DOI 10.1007/s10115-002-0094-1

Knowledge and Information Systems (2003) 5: 368–386

Knowledge and Information Systems

Modeling Spatial-Temporal Data with a Short Observation History

Dragoljub Pokrajac[1], Reed L. Hoskinson[2] and Zoran Obradovic[1]

[1]Center for Information Science and Technology, Temple University, Philadelphia, PA, USA
[2]Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID, USA

Abstract. A novel method is proposed for forecasting spatial-temporal data with a short observation history sampled on a uniform grid. The method is based on spatial-temporal autoregressive modeling where the predictions of the response at the subsequent temporal layer are obtained using the response values from a recent history in a spatial neighborhood of each sampling point. Several modeling aspects such as covariance structure and sampling, as well as identification, model estimation and forecasting issues, are discussed. Extensive experimental evaluation is performed on synthetic and real-life data. The proposed forecasting models were shown capable of providing a near optimal prediction accuracy on simulated stationary spatial-temporal data in the presence of additive noise and a correlated model error. Results on a spatial-temporal agricultural dataset indicate that the proposed methods can provide useful prediction on complex real-life data with a short observation history.

Keywords: Autoregressive models; Forecasting; Short observation history; Spatial-temporal data; Temporal and spatial database processing

1. Introduction

Emerging interest in spatial-temporal machine learning and data-mining (Roddick et al., 2001) is a consequence of advances in data acquisition, retrieval, and knowledge discovery technology along with an increased demand for applications in areas such as remote sensing (Kafatos, 1999), precision agriculture (Robert, 1999), and medical imagery (Megalooikonomou et al., 2000). Given historic spatial observations, in spatial-temporal prediction the goal is to find techniques to predict future values of certain attributes or the response (Cressie and Majure, 1997; Pokrajac and Obradovic, 2001a). However, due to specific properties of spatial-temporal datasets, numerous other issues involving prediction techniques arise.

Received 11 May 2001
Revised 26 Apr 2002
Accepted 5 June 2002

In addition to data domains characterized by long observation history (e.g. daily values of meteorological and oceanographic data exist for a hundred years or more, Jones et al., 1986), spatial-temporal prediction should often be performed on datasets that contain a relatively small number of temporal layers, as is usually the case in precision agriculture and medical imagery. In these domains, in addition to forecasting future attribute values, spatial-temporal prediction can be applied for data compression. Namely, it may be possible to predict certain attribute values based on their historical values and to store the attribute values only on locations where predicted and true values significantly differ (Pokrajac et al., 2002, in press). Similarly, instead of taking samples at each location and at each time instance, it might be possible to reduce spatial and/or temporal sampling resolution, and to approximate missing sample values by suitable interpolation/extrapolation of available data. To properly address these and other emerging problems, it is necessary to establish appropriate methods for attribute prediction on spatial-temporal data based on the exploitation of attribute correlation in space and time. Due to the strong presence of spatial correlation in data, conventional time-series models (Box et al., 1994) cannot provide desirable accuracy for spatial-temporal prediction. As alternatives, methods founded on geostatistical and state-space approaches are emerging.

Following geostatistical approach (Chilès and Delfiner, 1999), the spatial-temporal prediction can be performed through *kriging* – an estimation procedure based on a generalized least squares algorithm where an estimated attribute value is a linear combination of available attribute values. Originally, kriging was successfully applied for spatial interpolation (Denman and Freeland, 1985; Whelan et al., 1996; Kerry and Hawick, 1998). Recently, there have been numerous efforts to apply the principles of kriging to spatial-temporal domains. The effects of temporal and spatial correlation on attributes can be considered separately (Carrat and Valleron, 1992; Posa, 1995; Campling et al., 2001), but this method lacks firm theoretical foundation (Rouhani and Myers, 1990). An alternative is the spatial-temporal kriging where spatial and temporal dependences are modeled simultaneously (Chilès and Delfiner, 1999; Olea, 1999). The crucial problem here is estimation of valid spatial-temporal attribute statistics when disparity exists in the number of available spatial samples and the length of temporal history (Rouhani and Myers, 1990; Buxton and Pate, 1999). In addition, with a large number of spatial samples, kriging may become computationally prohibitive (Kerry and Hawick, 1998) and, when the number of temporal layers is small, can result in ill-conditioned linear systems (Rouhani and Myers, 1990). Hence, spatial-temporal kriging and its variants (Addink and Stein, 1999) are not applicable for forecasting in domains with a large amount of spatial samples but a comparatively short observation history.

Modeling based on the state space paradigm (Harvey, 1989; Brown and Hwang, 1993) is a powerful way of implementing a probabilistic framework to spatial-temporal processes, particularly applicable for domains such as meteorology, where data incrementally arrives at the prediction system. In space-time Kalman filtering (Wikle and Cressie, 1999), the observed variable consists of a spatial-temporal process and a spatially non-correlated measurement error. The spatial-temporal component is considered dependent on its values at the *immediate* previous time instance where the involved model coefficients vary in *space*. Attribute value at the consecutive time instance is predicted using a complicated recursive procedure for computing the optimal current estimate of the state vector based on available historical information (Harvey, 1989). Similar models have been applied by Stroud et al. (1999) for interpolation on meteorological and oceanographic data. The application of the state-space approach combined with stochastic simulations is demonstrated by Wikle et al. (2001) for parameter optimization in models with a complex structure ($\sim 10^5$ parameters!) on data with a long

temporal history. A further interesting approach is to combine the Kalman filtering and geostatistical modeling (Huang and Cressie, 1996; Mardia et al., 1998).

These and other techniques based on the state-space models are applicable for both uniform and non-uniform grids. They have an advantage over batch prediction models (such as those based on the ordinary least squares principle, Neter et al., 1985), assuming the availability of proper prior knowledge about the modeled process as well as a large number of temporal layers, so that the recursive nature of the models can be fully exploited (Brown and Hwang, 1993; Haykin, 1996). However, in applications of specific interest to us (precision agriculture and medical imaging), statistical parameters of the state-space model (correlation matrices of random shocks and measurement errors) are seldom known. In addition, Kalman filter misadjustment (consequence of poor correlation matrix initialization) can hardly be compensated due to a small number of temporal layers and the transient state of the filter will actually be longer than the available temporal span of the data. Hence, state-space models applied for prediction on spatial-temporal data with a short observation history are not likely to provide satisfactory results.

Alternative methods to geostatistical and state-space approaches involve prediction using spatial-temporal autoregressive models and its derivative **s**patial-**t**emporal auto-regressive models on a **u**niform **g**rid (STUG), initially introduced by Bennett (1979) without a proper theoretical underpinning. A special case of STUG, with a non-symmetric neighborhood structure (response at a spatial point dependent on response values only from a few pre-specified neighboring locations) was proposed by Kokaram and Godsill (1996). Similar to state-space models proposed by Wikle and Cressie (1999), STUG models assume spatial-temporal dependence in a uniform grid through a convolution (Smirnov, 1999) of filter coefficients and the response values at previous time instances. However, the STUG model considers only a *discrete* filtering process and filter parameters are *constant* in space. In addition, the model does *not* assume the existence of the prior knowledge about process correlations. Compared to geostatistical models that are essentially based on the moving average principle (Lindkvist and Lindqvist, 1997), shorter temporal history may be sufficient for a proper estimation of STUG – an autoregressive model that in contrast has a higher but still moderate number of parameters.

