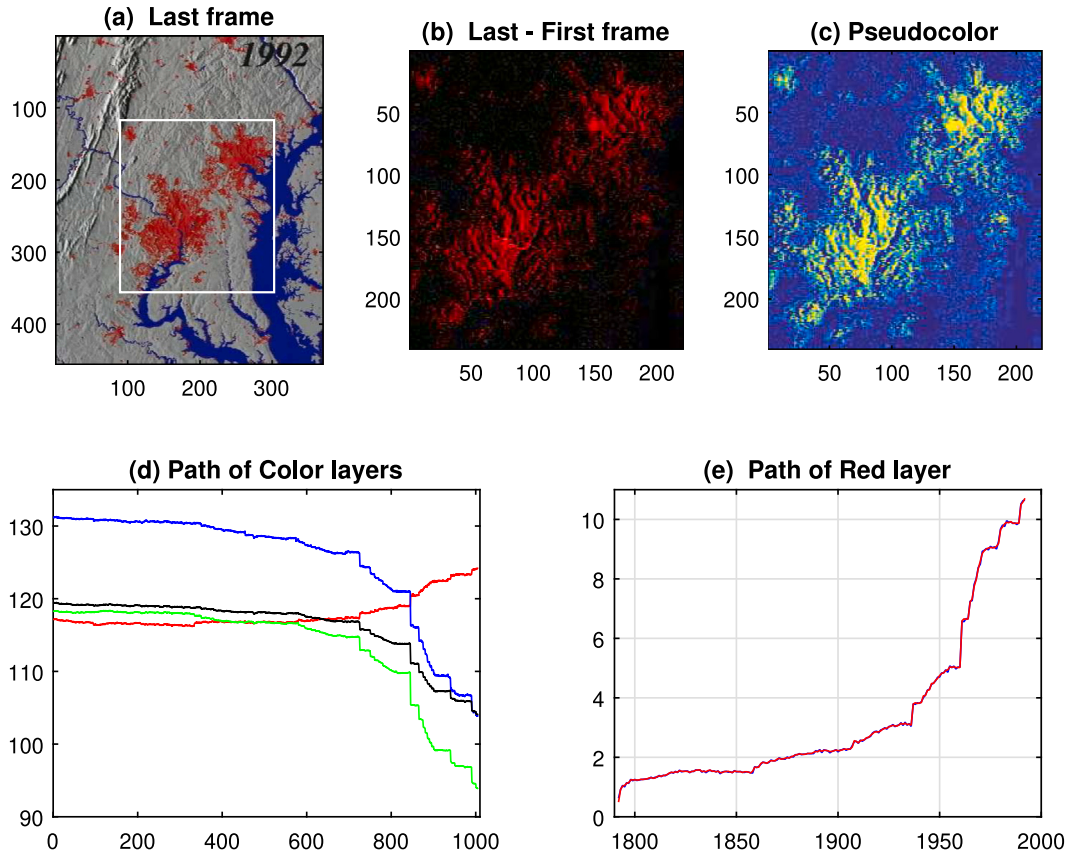


Fig. 2. Features of the Foresman and Masuoka (1996) video: (a) Last raw frame $\widetilde{\boldsymbol{Y}}_T$; (b) Transformed frame $\boldsymbol{Y}_T = (\widetilde{\boldsymbol{Y}}_T - \widetilde{\boldsymbol{Y}}_0)$; (c) Red layer of $\boldsymbol{Y}_T$ in pseudocolor; (d) Time path of spatial averages $\overline{\widetilde{Y}}_{lt}$ on the whole area; (e) Path of the red layer in the metropolitan area. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

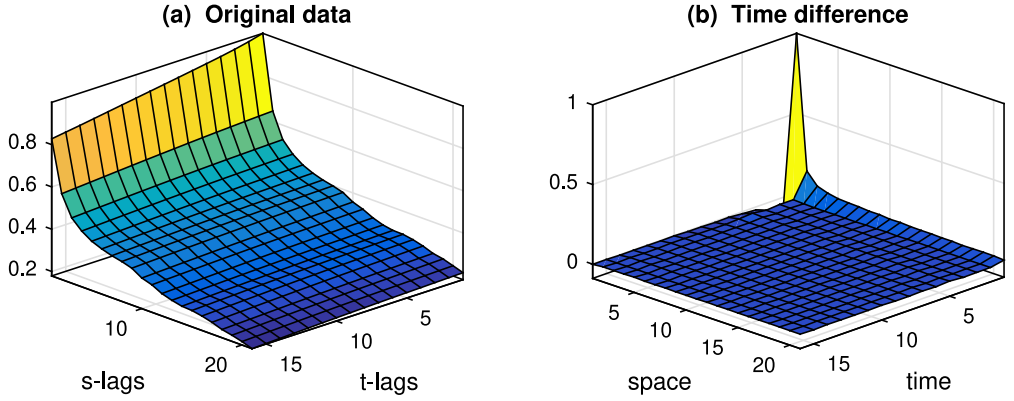


Fig. 3. Space–time Autocorrelation function of the data in Fig. 2c: (a) Original data Y_{ijt}; (b) Differenced series y_{ijt} (13a).

A preliminary data analysis, with descriptive statistics, regards the space–time ACR function $R(h,k) = \mathrm{Corr}\left(Y_{ijt}, Y_{i-h,j-h,t-k}\right)$. The estimates in Fig. 3 show a significant autocorrelation in the original data, which nearly disappears in the detrended series y_{ijt}. This means that the underlying STAR model of Y_{ijt} should be, at most, of second order ($p = 2$). The design of experiments for testing the performance of the methods is as follows: (1) Estimate the model (4) on the original and differenced series, with and without the contemporaneous components W (designed as triangular); (2) Leave the last two images (1991 and 1992) out of the estimation phase and use them for forecasting evaluation, with the statistics (8). The specific treatment of the components W is due to their peculiar role in forecasting, as they complicate the prediction algorithms and may affect their performance. The resulting LS estimates (9) are summarized in Table 2.