In this paper, we introduce a spatial-temporal autoregressive modeling on a uniform grid for phenomena with a short temporal history but a large number of spatial samples per temporal instance. In contrast to previous variants of STUG models, here an attribute depends on its values from *all* neighboring locations on rectangular lattices within a pre-specified distance. Moreover, unlike *ad hoc* attempts considered in prior studies (Bennett, 1979; Kokaram and Godsill, 1996), here we provide a comprehensive theoretical foundation of the STUG model. In addition to the model verification by experiments on real-life data (from the precision agriculture domain), here, by using synthetic datasets, we extensively discuss the influence of various data characteristics (e.g. random shocks correlation, the presence of measurement errors) on forecasting accuracy.

In Section 2, a theoretical foundation is provided for spatial-temporal autoregressive models on a uniform grid, including identification, coefficient estimation, and forecasting for different sampling techniques. In Section 3, results of an extensive experimental model evaluation for prediction on synthetic spatial-temporal data are presented. This is followed by reporting results on real-life agricultural data with a short observation history in Section 4 and a discussion in Section 5.

2. Methodology

In this section, we provide theoretical foundations for the STUG model based on statistical and image-processing techniques. Following the Box–Jenkins framework (Box et al., 1994), for three different spatial-temporal sampling strategies we discuss model identification, parameter estimation, and forecasting future response values. Derivations, including a multidimensional generalization and stationarity criteria, are provided in Pokrajac and Obradovic (2001b).

2.1. Model Specification

We define the STUG model on a uniform *two-dimensional* rectangular grid determined by sampling distances Δ in both spatial directions. Samples are taken at time instances $k\tau$ where τ is a time sampling interval and $k \in \{1, 2, \ldots\}$. For a temporal data layer corresponding to a specific time instance $t \cdot \tau$, the value of the random process $f_t(m, n)$ at each spatial location $(m\Delta, n\Delta)$, where m, n are integers, depends on samples taken in recent history from the same location and its spatial neighborhood. The history corresponds to p previous temporal layers sampled at time instances $(t-p)\tau, \ldots, (t-1)\tau$. We specify the spatial neighborhood as a $2L\Delta \times 2L\Delta$ square bounded by locations $((m \pm L)\Delta, (n \pm L)\Delta)$. Parameters p and L will be referred to as *temporal* and *spatial* orders of a STUG (p, L) model (for $p = 2$ and $L = 3$, STUG (2,3) model is illustrated in Fig. 1).

Symbolically, we express the value of a STUG (p, L) process $f_t(m, n)$ as

$$f_t(m,n) = \sum_{j=1}^{p} \sum_{k=-L}^{L} \sum_{l=-L}^{L} f_{t-j}(m-k, n-l) \cdot \phi_j(k,l) + a_{STUG,t}(m,n) \qquad (1)$$

Here, $a_{STUG,t}(m, n)$ is referred to as the error term and is defined as

$$a_{STUG,t}(m,n) = \sum_{k=-L}^{L} \sum_{l=-L}^{L} a_t(m-k, n-l) \cdot \phi_0(k,l) \qquad (2)$$

where $a_t(m, n)$ are spatially and temporary uncorrelated zero-mean Gaussian random 'shocks' with variance σ_a^2. Due to this assumption, STUG processes have zero means.

In Eqs (1) and (2), $\phi_j(k, l)$ denote model coefficients that can be represented as following coefficients matrices Φ_j, $j = 0, \ldots, p$:

$$\Phi_j = \begin{bmatrix} \phi_j(-L,-L) & \ldots & \phi_j(-L,L) \\ \ldots & \ldots & \ldots \\ \phi_j(L,-L) & \ldots & \phi_j(L,L) \end{bmatrix} \qquad (3)$$

Stationarity of a STUG process (spatial and temporal invariance of a distribution (Harvey, 1981; Box et al., 1994)) is determined by particular values of the model coefficients $\phi_j(k, l)$. Stationarity criteria can be derived from the theory of three-dimensional recursive filters (Smirnov, 1999).

Covariance structure. Formulas for variance σ_f^2 of a STUG process f and the variance $\sigma_{a_{STUG}}^2$ of a model error $a_{STUG,t}(m, n)$ can be obtained using the theory of multidimensional filtering. For two sets of samples from a stationary STUG process f, separated by m and n sampling intervals in spatial directions and by a temporal shift of k sampling intervals τ, auto-covariance coefficients $c_{FF,k}(m, n)$ defined in Bennett (1979) are related to each other through a *Yule–Walker system* defined in Pokrajac and Obradovic (2001b).

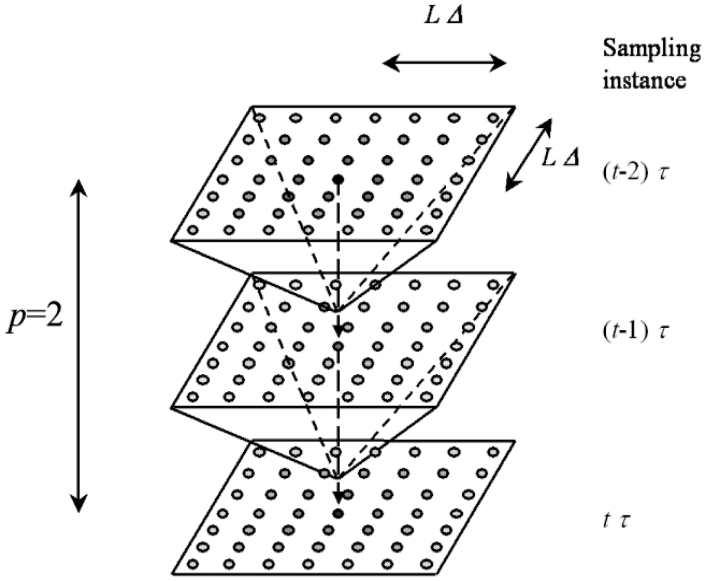


Fig. 1. Response dependence on neighboring samples in a recent history for a spatial-temporal model on a uniform grid with spatial and temporal orders $p = 2$, $L = 3$ (STUG(2, 3)).

In addition, autocorrelations $\rho_{FF,k}(m, n)$ are defined as quotients of auto-covariance coefficients $c_{FF,k}(m, n)$ and the variance σ_f^2 (Bennett, 1979).

2.2. Identification

To determine the temporal order p and the spatial order L of a STUG process f, autocorrelation and partial autocorrelation functions could be used as suggested by Bennett (1979). This approach is based on non-spatial time-series identification (e.g. Box et al., 1994) and requires a lot of time, advanced user skill, and cannot easily be made automatic when employed on real-life data (Harvey, 1981). An alternative is to apply model selection criteria based on summary statistics of residuals computed from a fitted model (Wei, 1994).

Using the Akaike information criterion (AIC) (e.g. Shibata, 1989) identified spatial and temporal orders of the observed process are chosen as (p, L) pair that minimize $AIC(p, L)$ value, defined for a STUG (p, L) model as

$$AIC(p, L) = \tfrac{1}{2} N \log(2\pi(1 - \hat{R}^2)) + p(2L + 1)^2 \tag{4}$$

where N is the total number of samples on which the model prediction and estimation are performed, and

$$\hat{R}^2 = 1 - \frac{\hat{\sigma}^2_{a_{STUG}}}{\hat{\sigma}^2_f} \tag{5}$$

is the estimated coefficient of determination (Neter et al., 1985).

Since the AIC criterion can overestimate the order of autoregressive processes (Wei, 1994) and does not provide asymptotically the true model orders when the sample size N approaches infinity (Rissanen, 1989), we also consider minimum-description

length (MDL) criterion (Rissanen, 1989). Here, p and L are chosen to minimize the $MDL(p, L)$ value defined for a STUG model as

$$MDL(p, L) = N \log(2\pi(1 - \hat{R}^2)) + p(2L + 1)^2 \log(N) \tag{6}$$

Other residual-based selection criteria, such as the Bayesian Akaike information criterion (Wei, 1994) and PHI criterion (Pukkila and Krishnaiah, 1988) can also be adapted for identification of STUG models. The comparison of these and other selection methods (McQuarrie and Tsai, 1998; Ljung, 1999), as applied for spatial-temporal process identification, will be the subject of a separate study.

2.3. Parameter Estimation

To estimate parameters of a STUG (p, L) model, we assume that samples are available from up to N_t successive temporal layers. Each temporal layer consists of $N_x \times N_y$ samples taken on a uniform $\Delta \times \Delta$ spatial grid that covers a spatial area of size $(N_x - 1)\Delta \times (N_y - 1)\Delta$. However, the actual number of available examples and the choice of estimation technique depend on the applied sampling strategy (Fig. 2).

In *ordinary sampling*, available data for model estimation consist of $N_x \times N_y \times N_t$ examples $f_t(m, n)$ where $t = 1, \ldots, N_t$; $m = 0, \ldots, N_x - 1$; $n = 0, \ldots, N_y - 1$ (see Fig. 2a). This sampling strategy is among the most common but it results in a high number of examples that may be prohibitive in a real-world application. As an alternative that reduces the number of examples to a half, we suggest sampling schemes with temporal and spatial interlacing. In *temporal interlacing*, instead of taking samples from all N_t temporal layers, even-index temporal layers are skipped (see Fig. 2b). However, in *spatial interlacing*, the spatial sampling grid is at a twice-lower resolution in m and n directions and adjacent temporal layers are complementary with samples taken at spatial-temporal locations where the sum of indexes m, n, and t has constant parity (see Fig. 2c). Observe that the number of samples can also be reduced by an ordinary sampling with a smaller spatial resolution. For a sampling interval 2Δ this is illustrated in Fig. 2(d). Ordinary sampling with the reduced sampling resolution will not be discussed in the rest of this study since in this case modeling is the same as with data ordinarily sampled on a sampling interval Δ.

Model parameters are estimated only for non-boundary samples $f_t(m, n)$ where *all* relevant neighbors (within a $2L\Delta \times 2L\Delta$ square neighborhood and from up to p temporal layers in the past) exist among available examples. Generally, parameter estimation is possible only if the set of non-boundary samples is not empty, i.e. when $N_x, N_y > 2L$ and $N_t > p$.

2.3.1. Ordinary Sampled Data

Model parameters on ordinary sampled data can be estimated using the Yule–Walker method (Box et al., 1994) and least squares (LS) estimation (Neter et al., 1985).

In the Yule–Walker method, parameter values are obtained by estimating process covariances $c_{FF,k}(m, n)$ (by averaging products $f_{k'-k}(m' - m, n' - n) \cdot f_{k'}(m', n')$ for all values of m', n' and k' where both corresponding samples are available) and solving a system of Yule–Walker equations.

We consider two LS techniques to provide consistent estimation (Wei, 1994) of STUG model parameters. The *averaged LS* estimates STUG model coefficients for each non-boundary location and averages these location-wise estimates. In contrast, in a *non-averaged LS* an estimate is computed directly from a single regression system.

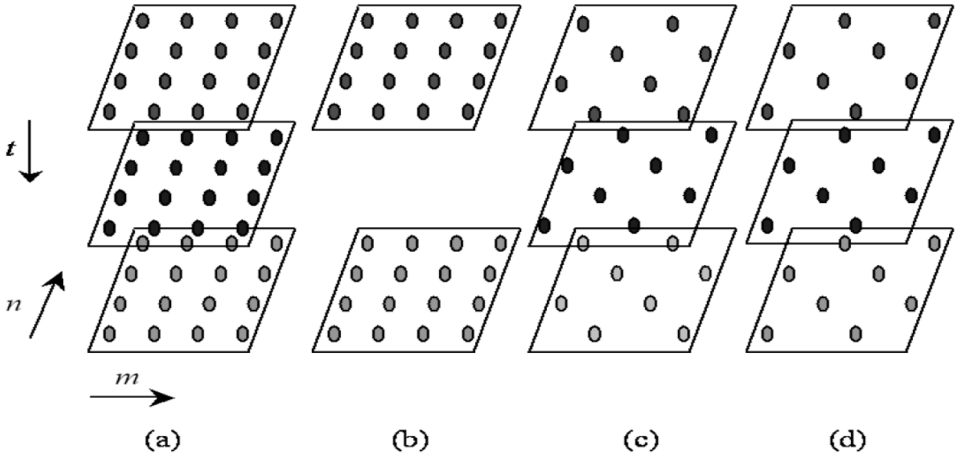


Fig. 2. An illustration of various sampling techniques for spatial-temporal data on a uniform grid; (a) Ordinary sampling (with no change in sampling resolution). (b) Sampling with temporal interlacing. (c) Sampling with spatial interlacing. (d) Ordinary sampling with the reduced resolution.

The Yule–Walker estimation is doable on data with shorter observation history in comparison to methods based on least-square estimation (the minimal number of temporal layers $N_{t,min}$ necessary to estimate the STUG model using the Yule–Walker, averaged and non-averaged LS methods, is $p+1$, $p+(2L+1)^2p$ and $p+\lceil(2L+1)^2p/((N_x-2L)\cdot(N_y-2L))\rceil$, respectively, (Pokrajac and Obradovic, 2001b)). At the same time the Yule–Walker estimation is asymptotically faster compared to averaged and non-averaged LS estimation with $O(p^2L^6+pL^2N_xN_yN_t)$ vs. $O(p^3L^6N_xN_y+p^2L^4N_xN_yN_t)$ and $O(p^3L^6+p^2L^4N_xN_yN_t)$ costs.