Owing to the large sample-size N = (240-4)(220-4)(111-4) = 5,454,432, the t-type statistics (in parentheses) are inflated and all regressors Y, X, W, Z are statistically significant. Their actual importance may be judged by the absolute size of their coefficients, which are related to auto-correlations; the minimum size should be $|\hat{\delta}_k| \geq 0.1$, which corresponds to $|t_k| > 100$. The fitting performances of the models in Table 2 are very close, with the main role of the individual AR components $Y_{ij,t-k}$. Despite the presence of residual outliers, the indices $R^2 \approx 0.95$ are excellent; one decreases to 0.82 by transforming the data in logarithms.

Table 2
LS estimates (and t-type statistics) (9), (11) of the parameters of the model (4) on original and differenced series, with and without the contemporaneous components W (2a), (3a), on the data of Fig. 2 in the period 1881–1990. The best results are in bold.

Param.	Y no W	Y with W	y no W	y with W	$\log(Y+1)$
α_0	0.049 (37.7)	0.049 (38.2)	0.090 (79.3)	0.045 (41.6)	0.012 (38.6)
ϕ_1	0.676 (1400)	0.674 (1411)	−0.334 (−676)	−0.346 (−738)	0.566 (1303)
ϕ_2	0.258 (530)	0.259 (538)	−0.135 (−272)	−0.143 (−303)	0.290 (667)
θ_1	.	0.101 (327)	.	0.442 (726)	0.146 (331)
θ_2	.	−0.019 (−70.6)	.	0.047 (63.6)	−0.002 (−5.34)
β_{11}	0.129 (126)	0.079 (77.2)	0.123 (118)	0.206 (207)	0.021 (22.9)
γ_1	0.019 (20.3)	−0.007 (−7.20)	0.031 (32.8)	0.049 (55.4)	−0.011 (−13.5)
β_{21}	0.041 (47.8)	0.046 (54.7)	0.049 (56.7)	0.033 (39.5)	0.027 (33.1)
β_{12}	−0.052 (−50.4)	−0.069 (−67.4)	0.080 (75.8)	0.109 (110)	−0.017 (−19.1)
γ_2	−0.029 (−31.7)	−0.032 (−35.5)	0.025 (26.2)	0.022 (25.0)	−0.022 (−25.7)
β_{22}	−0.032 (−37.7)	−0.025 (−29.7)	0.035 (39.6)	0.019 (23.3)	−0.003 (−3.64)
σ_e^2	7.006	6.855	7.012	**6.269**	0.247
R^2	0.946	0.947	0.945	**0.951**	0.817
BIC	1.947	1.925	1.948	**1.836**	.

Table 3
Evaluation of the out-of-sample forecasts of the STAR models in Table 2 with the MSFE statistics (8) on the frames 1991, 1992 of the Chesapeake area video.

Estimator	Year ℓ	Y	$Y+W$	y	$y+W$	$\log(Y+1)$	max R^2_ℓ
LS (9)	1991	**9.32**	9.62	10.15	10.16	19.76	**0.970**
" "	1992	**16.64**	17.53	19.57	19.65	33.08	0.946

Table 4
Estimation and forecast statistics of STAR(p) models applied to the data of Fig. 5. Legends are as in Tables 3 and 4; the values in parentheses of MSFE are R^2 (8b).

Param.	$p=1$ no W	$p=1$ with W	$p=2$ no W	$p=2$ with W
α_0	5.310 (258)	1.095 (82.0)	2.976 (149)	0.816 (62.2)
ϕ_1	0.587 (667)	0.609 (1076)	0.526 (441)	0.495 (634)
ϕ_2	.	.	0.217 (184)	0.204 (264)
θ_1	.	0.780 (2235)	.	0.913 (1560)
θ_2	.	.	.	−0.197 (−334)
β_{11}	0.347 (373)	−0.402 (−587)	0.019 (7.96)	−0.450 (−285)
β_{12}	.	.	0.101 (56.2)	0.139 (116)
β_{21}	.	.	−0.006 (−2.48)	−0.185 (−120)
β_{22}	.	.	0.109 (61.9)	0.071 (61.5)
σ_e^2	237.2	97.96	213.7	**91.25**
R^2	0.824	0.927	0.841	**0.932**
BIC	5.468	4.586	5.364	**4.514**
MSFE 2021	177.6 (0.892)	294.9 (0.821)	**167.4 (0.898)**	244.1 (0.852)
MSFE 2022	316.6 (0.818)	589.2 (0.660)	**283.7 (0.837)**	420.6 (0.758)

The predictions of the models in Table 2 are evaluated on the images $\boldsymbol{Y}_{1991}, \boldsymbol{Y}_{1992}$ kept out of estimations. They are computed with the function (7) and then evaluated with the MSFE statistic (8), with $T' = 1990$. The results are displayed in Table 3; as is natural, the errors increase with the forecast span $\ell = 2$. The best model is that of the original series Y, and without the contemporaneous terms. Although W have a unilateral structure (which makes LS unbiased), their effect on fitting is negligible and on forecasting is slightly negative. Computational details are in Appendix A.

The results in Table 3 are excellent if evaluated with the index R^2_ℓ (8b), which are even greater than those of the in-sample fitting in Table 2. The forecasts of the best model are displayed in Fig. 4, together with the corresponding ground data 1991, 1992. We recall that images regard the first (red) layer without the geographical background and displayed in pseudocolor; the corresponding forecast errors in Fig. 4c,f do not show significant spatial patterns. To generate long-term forecasts (e.g. 10, 20 years ahead), one should consider only the decadal frames 1792, 1802 ... 1992, but the sample size reduces to $T = 21$.

The major application of STAR models to urban planning is monitoring the phenomena of urban sprawl and soil occupation, especially in vulnerable and risky areas, such as wetlands (see Foresman et al., 1997). At a wider scale, long-term projections may identify gravitational dynamics and clustering tendencies of groups of cities, so as to plan transportation infrastructures at the regional level. On the contrary, the major limitation of remote-sensing forecasts is in historical and built-up parts of cities, which change slowly and require ground informations. These can be integrated into the STAR framework, by "gridding" the microdata of buildings, households, stations, etc.