2.3.2. Temporal Interlacing

When estimated on data with temporal interlacing, model (1) on processes with temporal order $p=1$ can be represented as a constrained STUG $(1, 2L)$ model:

$$f_t(m,n)=\sum_{k=-2L}^{2L}\sum_{l=-2L}^{2L} f_{t-2}(m-k,n-l)\cdot\phi_1^{\#}(k,l)+a_t^{\#}(m,n) \tag{7}$$

where

$$\phi_1^{\#}(k,l)=\sum_{k'=-L}^{L}\sum_{l'=-L}^{L}\phi_1(k-k',l-l')\cdot\phi_1(k',l') \tag{8}$$

$$a_t^{\#}(m,n)=\sum_{k'=-L}^{L}\sum_{l'=-L}^{L}\phi_0(k',l')\cdot a_{STUG,t-1}(m-k',n-l')+a_{STUG,t}(m,n) \tag{9}$$

Non-linear parameter estimation of the model (7)–(9) using Gauss–Newton regression is discussed in Pokrajac and Obradovic (2001b). Particularly, the variance of equivalent model errors $a_t^{\#}(m,n)$ is shown to be larger than the variance of the original STUG model error term $a_{STUG,t}(m,n)$. Therefore, models estimated on data with temporal

interlacing can have lower prediction accuracy as compared to the direct modeling using Eq. (1) on ordinary sampled data.

2.3.3. Spatial Interlacing

When parameters are estimated on spatially interlaced data, only samples $f_t(m,n)$ where the sum of spatial and temporal indices m, n and t have a constant (even) parity are available for model estimation. Hence the model (1) becomes

$$f_t(m,n) = \sum_{\substack{j=1,\dots,p \\ -L' \le k,l \le L' \\ -L' \le k+l \le L' \\ j+k+l \text{ even}}} \sum \sum f_{t-j}(m-k,n-l) \cdot \phi_j(k,l) + a_t^*(m,n) \tag{10}$$

where a current sample value is considered as influenced by samples within a Manhattan distance $L'\Delta$ from the sample position, and the error term $a_t^*(m,n)$ incorporates unavailable samples. The model (10) can be easily estimated using a non-averaged least-squares optimization. However, if model coefficients do not satisfy the condition

$$\phi_j(k,l) = 0, \text{ for } j+k+l \text{ odd} \tag{11}$$

the model estimated on spatially interlaced data using Eq. (10) may have prediction accuracy significantly smaller than the accuracy obtained by the model (1) estimated on ordinary sampled data.

2.4. Forecasting

For ordinary sampling and temporal interlacing, the forecasting on the subsequent temporal layer $N_t + 1$ can be performed as

$$\hat{f}_{N_t+1}(m,n) = \sum_{j=1}^{p} \sum_{k=-L}^{L} \sum_{l=-L}^{L} f_{N_t+1-j}(m-k,n-l) \cdot \hat{\phi}_j(k,l) \tag{12}$$

on locations that satisfy $m = L, \dots, N_x - L; n = L, \dots, N_y - L$. With spatially interlaced samples, forecasting is performed as

$$\hat{f}_{N_t+1}(m,n) = \sum_{\substack{j=1,\dots,p \\ -L' \le k,l \le L' \\ -L' \le k+l \le L' \\ j+k+l \text{ even}}} \sum \sum f_{N_t+1-j}(m-k,n-l) \cdot \hat{\phi}_j(k,l) \tag{13}$$

on the non-boundary locations where the sum $m + n + N_t + 1$ of indexes is even. Analogous to non-spatial time-series (Box et al., 1994), the estimated prediction error $\hat{\sigma}^2_{a_{STUG}}$ and process variance $\hat{\sigma}^2_{\hat{f}}$ can be computed by averaging point-wise errors $(\hat{f}_{N_t+1}(m,n) - f_{N_t+1}(m,n))^2$ and squared responses $f_{N_t+1}(m,n)^2$ through all locations at the time instance $N_t + 1$ where the forecasting is performed.

Forecasting accuracy of particular models on a specific dataset will be compared using the coefficient of determination R^2 estimated as Eq. (5), where the larger scores correspond to the more accurate prediction with 1 corresponding to a perfect and 0 to a

trivial mean predictor. Considered prediction models do not contain an intercept, so R^2 can be negative (Davidson and MacKinnon, 1993). In the experiments with synthetic data (Section 3) when *true* values of model coefficients $\phi_j(k,l)$ are known, the coefficient of determination will also be compared to its theoretical maximum $R^2_{theoretical}$. This optimal value can be computed by using Eq. (5), where the estimated variances of a STUG series and the estimated model error are replaced by their theoretical values calculated as functions of *known* model coefficients $\phi_j(k,l)$.

3. Experimental Results on Simulated STUG Series

Here we present the main results of experiments with simulated STUG series (see Pokrajac and Obradovic, 2002, for more detailed results).

3.1. Experiment Set-Up

Experimental data consisted of $N_t + 1$ temporal layers, each with $N_x \times N_y$ spatial points on a uniform grid. Models were estimated using data from the first N_t temporal data layers. Prediction accuracy was measured by the coefficient of determination ($\hat{R}^2$), estimated using Eq. (5) on the $N_t + 1$th temporal layer. Unless specified otherwise, each experiment was repeated 10 times (each time on other simulated data that followed the specifications) and the estimated mean and standard deviation of $\hat{R}^2$ from these repetitions were reported in the form $mean(\hat{R}^2) \pm std(\hat{R}^2)$.

To explore properties of the proposed methods on data collected using ordinary sampling and temporal and spatial interlacing, we simulated eight different STUG (p, L) series. Using coefficient matrices Φ_1, Φ_2 shown in Table 1, we varied properties of simulated spatial-temporal series. In addition, for each series, we independently varied the matrix Φ_0 that determines correlation properties of a STUG model error (Pokrajac and Obradovic, 2001b). Unless otherwise specified, presented results were obtained for spatially uncorrelated model errors, corresponding to

$$\phi_0(k,l) = \begin{cases} 1, & \text{if } k = l = 0 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

The stationarity of simulated STUG series was verified using the stability criteria (Pokrajac and Obradovic, 2001b). Simulated STUG series $S1$–$S3$ had spatial and temporal orders both equal to one. Series $S4$ and $S5$ were representatives of STUG $(p = 1, L = 2)$ series. With the same values of non-zero coefficients in Φ_1, the series $S4$ and $S5$ had a different correlation structure, which influenced the prediction accuracy achievable through the considered methods. The series $S6$ was an example of a STUG $(p = 1, L = 3)$ series with the coefficient matrix structure very similar to that of series S4. Unlike series $S1$–$S6$, in series $S7$ and $S8$ the current response values depended on the response from the two previous time layers $(p = 2, L = 1)$.