3.2. Application to the Amazon area

Computing the forecasts of Fig. 4 in the format of Fig. 2a requires a lot of work; namely, fitting STAR models to the data of the green and blue channels; encoding their forecasts in `uint8` and overlying (and clipping) the color image to the digital terrain model (DTM) of the area with GIS software. Instead, we develop an application to recent remote-sensing data, as the videos of Google Earth (2023), edited with LandSat 5–9 and Copernicus images.

Recent videos of historical cities are not interesting as the great urban expansion has stopped and only minor transformations are occurring. Therefore, we consider the Iquitos city in Peru from 1984 to 2022, in low-resolution (240p) on an area of about 18 × 22 km. This yields a data array in the grayscale of size $[n, m, T] = [288, 368, 39]$, which provides $N = 4{,}133{,}376$ observations. Decadal frames are shown in Fig. 5; notice the significant displacement of the Amazon river bed in 40 years.

For the sake of comparison, we fit the data of Fig. 5 in grayscale, with STAR models of orders $p = 1, 2$, with and without the simultaneous components W, and by aggregating the spatial components X, Z in Eqs. (2), (3). The results are reported in Table 4, together with the forecast statistics on the last two frames. Unlike the Chesapeake application, one may note that simultaneous components do improve the in-sample fitting, but also worsen the out-of-sample forecasting. STAR forecasts then prefer parsimonious models, and their reliability is supported by the large R^2 indexes; the display of $\widehat{\boldsymbol{Y}}_{2021}$ in pseudocolor is given in Fig. 6.

4. Advanced aspects

4.1. Adaptive algorithms

Robust estimates. An important issue in model estimation is the presence of anomalous observations Y^o_{ijt}, which belong to a long-tailed distribution $f_o(Y)$. These data generate biased parameter estimates and wrong series forecasts; therefore, Y^o_{ijt} must be detected and censored. A simple method is based on the identification of residuals $\hat{e}^o_{ijt} = \left(Y^o_{ijt} - \hat{\boldsymbol{\delta}}'_N \boldsymbol{\xi}_{ijt}\right)$, which fall outside the assumed model $\mathrm{N}(0, \sigma_e)$ with 95% probability, i.e. $\left|\hat{e}^o_{ijt}\right| > 2\,\sigma_e$. This test requires an unbiased estimator of the residual variance, such as that based on the mean absolute deviation (MAD): $\tilde{\sigma}_e = 1.4825\ \mathrm{median}\left(\ \left|\hat{e}_{ijt} - \mathrm{median}(\hat{e}_{ijt})\right|\ \right)$. It is

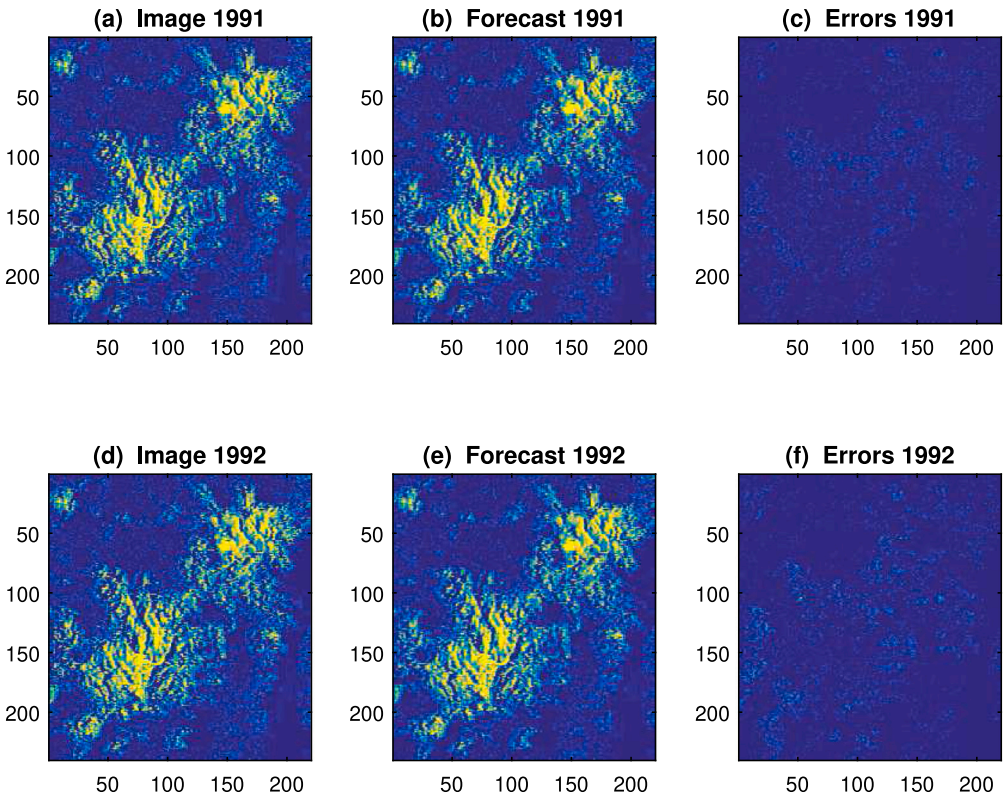


Fig. 4. Display of out-of-sample forecasts of the best model in Table 3: (a–c) Year 1991; (d–f) Year 1992; (a, d) Ground images; (b, e) Forecasts (7); (c, f) Forecast errors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Fig. 5. Frames of the timelapse video of Iquitos city of Google Earth (2023).