Initially, we compared three proposed methods – the Yule–Walker estimation, non-averaged least-squares, and averaged least-squares estimators – on ordinary sampled data. To determine the influence of the sample size on estimation quality, we gradually increased the size of a spatial grid specified by N_x, N_y for a series with a short time history, and compared prediction accuracy using models obtained by the three proposed methods. The accuracy was better and the standard deviation of accuracy in repeated experiments was smaller with larger spatial grid sizes. Model accuracy did not change significantly when N_x, N_y were greater than 30 and became practically constant with

Table 1. Coefficient matrices for simulated STUG series $S1$–$S8$

$S1$	$S2$	$S3$	$S4$	$S5$
$\Phi_1 \begin{bmatrix} .14 & -.21 & -.13 \\ -.22 & .4 & -.24 \\ .12 & .23 & -.11 \end{bmatrix}$	$\begin{bmatrix} 0 & -.1 & 0 \\ -.1 & .5 & -.1 \\ 0 & -.1 & 0 \end{bmatrix}$	$\begin{bmatrix} -.1 & .1 & -.1 \\ .1 & .15 & .1 \\ -.1 & .1 & -.1 \end{bmatrix}$	$\begin{bmatrix} .2 & 0 & 0 & 0 & -.2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -.2 & 0 & 0 & 0 & .2 \end{bmatrix}$	$\begin{bmatrix} 0 & .2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -.2 \\ 0 & 0 & .15 & 0 & 0 \\ 0 & 0 & 0 & 0 & .2 \\ 0 & 0 & 0 & -.2 & 0 \end{bmatrix}$

$S6$	$S7$	$S8$
$\Phi_1 \begin{bmatrix} .2 & 0 & 0 & 0 & 0 & 0 & -.2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .15 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.2 & 0 & 0 & 0 & 0 & 0 & .2 \end{bmatrix}$	$\Phi_1 \begin{bmatrix} .1 & 0 & .1 \\ 0 & .2 & 0 \\ .1 & 0 & .1 \end{bmatrix}$ $\Phi_2 \begin{bmatrix} -.2 & 0 & -.2 \\ 0 & 0 & 0 \\ -.2 & 0 & -.2 \end{bmatrix}$	$\begin{bmatrix} .1 & 0 & .1 \\ 0 & .2 & 0 \\ .1 & 0 & .1 \end{bmatrix}$ $\begin{bmatrix} 0 & 0 & 0 \\ 0 & -.2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$N_x = N_y = 50$. Therefore, the subsequent experiments were performed only on a 50×50 spatial grid.

The accuracy of spatial-temporal models estimated by the *averaged* LS was consistently lower than the accuracy of *non-averaged* LS models that were estimated with significantly fewer examples. Since the Yule–Walker estimation provided comparable or better results than non-averaged LS with considerably smaller computational time requirements (see Section 2.3.1), results of further experiments on ordinary sampled data are shown only for the Yule–Walker method.

In the rest of this section each simulated series consisted of six temporal layers ($N_t = 5$), corresponding to a typical temporal span of real-life spatial-temporal data in agriculture (Hoskinson et al., 1999) (see also Section 4 of this paper).

3.2. Model Identification

Five identification experiments were repeated per series and the results were averaged. Each experiment consisted of simulation of the specified series and subsequent application of AIC (Eq. 4) and MDL (Eq. 6) criteria on simulated data. Both AIC and MDL selection criteria provided the 100% correct model identification when series $S4$–$S6$ were identified. Moreover, by the MDL criterion, series $S2$ and $S3$ were correctly identified in all five experiments, while the AIC criterion provided correct identification in 60% and 80% cases, respectively. Identification was difficult when correlations $|\rho_{FF,k}(m, n)|$ for $|m|, |n| > 1$ and time delay $k = 1$ were significant (e.g. on series $S1$ we achieved 20% correct identification using AIC and 40% with MDL criterion). Also, when identifying STUG ($p = 2, L = 1$) series ($S7$, $S8$) both selection criteria were conservative in providing temporal order ($p = 1$) smaller than the actual.

Influence of additive measurement error. In real-world applications, spatial-temporal data are frequently contaminated with noise, due to data acquisition or to the nature of the observed series. To investigate identification of STUG (1, 1) models in the presence of a measurement error, we repeated identification experiments with a controlled amount of additive uncorrelated Gaussian noise (Papoulis, 1991), with variance ranging from 10% to 100% of the model variance σ_f^2. Generally, the higher measurement error implied the less accurate model identification, especially if the identification accuracy in the *absence* of measurement errors was already low. With measurement error variance of 10% σ_f^2 series $S1$ was correctly identified in only 1 out of 10 attempts (both AIC and MDL experiments accounted), while with the error variance of 100% σ_f^2 neither criterion provided a correct identification.

Identification of series with spatially correlated model errors. Experimental results suggest that the spatial correlation of model errors, introduced by coefficients $\phi_0(k, l)$ different from values at Eq. (14), may have a negative effect on model identification. When errors were highly spatially correlated the selection criterion tended to recognize a higher temporal order p while underestimating the spatial order L. This can be explained by behavior of estimated autocorrelations $\rho_{FF,k}(m, n)$ for samples shifted in time ($k = 1, 2$).

3.3. Parameter Estimation and Forecasting

Experiments on parameter estimation and prediction accuracy were performed assuming the correct orders of estimated STUG series had been identified.

Ordinary sampling. When the Yule–Walker estimation was applied to ordinary sampled simulated STUG series $S1$–$S8$, we were always able to obtain a useful prediction model ($\hat{R}^2 > 0$) with small variation in prediction accuracy ($std(\hat{R}^2) \leq 0.02$), as reported at Table 2. The accuracy was close to the theoretical optimum obtained by computing $R^2_{theoretical}$.

An introduction of measurement error led to decrease in the prediction accuracy. The higher error resulted in the smaller accuracy. The average estimated coefficient of determination for series $S1$, which was 0.64 in the absence of noise (Table 2), decreased to 0.55 in the presence of 10% noise and was merely 0.20 when the variances of noise and of the STUG series were equal (100% noise) with standard deviation of estimated $\hat{R}^2$ smaller than 0.02. Similar results were obtained for series $S2$ and $S3$.

The presence of spatially correlated model error introduced by values of the coefficients $\phi_0(k, l)$ different from Eq. (14) decreased theoretical prediction accuracy $R^2_{theoretical}$. For such data properties, the estimated mean of prediction accuracy $\hat{R}^2$ decreased while the variance of prediction accuracy slightly increased, as anticipated from the identification results discussed in Section 3.2. Hence, although we were able to achieve $\hat{R}^2$ close to the theoretical optimum, the prediction accuracy in the presence of a high model error correlation may be unsatisfactory.

Temporal interlacing. Model estimation was performed using the Gauss–Newton regression with 100 iterations on series $S1$–$S8$. Ten randomized realizations were generated for each simulated spatial-temporal series, and for each realization the model estimation and the model evaluation were repeated 10 times due to a stochastic character of the estimation technique, giving a total of 100 estimation experiments.

For series $S1$ and $S2$ useful predictors were obtained with prediction accuracies close to those obtained with ordinary sampled data, as shown in Table 2. However, using the proposed method on data with temporal interlacing, no useful prediction models could be estimated for series with a relatively small $R^2_{theoretical}$ ($S3$), or spatial orders $L > 1$ ($S4$–$S6$) where prediction was incorrect ($\hat{R}^2 < 0$), presumably due to an inadequate initialization of the applied iterative estimation procedure. In addition, since the estimation method is designed for series with temporal order $p = 1$, series $S7$ and $S8$ could not be properly forecasted.

When a non-correlated measurement error was present, the proposed method resulted in useful prediction models for STUG (1, 1) series on temporal interlaced data if models on ordinary sampled data had at least moderate prediction accuracy. Thus, for series $S1$, accuracy on temporal interlaced samples was close to or sometimes even better than the accuracy obtained using ordinary sampled data. On data with 10% measurement error variance, models on temporal interlaced data resulted in $\hat{R}^2$ of 0.53 ± 0.09, while with 100% measurement error variance we achieved $\hat{R}^2$ of 0.17 ± 0.04. Similar results were obtained for series $S2$. In contrast, on series $S3$–$S8$ where prediction accuracy was already low in the absence of the measurement error, no useful model on temporal interlaced data could be estimated when the measurement error existed.

The presence of spatially correlated model errors within temporal interlaced samples for the majority of examined series had similar effects as in the case of ordinary sampled data, where the forecasting accuracy was significantly reduced due to highly correlated model errors.

Spatial interlacing. On spatially interlaced data, we estimated model (10) using non-averaged least-squares (see Section 2.3.3). For all series, estimated models were useful for forecasting ($\hat{R}^2 > 0$, see Table 2). When the spatial or the temporal order of a

Table 2. Spatial-temporal prediction accuracy $\hat{R}^2$ on STUG series $S1$–$S8$ using ordinary sampling, temporal and spatial interlacing, and the optimal accuracy $R^2_{theoretical}$

Production accuracy	Series							
	$S1$	$S2$	$S3$	$S4$	$S5$	$S6$	$S7$	$S8$
Optimal	0.65	0.38	0.13	0.31	0.22	0.32	0.34	0.65
Ordinary sampling	0.64 ± 0.02	0.38 ± 0.02	0.11 ± 0.01	0.28 ± 0.02	0.21 ± 0.02	0.25 ± 0.01	0.34 ± 0.02	0.65 ± 0.01
Temporal interlacing	0.63 ± 0.01	0.38 ± 0.02				< 0		
Spatial interlacing	0.52 ± 0.03	0.21 ± 0.02	0.06 ± 0.04	0.10 ± 0.03	0.19 ± 0.03	0.05 ± 0.04	0.23 ± 0.02	0.57 ± 0.02

simulated series was larger than 1 (series $S4$–$S8$) models trained on spatially interlaced data outperformed those on temporal interlaced data, approaching in some cases the accuracy achieved on ordinary sampled data (Table 2).

In the presence of uncorrelated measurement errors or spatially correlated model errors, the models estimated on spatially interlaced data behaved similarly to those estimated on ordinary sampled data. When uncorrelated measurement errors were present, the prediction accuracy decreased proportional to the variance of the additive measurement error. In the presence of correlated model errors, the accuracy of the proposed model, Eq. (10), significantly dropped but the proposed estimation technique still provided useful forecasting models.

4. Experiments on Real-Life Data

The applicability of the proposed method on real-life datasets was evaluated on agricultural data obtained from Idaho National Engineering and Environmental Laboratory (INEEL) field research (Hoskinson et al., 1999; Pokrajac et al., 2002, in press). The dataset consisted of 12 soil attributes (concentrations of boron (B), copper (Cu), iron (Fe), potassium (K), manganese (Mn), nitrogen (N), phosphorus (P), sulfur (S), and zinc (Zn) and soil salinity (sa), cation exchange capacity (CEC), and pH value (pH) in the top 1 foot (30.5 cm) of soil) provided on a uniform spatial grid of $10.66 \times 10.66\,\mathrm{m}^2$ covering the total area of 60.04 ha, during four consecutive years (1995–1998). Within each year, soil attributes were sampled several times at time instances that corresponded to specific phases in the crop development (fertilization and planting, growing season, just prior to harvesting). In this paper, for each attribute we considered two spatial-temporal series, corresponding to the summer and fall seasons. For all attributes, an individual temporal data layer corresponded to data collected in a specified year and season and consisted of 84×65 examples.

Prior to the application of our methods to a particular attribute, the mean values were estimated for each temporal layer and subtracted from corresponding sample values, to obtain data that satisfy the zero-mean property of the proposed STUG model (see Section 2.1). Observe that a similar normalization procedure has been applied by Wikle and Cressie (1999). This was followed by estimation of spatial and temporal statistics of the normalized data. The parameters of estimated spatial variograms (Chilès and Delfiner, 1999) expressed significant non-stationarity in the observed period of four years, which prevented the application of geostatistical prediction methods (Posa, 1995). On the other side, in spite of significant values of computed temporal autocorrelations $\rho_{FF,k}(0,0)$, $k = 1, 2, 3$ for the same season samples of attributes collected in different years, we were not able to obtain satisfactory prediction results using non-spatial methods due to a small number of available temporal layers (insufficient temporal history) and non-stationarity of data.

For each attribute/season combination, STUG models were trained on samples from years 1995–1997 and tested on data from the consecutive year: 1998. As in Section 3, prediction accuracy was evaluated through estimated coefficient of determination $\hat{R}^2$ (Eq. 5). To directly compare the influence of the order selection, we chose to evaluate STUG models with various values of spatial and temporal orders, instead of performing model identification. For ordinary sampling and spatial interlacing we varied both temporal and spatial orders (such that $p = 1, 2$ and L (or L' for spatial interlacing) $= 0, 1, 2, 3$) while in the case of temporal interlacing we only estimated STUG models with temporal order $p = 1$ (for which specifically the estimation procedure in Section 2.3.2 has been developed). For ordinary sampling as well as for spatial interlacing, the application of models with $L = 0$ (corresponding

to non-spatial modeling) or $L' = 0$ led to accuracy similar to a trivial mean predictor. For temporal interlacing, modeling with $L = 0$ or $L > 1$ consistently failed to provide useful predictors.

When modeling was applied on ordinary sampled data, the proposed technique was capable of providing useful prediction models ($\hat{R}^2 > 0$) on a relatively small subset of observed attributes/seasons. Overall, the best performance was achieved using the STUG ($p = 1, L = 2$) model and the highest predictability out of all attributes was obtained for Mn in summer, with $R^2 = 0.24$. Except for pH, where acceptable prediction accuracy (with best R^2 in the range of 0.14–0.22) was obtained for summer and fall, the proposed model could not provide good prediction consistently for *both* seasons. Our hypothesis is that the predictability of attributes in different seasons was unstable due to the diverse properties (temporal and spatial correlation) of various unobserved factors that may have been involved in distinct phases of plant development. In addition, we should observe that in spite of the limited *overall* prediction quality, there were parts of the field where the proposed method did perform well (Pokrajac et al., 2002, in press), which indicates the spatial non-stationarity of the analyzed dataset and potential improvements that might be achievable by heterogeneous prediction models (Pokrajac et al., 1999).

The application of models estimated on temporal interlaced data could provide prediction with the coefficient of determination up to 0.32 (for pH in fall). Prediction using temporal interlaced data, in addition to this attribute/season combination, overperformed modeling on ordinary sampled data on salinity in both observed seasons (R^2 was 0.07), and on P in fall (R^2 of 0.28), while both models performed similarly on P and pH in summer (R^2 in the range of 0.14–0.15). Generally, predictions in the fall were better than in the summer. Also, temporal interlacing was able to provide useful prediction models (R^2 in the range of 0.14–0.28) for the attribute/season combinations (Fe and K in fall) where autocorrelation $\rho_{FF,k}(0, 0)$ *increases* with the temporal lag k.