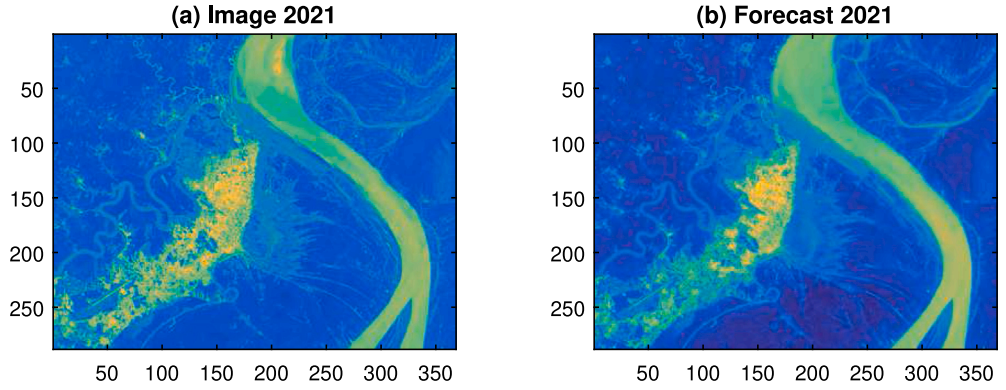


Fig. 6. Frame 2021 of Iquitos Google Earth video and its STAR forecast.

known that the MAD estimator is unbiased under the assumption of Gaussian residuals.

Once an outlier is detected, the corresponding observation must be replaced by its in-sample prediction $\widetilde{Y}_{ijt} = \left(Y_{ijt} - \widehat{e}^{o}_{ijt}\right)$ and the LS estimates may be recomputed as in (9a), using the cleaned series $\widetilde{Y}_{ijt}$, to obtain $\widetilde{\delta}_N$. This solution can be iterated and is preferable to robust methods which minimize the sum of censored residuals $\sum_{ijt} \rho(e_{ijt})$, where $0 < \rho(x) \ll \infty$ (e.g. Bustos et al., 2009; Bayer et al., 2021). Indeed, this approach is computationally demanding and is robust only against outliers in the dependent variable Y_{ijt}, but not in the regressors ξ_{ijt}. Instead, in the outlined approach, the cleaned series $\widetilde{Y}_{ijt}, \widetilde{\xi}_{ijt}$ can

Table 5
Performance of the LS estimators (9) and (15) applied to the model (4) (with order $p = 1$, triangular component W and coefficients $|\delta_i| = 0.5$), contaminated by 6 additive outliers; Mean, RMSE, P-value are over 500 replications.

Param.	Value	LS (9)			Robust LS (15)		
		Mean	RMSE	N-test	Mean	RMSE	N-test
α_0	0.5	−0.209	0.710	>0.5	0.461	0.058	>0.5
ϕ_1	−0.5	−0.001	0.499	>0.5	−0.482	0.031	>0.5
θ_1	0.5	0.001	0.499	0.211	0.456	0.055	0.317
β_1	−0.5	−0.001	0.499	0.078	−0.469	0.057	0.105
γ_1	0.5	0.001	0.499	0.468	0.510	0.044	>0.5
Abs. Ave.	0.5	0.043	0.541	0.351	0.476	0.049	0.384

be computed in *parallel* as

$$\widetilde{Y}_{ijt} = \begin{cases} \widehat{\boldsymbol{\delta}}_N' \widetilde{\boldsymbol{\xi}}_{ijt}, & \text{if} \quad \left|\widehat{e}_{ijt}\right| > \kappa\, \sigma_e, \\ Y_{ijt}, & \text{otherwise}, \end{cases} \tag{15}$$

where $\kappa > 0$ is a tuning constant that depends on the probability tails of the Normal distribution (e.g. $\kappa = 2.57$ for 1%).

In summary, the "linear" robust algorithm is as follows:

(1) Get initial LS estimates of parameters and residuals $\widehat{\boldsymbol{\delta}}_N, \widehat{e}_{ijt}$;

(2) Get the MAD estimate $\widetilde{\sigma}_e$ and detect (ijt) where $\left|\widehat{e}_{(ijt)}\right| > 3\,\widetilde{\sigma}_e$;

(3) Replace $Y_{(ijt)}$ with $\widetilde{Y}_{(ijt)} = \widehat{\boldsymbol{\delta}}_N' \widetilde{\boldsymbol{\xi}}_{(ijt)}$ and recompute $\widetilde{\boldsymbol{\delta}}_N$ with (9);

finally, iterate the entire procedure if necessary.

To check the robustness of the above method, we extend the simulation experiment of Table 1 (on a first-order model with triangular W component), by contaminating the series Y_{ijt} with 6 additive outliers in a fixed position, one every two frames. Table 5 shows that the LS estimator (9) is strongly biased, whereas the method (15) approaches the ground values.

Varying Parameters. At the computational level, the matrix of squared regressors $\boldsymbol{R}_N$ in (9a), (11a) can be computed recursively to save computer memory as

$$\boldsymbol{P}_{i,j,t} = \boldsymbol{P}_{i,j-1,t} + \boldsymbol{\xi}_{ijt}\boldsymbol{\xi}_{ijt}', \quad j = p+1 \ldots m-p, \tag{16a}$$

$$\boldsymbol{Q}_{i,t} = \boldsymbol{Q}_{i-1,t} + \boldsymbol{P}_{i,m,t}, \quad i = p+1 \ldots n-p, \tag{16b}$$

$$\boldsymbol{R}_t = \boldsymbol{R}_{t-1} + \boldsymbol{Q}_{n,t}, \quad t = p+1 \ldots T, \tag{16c}$$

where the entries W, X, Z of the vector $\boldsymbol{\xi}_{ijt}$ are computed with Eqs. (2)–(3). This calculation is useful for developing adaptive LS algorithms.

The space–time variability of parameters $\boldsymbol{\delta}_{ijt}$ may be due to external factors, that have been omitted in the STAR modeling; or nonlinear dynamics, which are approximated by the linear representation. Methods for estimating $\boldsymbol{\delta}_{ijt}$ can be obtained by adapting the LS algorithms (9), (11), as in Young (2011); in particular, the recursive expression (16c) is suitable for time-varying parameters $\boldsymbol{\delta}_t$, combined with weighting the observations:

$$\boldsymbol{R}_t = \lambda\, \boldsymbol{R}_{t-1} + \sum_{i=p+1}^{n-p}\sum_{j=p+1}^{m-p} \boldsymbol{\xi}_{ijt}\boldsymbol{\xi}_{ijt}', \qquad \boldsymbol{R}_0 = \mathbf{O}, \tag{17a}$$

$$\boldsymbol{S}_t = \lambda\, \boldsymbol{S}_{t-1} + \sum_{i=p+1}^{n-p}\sum_{j=p+1}^{m-p} \boldsymbol{\xi}_{ijt} Y_{ijt}, \qquad \boldsymbol{S}_0 = \mathbf{0}, \tag{17b}$$

$$\widehat{\boldsymbol{\delta}}_t = \boldsymbol{R}_t^{-1}\boldsymbol{S}_t, \qquad t = p+1 \ldots T, \tag{17c}$$

where $0 < \lambda < 1$ gives more weight to the recent observations and avoids the convergence of (17c) toward a constant value. Notice that when $\lambda = 0$ and $p = 1$, then each $\widehat{\boldsymbol{\delta}}_t$ is the result of the regression of the frame $\boldsymbol{Y}_t$ on $\boldsymbol{Y}_{t-1}$.