Modeling on spatial interlaced data resulted in rather high prediction accuracy, where estimated R^2 ranged up to 0.99 (for S in summer). For both seasons, prediction was good for N (best R^2 was in the range of 0.80–0.87) and B (best R^2 0.51 and 0.84). For the summer season, besides S, high prediction accuracy was achieved for Fe and salinity (R^2 in the range of 0.67–0.93). For each attributes/season combination where modeling on spatial interlaced data provided a useful prediction, it also outperformed modeling on ordinary sampled data and modeling with temporal interlacing. We hypothesize that by using spatial interlacing we were able to exploit the spatial correlation of samples taken at the same time instance without the risk of considering many almost identical samples (as we would in the case of ordinary sampled data where two immediate neighbors have a high degree of similarity). Jagger et al. (2002) made similar hypotheses when concluding that their estimation procedure failed to converge due to high spatial correlations of neighborhood samples. For the majority of cases, there was no significant improvement of model performance due to adoption of higher spatial ($L' > 1$) and temporal ($p > 1$) orders. Our hypothesis is that here potential benefits due to the application of more complex models were offset by problems due to estimation of the higher number of parameters (caused by spatial non-stationarity of the data). However, we should point out that the best results for certain attributes were achieved with higher spatial or temporal orders ($p = 1$ and $L' = 3$, for B and Mn in fall; $p = 2, L' = 1$ for Fe in summer and CEC in fall), which can be an impetus for a follow-up analysis of these attributes by soil science specialists.

5. Conclusions

We have provided a theoretical foundation and experimental evaluation for prediction in spatial-temporal data with a short history using **s**patial-**t**emporal auto-regressive models on **u**niform **g**rid (STUG) where the response value is influenced by its pre-specified spatial and temporal neighborhood.

In the theoretical part of the study, we considered definition, model identification, and covariance structure of the proposed data model. In addition to ordinary sampling where all samples from the observed spatial-temporal grid are available, we investigated estimation and forecasting techniques on data sampled with the density reduced along either temporal or spatial directions.

The validity of the proposed estimation techniques was first evaluated on synthetic data that was generated to satisfy STUG model assumptions, with prediction accuracy measured through the average coefficient of determination (R^2). On ordinary sampled data, the techniques based on information theoretic criteria have been shown capable of providing acceptable model identification and we achieved prediction accuracy close to the theoretical maximum in spite of a relatively small number of temporal data layers available for model estimation. In the presence of spatially correlated model errors or measurement errors, the identification was more challenging but the proposed prediction methods were still able to approach the optimal performance. Sampling with spatial interlacing was more appropriate for synthetic data with higher spatial or temporal orders. In contrast, on data where the response at a sampling location depended only on its immediate neighbors in space and the immediate predecessors in time, models on temporal interlaced data were superior to those on spatial interlaced data and, in fact, these results approached the accuracy on ordinary sampled data.

To verify the applicability of the proposed method to real-life datasets with a short observation history, we performed forecasting experiments on agricultural spatial-temporal data consisting of 12 attributes observed in two seasons through four consecutive years. Although modeling on ordinary sampled data was capable of providing useful predictions only on a relatively small subset of observed attributes/seasons, significant accuracy improvements were obtained when sampling density was reduced by spatial interlacing. While we still do not have the complete explanation for such phenomena, our experimental results clearly indicate the appropriateness of the proposed STUG model for forecasting data with a short temporal span.

Our work in progress includes extensive empirical comparison of the proposed STUG methods with the alternative prediction techniques on spatial-temporal data with short observation history where theory suggests the advantage of the STUG modeling. Also, we emphasize the importance of additional experimentation on real-world datasets outside of agriculture, including medical data, 3-D image sequences and remote sensing data. Further theoretical development of the proposed techniques includes their adaptation for forecasting on distributed spatial-temporal datasets. In addition, supplementary research is necessary to develop novel model identification techniques and to simplify the stationarity criteria for spatial-temporal modeling explored in this study.

Acknowledgements. The authors are thankful to Dragan Obradovic, Aleksandar Lazarevic, and Yelena Quigley for discussing Kalman filters, providing useful comments and help with style polishing. In addition, the authors express their gratitude to the anonymous reviewers whose comments significantly improved the structure and overall quality of the paper. Work was in part supported by INEEL LDRD Program under DOE Idaho Operations Office Contract DE-AC07-99ID13727.

References

Addink EA, Stein A (1999) A comparison of conventional and geostatistical methods to replace clouded pixels in NOAA-AVHRR images. International Journal of Remote Sensing 20(5): 961–977

Bennett RJ (1979) Spatial time series. Pion, London

Box G, Jenkins G, Reinsel G (1994) Time series analysis. Prentice-Hall, Englewood Hills, NJ

Brown RG, Hwang PYC (1993) Introduction to random signals and applied Kalman filtering (with Matlab exercises and solutions) (3rd edn). Wiley, New York

Buxton BE, Pate AD (1999) Estimation of joint temporal-spatial semivariogram. In Proceedings of fifth international geostatistics congress, Wollongong, Australia, September 1996, pp 150–161

Campling P, Gobin A, Feyen J (2001) Temporal and spatial rainfall analysis across a humid tropical catchment. Hydrological Processes 15(3):359–375

Carrat F, Valleron A-J (1992) Epidemiologic mapping using the 'Kriging' method: application to an influenza-like illness epidemic in France. American Journal of Epidemiology 135:1293–1300

Chilès J, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Cressie N, Majure JJ (1997) Spatio-temporal statistical modeling of livestock waste in streams. Journal of Agricultural, Biological and Environmental Statistics 2(1):24–47

Davidson R, MacKinnon JG (1993) Estimation and inference in econometrics. Oxford University Press, New York

Denman KL, Freeland HJ (1985) Correlation scales, objective mapping and a statistical test of geostrophy over the continental shelf. Journal of Marine Research 43:517–539

Harvey AC (1981) Time series models. Wiley, New York

Harvey AC (1989) Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge, UK

Haykin S (1996) Adaptive filter theory (3rd edn). Prentice-Hall, Englewood Cliffs, NJ

Hoskinson RL, Hess JR, Alessi RS (1999) Temporal change in the spatial variability of soil nutrients. In Proceedings of 2nd European conference on precision agriculture, Odense, Denmark, pp 61–70

Huang H-C, Cressie N (1996) Spatio-temporal prediction of snow water equivalent using the Kalman filter. Computational Statistics and Data Analysis 22:159–175

Jagger T, Niu X-F, Elsner JB (2002) A space-time model for seasonal hurricane prediction. Inter- national Journal of Climatology 22:451– 465

Jones PD, Raper R, Bradley RS, et al. (1986) Northern hemisphere surface air temperature variations: 1851–1984. Journal of Climate and Applied Meteorology 25:161–179

Kafatos M (1999) Data access, querying, analysis and data mining in a distributed framework for earth system science support. In Proceedings of Geocomputation '99, Fredericksburg, VA, CD-ROM, paper 022

Kerry KE, Hawick KA (1998) Kriging interpolation on high-performance computers. In Proceedings of high-performance computing and networks (HPCN) Europe '98, Amsterdam, 1998, also available on http://acsys.adelaide.edu.au/reports/035/abs-035.html

Kokaram A, Godsill S (1996) A system for reconstruction of missing data in image sequences using sampled 3D AR models and MRF motion priors. In Proceedings of European conference on computer vision 1996, Springer Verlag, pp 613–624

Lindkvist L, Lindqvist S (1997) Spatial and temporal variability of nocturnal summer frost in elevated complex terrain. Agricultural and Forest Meteorology 87(2–3):139–153

Ljung L (1999) System identification: theory for the user (2nd edn). Prentice-Hall, Englewood Cliffs, NJ

Mardia KV, Goodall C, Redfern C, et al. (1998) The kriged Kalman filter. Test 7(2):217–252

McQuarrie ADR, Tsai C-L (1998) Regression and time series model selection. World Scientific, Singapore

Megalooikonomou V, Ford J, Li S, et al. (2000) Data mining in brain imaging. Statistical Methods in Medical Research 9(4):359–394

Neter J, Waserman W, Kuther M (1985) Applied linear statistical models (regression, analysis of variance, and experimental Designs) (2nd edn). Richard D. Irwin, Homewood, IL

Olea RA (1999) Geostatistics for engineers and earth scientists. Kluwer, Boston, MA

Papoulis A (1991) Probability, random variables, and stochastic processes (3rd edn). McGraw-Hill, New York

Pokrajac D, Fiez T, Obradovic D, et al. (1999) Distribution comparison for site-specific regression modeling in agriculture. In Proceedings of 1999 international joint conference on neural networks, Washington, DC, CD-ROM, No. 346, Section 10.9

Pokrajac D, Obradovic Z (2001a) Improved spatial-temporal forecasting through modeling of spatial residuals in recent history. In Proceedings of first international SIAM conference on data-mining, Chicago, 2001, CD-ROM, No. 9

Pokrajac D, Obradovic Z (2001b) Spatial-temporal autoregressive model on uniform sampling grid. Technical report CIS TR 2001-05, Computer and Information Sciences Department,

College of Science and Technology, Temple University, Philadelphia, PA, also available on http://www.ist.temple.edu/pokie/techreport2001.ps

Pokrajac D, Obradovic Z (2002) Experiments with spatial-temporal autoregressive models on uniform grid. Technical report CIS TR 2002-01, Computer and Information Sciences Department, College of Science and Technology, Temple University, Philadelphia, PA, also available on http://www.ist.temple.edu/pokie/techreport2002.ps

Pokrajac D, Hoskinson RL, Lazarevic A, et al. (in press) Spatial-temporal techniques for prediction and compression of soil fertility data. In Proceedings of 6th international conference on precision agriculture, Minneapolis, MN

Posa D (1995) A simple description of spatial-temporal processes. Computational statistics and Data Analysis 15(4):425–437

Pukkila TM, Krishnaiah PR (1988) On the use of autoregressive order determination criteria in univariate white noise tests. IEEE Transactions on ASSP 36(5):764–774

Rissanen J (1989) Stochastic complexity in statistical enquiry. Series in Computer Science 15, World Scientific, Singapore

Robert PC (ed) (1999) Proceedings of 4th international conference on precision agriculture. ASA-CSSA-SSSA, Madison, WI

Roddick JF, Hornsby K, Spiliopoulou M (2001) An updated bibliography of temporal, spatial and spatio-temporal data mining research. In Roddick JF, Hornsby K (eds). Lecture Notes in Artificial Intelligence 2007, Springer, Berlin, pp 147–164

Rouhani S, Myers DE (1990) Problems in space-time kriging of geohydrological data. Mathematical Geology 22(5):611–623

Shibata R (1989) Statistical aspects of model selection. In Willems JC (ed). From data to model. Springer, Berlin, pp 215–240

Smirnov A (1999) Processing of multidimensional signals. Springer, Berlin

Stroud J, Miller P, Sansó B (1999) Dynamic models for spatio-temporal data. Technical report working paper 99-20, Institute of Statistics and Decision Sciences, Duke Unversity, Durham, NC

Wei WW (1994) Time series analysis, univariate and multivariate methods. Addison-Wesley, Redwood City, CA

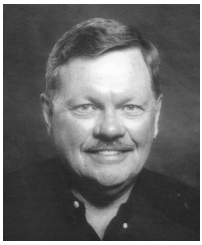
Whelan BM, McBratney AB, Viscarra Rossel RA (1996) Spatial prediction for precision agriculture. In Robert PC, Rust RH, Larson WE (eds). Proceedings of 3rd international conference on precision agriculture, ASA-CSSA-SSSA, Madison, WI, pp 331–342

Wikle CK, Cressie N (1999) A dimension-reduced approach to space-time Kalman filtering. Biometrika 86:815–829

Wikle CK, Milliff RF, Nychka D, et al. (2001) Spatio-temporal hierarchical Bayesian nodeling: tropical ocean surface winds. Journal of the American Statistical Association Applications and Case Studies 96(454):382–397

Author Biographies

Dragoljub Pokrajac received a BS degree in Electrical engineering and an MS degree in Telecommunications from the University of Nis, Serbia and Montenegro, in 1993 and 1997, respectively. From 1998 to 2000, he attended a PhD program in Computer Science at Washington State University, Pullman, WA. He received a PhD degree at Temple University, Philadelphia, PA, in 2002. He is currently an adjunct professor at Temple University. Dr Pokrajac has authored or co-authored one book, six journal papers and numerous conference papers and is a co-author of one patent application. His research interests include spatial-temporal data mining in brain imaging, remote sensing, and precision agriculture, as well as machine learning, statistics and neural networks.



Reed Hoskinson received a bachelor's degree with majors in Physics and Mathematics from the University of Iowa, and a master's and PhD in Ecology from the University of Minnesota, and has worked at the Idaho National Engineering and Environmental Laboratory (INEEL) for 18 years. Reed has worked in scientific programming for over 35 years, including modeling of agricultural ecosystems and statistical analyses of ecological data, and development of major software systems. He was one of the founders of the INEEL Spatial Analysis Laboratory for GIS applications, and later was the manager of the laboratory. Reed has been a Principal Investigator for INEEL precision agriculture research for the past 10 years. He has over 35 agriculture-related publications, and eight agriculture-related patents and patent applications.

Zoran Obradovic is the Director at the Center for Information Science and Technology and a Professor of Computer and Information Sciences at Temple University, Philadelphia. His research interests focus on solving challenging bioinformatics, biomedical, geoinformation sciences and computational finance problems by developing and integrating data mining and statistical learning technology for an efficient knowledge discovery at large databases. Funded by NSF, NIH, DOE and industry, during the last decade he has contributed to about 130 refereed articles on these and related topics and to several academic and commercial software systems. He is/was an editorial board member at the *Multiple Valued Logic*, *Journal of Computational Intelligence in Finance* and the *IEEE Transactions on Education* and has organized several journal special issues and workshops. For more details see www.ist.temple.edu/∼zoran.

Correspondence and offprint requests to: Dragoljub Pokrajac, Delaware State University, Computer Science Department, 1200 N Dupont Hwy, Dover, DE 19904, USA. Email: pokie@ist.temple.edu