By adapting the formula (9b): $\widehat{\boldsymbol{\delta}}_N = \boldsymbol{\delta} + \boldsymbol{R}_N^{-1}\sum_{ijt}\boldsymbol{\xi}_{ijt}e_{ijt}$, one can rewrite the estimator (17c) with the prediction errors as

$$\widehat{e}_{ijt} = \left(Y_{ijt} - \widehat{\boldsymbol{\delta}}_{t-1}'\boldsymbol{\xi}_{ijt}\right), \qquad \widehat{\boldsymbol{\delta}}_0 = \mathbf{0}, \tag{18a}$$

$$\widehat{\boldsymbol{\delta}}_t = \widehat{\boldsymbol{\delta}}_{t-1} + \boldsymbol{R}_t^{-1}\sum_{i=p+1}^{n-p}\sum_{j=p+1}^{m-p} \boldsymbol{\xi}_{ijt}\,\widehat{e}_{ijt}, \tag{18b}$$

$$\widehat{\sigma}_t^2 = \lambda\,\widehat{\sigma}_{t-1}^2 + (1-\lambda)\,\widehat{e}_{ijt}^2, \qquad \widehat{\sigma}_0^2 = 0, \tag{18c}$$

where $\boldsymbol{R}_t$ is computed in (17a), and (18c) is an exponentially weighted variance. The algorithm (18b) can be made robust by trimming the prediction errors (18a); accordingly, also the regressors may be filtered as in Eq. (15), to obtain

$$\widehat{e}_{ijt} = \left(Y_{ijt} - \widehat{\boldsymbol{\delta}}_{t-1}'\widetilde{\boldsymbol{\xi}}_{ijt}\right), \tag{19a}$$

$$\widetilde{e}_{ijt} = \text{sign}(\widehat{e}_{ijt})\,\kappa\,\widehat{\sigma}_{t-1}, \quad \text{if} \quad \left|\widehat{e}_{ijt}\right| > \kappa\,\widehat{\sigma}_{t-1}, \tag{19b}$$

$$\widetilde{Y}_{ijt} = \widehat{\boldsymbol{\delta}}_{t-1}'\widetilde{\boldsymbol{\xi}}_{ijt}, \quad \text{if} \quad \left|\widehat{e}_{ijt}\right| > \kappa\,\widehat{\sigma}_{t-1}, \tag{19c}$$

The selection of the coefficients λ, κ of the joint algorithm (18)–(19) can be done with a cross-validation criterion based on the uncensored prediction errors:

$$\left[\widehat{\lambda}, \widehat{\kappa}\right]_N = \arg\min \sum_{t=p+1}^{T}\sum_{i=p+1}^{n-p}\sum_{j=p+1}^{m-p}\left(Y_{ijt} - \widehat{\boldsymbol{\delta}}_{t-1}'\widetilde{\boldsymbol{\xi}}_{ijt}\right)^2, \tag{20}$$

however, for STAR models this solution is computationally very demanding and should be performed only on the most recent frames.

Furthermore, an estimator for spatially-varying parameters can be obtained by fitting the model (4) at each location ij with the algorithm (18), obtaining

$$\boldsymbol{R}_{ijt} = \lambda\, \boldsymbol{R}_{ij,t-1} + \boldsymbol{\xi}_{ijt}\boldsymbol{\xi}_{ijt}', \qquad \boldsymbol{R}_{ij0} = \mathbf{O}, \tag{21a}$$

$$\widehat{\boldsymbol{\delta}}_{ijt} = \widehat{\boldsymbol{\delta}}_{ij,t-1} + \boldsymbol{R}_{ijt}^{-1}\boldsymbol{\xi}_{ijt}\left(Y_{ijt} - \widehat{\boldsymbol{\delta}}_{ij,t-1}'\boldsymbol{\xi}_{ijt}\right), \tag{21b}$$

this is also time-varying when $\lambda < 1$. Through the components W, X, Z, the estimator (21) involves the contiguous data $i \pm p, j \pm p$; however, it can be made more spatially smoothed by computing $\boldsymbol{R}_{ijt}, \boldsymbol{S}_{ijt}$ on a neighbor of ij as

$$\widetilde{\boldsymbol{R}}_{ijt} = \lambda\, \widetilde{\boldsymbol{R}}_{ij,t-1} + \sum_{h=i-d}^{i+d}\sum_{k=j-d}^{j+d} \lambda^{|i-h|+|j-k|}\, \boldsymbol{\xi}_{hkt}\,\boldsymbol{\xi}_{hkt}', \qquad \widetilde{\boldsymbol{R}}_{ij0} = \mathbf{O}, \tag{22}$$

where $2d \ll n, m$ is a square window and λ also spatially weights the observations. Unfortunately, the algorithm (22) requires more border observations and cannot be managed recursively over space; hence, to obtain $\widetilde{\boldsymbol{\delta}}_{ijt}$ it may be preferable to smooth the local estimates (21b) with a Gaussian filter with bandwidth $\kappa > 0$

$$\widetilde{\boldsymbol{\delta}}_{ijt} = \sum_{h=1}^{n}\sum_{k=1}^{m} w_{hk}(i,j)\,\widehat{\boldsymbol{\delta}}_{hkt}, \tag{23a}$$

$$w_{hk}(i,j) \propto \exp\left\{-\left[(h-i)^2 + (k-j)^2\right]/2\kappa^2\right\}. \tag{23b}$$

4.2. Empirical application

We continue the application to the Chesapeake area to illustrate the adaptive methods and check their performance on the STAR model; owing to the results of Table 3, we only process the raw data Y. The robust method (15) needs the initial LS estimates in Table 2; the MAD statistics provides the variance $\widetilde{\sigma}_e^2 = 0.281$, which is much smaller than the LS one $\widehat{\sigma}_e^2 = 7.01$. This difference is a symptom of the presence of outliers and/or heteroskedasticity; hence, the application of the method (15) is in order. The resulting estimates are similar to those in Table 2, but with greater values of the AR parameters, e.g. $\widehat{\phi}_1 > 0.7$; we only report their MSFE statistics computed on the out-of-sample frames 1991, 1992 in Table 6. We note that the robust method (15) with $\kappa = 3$ improves the LS results of Table 3 by about –6%.

The second block of Table 6 provides the results of the time-varying parameters (17)–(19); Fig. 7 shows the path of recursive estimates for $\lambda = 0.9$ and 0 (sequential estimates, obtained by regressing $\boldsymbol{Y}_t$ on $\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}$). The jumps of parameters in Panel (a) occur in correspondence with the changes in the basic cartography (see Foresman et al., 1997); however, Panel (b) shows a more complex situation of

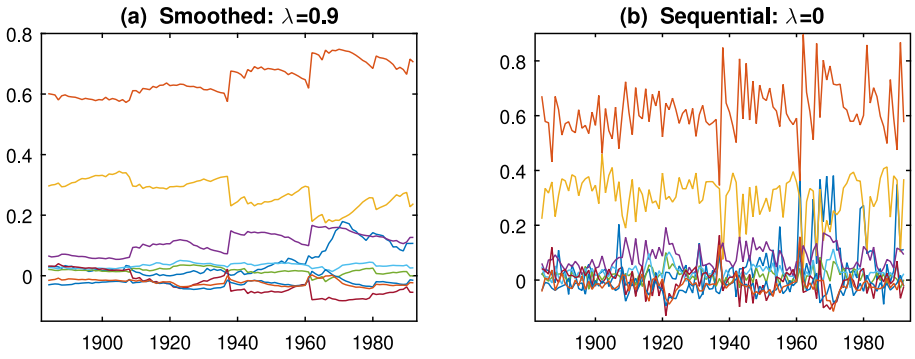


Fig. 7. Time path 1884–1992 of the adaptive estimates (17) of the parameters of the STAR model (4) without the component W: (a) Smoothed, with $\lambda = 0.9$; (b) Sequential, with $\lambda = 0$. The greatest values are those of the time AR parameters ϕ_1, ϕ_2.

Table 6
Evaluation of the out-of-sample forecasts of the STAR model estimated with adaptive algorithms (15)–(23). The 3 blocks correspond to: robust constant, time-varying and space–time varying parameters. Computational details are in Appendix A.

Estimator	Year	Y		$Y+W$	
		MSFE	R^2_ℓ	MSFE	R^2_ℓ
Robust (15)	1991	**8.76**	0.972	9.21	0.970
($\kappa = 3$)	1992	**16.37**	0.947	16.90	0.945
T-var (17)	1991	9.19	0.970	9.43	0.969
($\lambda = 0.97$)	1992	16.55	0.947	17.27	0.944
T-var (19)	1991	**8.36**	**0.973**	8.41	0.973
($\lambda, \kappa = 0.9, 2$)	1992	**16.33**	0.947	16.48	0.947
ST-var (23)	1991	15.06	0.952	18.20	0.941
($\lambda, \kappa = 1{,}0.5$)	1992	27.12	0.913	35.53	0.886
ST-var (22)	1991	10.51	0.967	.	.
($\lambda, d = 1{,}1$)	1992	17.81	0.945	.	.

nonstationarity. The time-varying forecast function is given by Eq. (7) with the parameters $\hat{\boldsymbol{\delta}}_t$ of the year $t = 1990$; these estimates slightly improve the forecasts of Table 3, but get better when they are made robust as in Eq. (19), obtaining –10% MSFE. A strong improvement is achieved by using the estimates $\hat{\boldsymbol{\delta}}_t$ at $t = 1991$, but this solution is *not* out-of-sample and benefits from the future information.

Lastly, the third block of Table 6 provides the results of the space–time varying estimates (21)–(23); in this context, the estimators are more complex and slow, and are unable to improve previous forecasts. The only interesting fact is the spatial pattern of the parameters $\hat{\boldsymbol{\delta}}_{ij,T-2}$; Fig. 8 displays those of α_0, ϕ_1 in pseudocolor. They reveal hidden features of the dataset; in particular, the spatial ACR is greater in the areas of urban growth. As before, the forecasts strongly improve by using the in-sample estimates $\hat{\boldsymbol{\delta}}_{ij,T-1}$; but these belong to the forecast horizon.

In summary, the best forecasting method is based on the robust time-varying algorithm (19), as it has the smallest MSFE statistics. However, the best R^2_ℓ indices of Tables 3 and 5 are similar; this explains why the forecasts of the method (19) resemble those in Fig. 4, and, therefore, they are not displayed here. Notice that $\max R^2_\ell > 0.97$ guarantees the reliability of the forecasts; a result that can hardly be achieved by the numerical methods mentioned in the introduction. To have a visual appearance similar to Fig. 2a, one should forecast also the blue and green layers of the video frames and finally add the DTM model of the area. Finally, to obtain long-term forecasts, one may focus on decadal frames, although the sample size drastically reduces.

Adaptive parameter estimates reflect hidden features of the Chesapeake video, but some interpretations are possible. The time jumps in Fig. 7 probably occur in correspondence with the changes in historical cartography, but they also represent specific phases of the urban development. However, the sequential estimates in Panel (b) display a more complex situation, with a leading role of the AR parameter $\hat{\phi}_{1t} > 0.5$. The spatial paths in Fig. 8 show that the constant $\hat{\alpha}_{0ij}$ may identify the two gravitational poles of the area; whereas the ACR parameter $\hat{\phi}_{1ij}$ is greater in zones where the urban growth has been fastest.

The extended numerical application to the Baltimore–Washington area has portrayed a fast urban expansion during the 19th century, due to population growth. Urbanization has spread into a wetland, forested and agricultural ecosystems (Foresman et al., 1997); however, with less than 1.5 million inhabitants, they did not take the form of megalopolises. Despite the recent population decline (–100,000 in the last 30 years), the urban expansion has continued, mostly as building replacement and urban project completions (see Google Earth maps). This positive trend is consistent with the forecasts of the STAR model in Fig. 4, whose out-of-sample performance, measured by R^2 indexes, is outstanding.

5. Conclusions

STAR modeling of remote sensing data is an effective method for representing the urban growth of cities, both at the spatial and temporal levels. Its major strength is the absence of a-priori assumptions about the growth dynamics, its main limitation is the need for real data; fortunately, these are available, with increasing coverage and resolution, from modern technologies. Out-of-sample forecasting is the main result of the method; it may provide alert signals about natural and social events in any area. In particular, uncontrolled soil occupation and illegal building constructions in vulnerable zones can be detected by segmenting the image forecasts. Policy actions can then be planned at the local and regional levels to manage the ongoing tendencies.

From the technical viewpoint, STAR modeling of timelapse videos requires regularity conditions on both the data and the models; the frames must be properly georeferenced, equalized and regularly distributed in time. Similarly, the model structure must avoid over-parametrization, by using pixel aggregation; simultaneous relationships between cells must be of recursive type, to have unbiased estimates of parameters and forecasts. In this context, the LS method can be made robust against outliers and adapted to conditions of varying parameters. In-sample forecasts may also be used to restore parts of the frames corrupted by noise or blurred by clouds.

From the statistical viewpoint, the paper has shown the minor importance of the regressors W concerning the simultaneous spatial relationships. These components are important in static spatial AR models (SAR); but in the dynamic context, they are not useful in forecasting, for the same reasons as they are not in vector AR models (VAR). In practice, STAR and VAR systems are strictly related and it is known that simultaneous components do not improve the forecasts in the latter (see Kilian and Lütkepohl, 2017 p.46). The best forecasts are provided by robust sequential estimators, suitable for slowly time-varying parameters; whereas spatially-varying algorithms only complicate the calculations.

Improvements in the STAR methodology can be pursued by including in the models exogenous variables that drive the cities' growth,

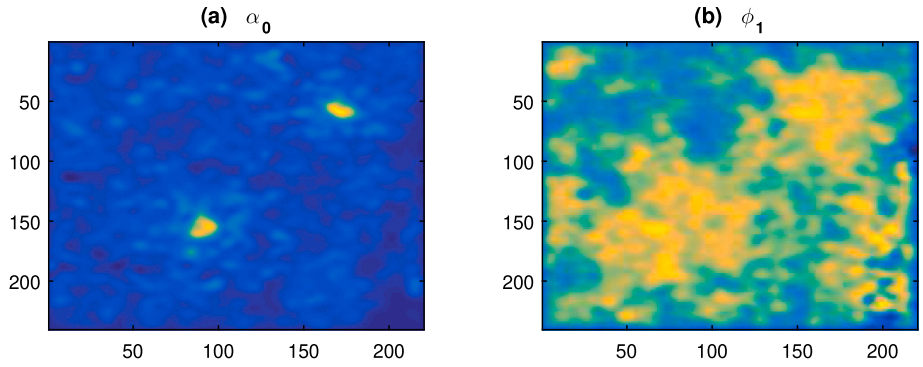


Fig. 8. Spatial variation of the coefficients α_0, ϕ_1 of the STAR model (4) without W at time t = 1990, estimated with the method (23) with $\lambda = 1$ and $\kappa = 1/2$.

Table A.1
Execution times of the methods described in the paper applied to the cropped Baltimore dataset ($240 \times 220 \times 110$) on a small notebook with MatLab R15a software.

Purpose	Methods	Results	Seconds	Function
Dataset editing	.	Fig. 2	37.6	`STAR_data.m`
Autocorrelation	.	Fig. 3	121.3	`STAR_acr.m`
Simulation exp.	.	Table 1	4.76	`STAR_sim.m`
OLS estimation	Eqs. (9), (11)	Table 2	6.15	`STAR_ols_fast.m`
Forecasting	Eqs. (7), (8)	Table 3	3.15	`STAR_for.m`
Robust est.	Eq. (15)	Table 5	7.35	`STAR_rob_fast.m`
Time-varying est.	Eq. (17)	Fig. 7	9.18	`STAR_seq_fast.m`
Space-varying est.	Eqs. (21)–(23)	Fig. 8	85+26	`STAR_var_smth.m`

such as population movements, real estate, energy consumption and climatic conditions. These variables can be organized in raster form by "gridding" the microdata of buildings, households, stations and may be modeled themselves as STAR systems. The introduction of nonlinear regressors (as Y^2_{ijt-1}) may also be useful; these components leave the models linear in the parameters, so that LS estimators can still be applied. To mitigate the model complexity, and reduce the computing time, one may segment the frames into a few groups (urban, forest and wetland) and fit each group with separate models.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could influence the work reported in this paper.

Acknowledgments

The author thanks the Editors for their assistance in the submission and the Reviewers for useful technical suggestions.

Appendix A. Computational aspects

This Appendix provides computational details of the methods discussed in the paper, with reference to the data of the case study. They are obtained with the software developed for the research: a set of MatLab scripts that provide a complete toolbox. This program is now offered as supplementary material, together with the dataset. Table A.1 summarizes the execution times performed on a small NoteBook (with DualCore 1.66 GHz, 64 Bit, 4 GB Ram) running MatLab R15a, under Windows system, in operative conditions (other apps opened, etc.).

Given the 1 billion pixels of the original video, the performances in Table A.1 are acceptable; further, they can be greatly improved with software optimization and a desktop computer. These devices may also support the processing of high-resolution videos, provided that memory requirements are satisfied. The maximum vector length that MatLab can manage in double precision is 2^{48} (about 281.5 trillion pixels), and an array of size 2^{35} requires 32 Gb RAM. It follows that recent laptops can manage timelapse videos with 100 frames in full HD (i.e. 2^{27}). However, as shown in the paper, spatial data are characterized by high anisotropy, which recommends restricting the analysis to homogeneous areas and periods. A viable way of fast computing may be segmenting the frames (e.g. in urban and rural areas) and applying separate STAR models to each group.

Appendix B. Supplementary material

Video data (in MP4 format) and computer codes (in MatLab language) can be found online at the link https://doi.org/10.1016/j.jag.2025.104451.

Data availability

Raw data (a MP4 video) are available at: https://svs.gsfc.nasa.gov/99. Old computer codes (a set of scripts in MATLAB language, R2015a or newer) were available at: https://figshare.com/s/0f849e8d2274da5fea9e. A faster version is provided as supplementary material, together with the dataset in MatLab format.

References

Acevedo, W., Masuoka, P., 1997. Time-series animation techniques for visualizing urban growth. Comput. Geosci. 23 (4), 423–435.
Ali, M.M., 1979. Analysis of stationary spatial–temporal processes: estimation and prediction. Biometrika 66, 513–518.
Batty, M., 2013. The New Science of Cities. MIT Press.
Bayer, D.M., Bayer, F.M., Gamba, P., 2021. A 3-D Spatiotemporal model for remote sensing data cubes. IEEE Trans. Geosci. Remote Sens. 59 (2), 1082–1093.
Bhattacharyya, B.B., Richardson, G.D., Franklin, L.A., 1997. Asymptotic inference for near unit roots in spatial autoregression. Ann. Stat. 28, 173–179.
Bustos, O., Ojeda, S., Vallejos, R., 2009. Spatial ARMA models and its applications to image filtering. Braz. J. Probab. Stat. 23 (2), 141–165.
Capoani, L., 2023. Review of the gravity model: origins and critical analysis of its theoretical development. SN Bus. Econ. 3 (95).
Deep, S., Saklani, A., 2014. Urban sprawl modeling using cellular automata. Egypt. J. Remote. Sens. Space Sci. 17 (2), 179–187.
Dentinho, T.P., Reid, N., 2021. Urban growth models: An application to American cities. Land Use Policy 111, 105160.
Falah, N., Karimi, A., Harandi, A.T., 2020. Urban growth modeling using cellular automata model and AHP (case study: Qazvin city). Model. Earth Syst. Environ. 6, 235–248.
Foresman, T.W., Masuoka, P., 1996. Baltimore-Washington: 200 Years of Urban Growth. Goddard Space Flight Center Scientific Visualization Studio, NASA, Available at https://svs.gsfc.nasa.gov/99.
Foresman, T.W., Pickett, S.T.A., Zipperer, W.C., 1997. Methods for spatial and temporal land use and land cover assessment for urban ecosystems and application in the greater Baltimore-Chesapeake region. Urban Ecosyst. 1, 201–216.
Google Earth, 2023. Iquitos Peru - Earth timelapse. Available at https://www.youtube.com/watch?v=ZHhByopdLY4.
Grillenzoni, C., 2004. Adaptive spatio-temporal models for satellite ecological data. J. Agric. Biol. Env. Stat. 9, 158–180.

Guerra, R.R., Bayer, F.M., Gamba, P., 2023. A 3D dynamical hybrid model: Coupling statistical and machine learning techniques to SAR time series prediction. In: IGARSS 2023: IEEE International Geoscience and Remote Sensing Symposium. Pasadena (CA), pp. 8342–8345.

Jaoude, G.A., Mumm, O., Karch, A., Carlow, V.M., 2022. Understanding land take in small and medium-sized cities through scenarios of shrinkage and growth using autoregressive models. Front. Built Environ. 8.

Kaviari, F., Saadi Mesgari, M., Seidi, E., Motieyan, H., 2019. Simulation of urban growth using agent-based modeling and game theory with different temporal resolutions. Cities 95, 102387.

Kilian, L., Lütkepohl, H., 2017. Structural Vector Autoregressive Analysis. Cambridge University Press.

Mojiri, A., Waghei, Y., Nili-Sani, H.R., Mohtashami Borzadaran, H.R., 2018. The stationary regions for the parameter space of unilateral second-order spatial AR model. Random Oper. Stoch. Equ. 26 (3), 185–191.

Mukhaiyar, U., Nabilah, F.T., Pasaribu, U.S., Huda, N.M., 2022. The space–time autoregressive modelling for the number of vehicles. J. Phys.: Conf. Ser. 2243, 012068.

Niu, X.-F., McKeague, I.W., Elsner, J.B., 2003. Seasonal space–time models for climate systems. Stat. Inference Stoch. Process. 6, 111–133.

Qiu, B., Zhou, M., Qiu, Y., Liu, S., Ou, G., Ma, C., Tu, J., Li, S., 2022. An integrated spatial autoregressive model for analyzing and simulating urban spatial growth in a garden city, China. Int. J. Env. Res. Public Heal. 19 (18), 11732.

Ribeiro, F.L., Rybski, D., 2023. Mathematical models to explain the origin of urban scaling laws. Phys. Rep. 1012, 1–39.

Shafizadeh-Moghadam, H., 2019. Improving spatial accuracy of urban growth simulation models using ensemble forecasting approaches. Comput. Environ. Urban Syst. 76, 91–100.

Sun, Q., Wang, S., Zhang, K., Ma, F., Guo, X., Li, T., 2019. Spatial pattern of urban system based on gravity model and whole network analysis in eight urban agglomerations of China. Math. Probl. Eng. 6509726.

Xu, Q., Zhu, A-X., Liu, J., 2023. Land-use change modeling with cellular automata using land natural evolution unit. CATENA 224, 106998.

Young, P.C., 2011. Recursive Estimation and Time-Series Analysis. Springer, Berlin.

Zhang, B., Li, F., Sang, H., Cressie, N., 2022. Inferring changes in Arctic sea ice through a spatio-temporal logistic autoregression fitted to remote-sensing data. Remore Sens. 14 (23), 5995.