Combining Regressive and Auto-Regressive Models for Spatial-Temporal Prediction

Dragoljub Pokrajac DPOKRAJA@EECS.WSU.EDU
Zoran Obradovic ZORAN@EECS.WSU.EDU

School of Electrical Engineering and Computer Science, Washington State University, Pullman WA 99164 USA

Abstract

A two-phased method for prediction in spatial-temporal domains is proposed. After an ordinary regression model is trained on spatial data, a prediction is adjusted by incorporating auto-regressive modeling of residuals in time. The prediction accuracy of the proposed method is evaluated on simulated agricultural data with a significant improvement of accuracy for both linear and non-linear regression models. The obtained experimental results suggest that when auto-regressive residual modeling is included, computationally more efficient linear regression models may predict almost as good as non-linear ones.

1. Introduction

Prediction of a continuous response variable in spatial-temporal domains, has recently drawn significant attention in the data analysis community (Cressie & Majure, 1997; Pace Kelly *et al* 1998). Spatial-temporal regression models learned on systematically collected values of driving attributes can contribute to a better understanding of complex phenomena studied in meteorology, oceanography, environmental science, precision agriculture and other domains. However, spatial-temporal modeling is often difficult due to various factors, including:

- Small number of available time layers;
- Low spatial sampling resolution on a uniform spatial grid and errors due to data interpolation;
- High influence of unobserved attributes;
- Non-stationarity in space and time;
- Non-linear dependence on driving attributes.

Typical spatial prediction methods have been developed assuming non-uniform event-driven sampling, where the objective is interpolation at different spatial positions (at current time). In contrast, we focus on a uniform grid and the prediction of unknown future response values.

Auto-regressive models use information from a spatial and a temporal neighborhood to perform a prediction at a specified location. Performance improvement as compared to ordinary regression models is often possible due to the postulated spatial and temporal correlation of data. However, an interpolation error in real-life data can introduce a false spatial similarity and therefore may have a negative impact on accuracy of an auto-regression model.

For non-linear phenomena, learning algorithms that model a response variable as a non-linear function of driving attributes may be superior to linear regressors. However, due to the presence of noise in data, an insufficient size of a training set, and an interpolation error, linear models can outperform non-linear ones in practice (Pokrajac, Obradovic & Fiez, 2000).

Finally, a majority of spatial-temporal learning algorithms were developed for stationary or time-constant processes. Data non-stationarity can significantly decrease the prediction quality and the applicability of such prediction models.

The purpose of this paper is to examine the effect of including auto-regressive modeling of ordinary regression residuals for learning on non-stationary spatial-temporal data sampled on a uniform grid. The proposed method combines linear or non-linear non-spatial and non-temporal regression models learned on data collected at particular time moments with spatial-temporal auto-regression of residuals.

After the survey of related work presented in Section 2, the proposed methods for spatial-temporal prediction is described in Section 3, along with a technique for experimental data generation. The obtained results are reported in Section 4, followed by conclusions and directions for future work discussed in Section 5.

2. Related Work

2.1 Spatial Auto-regression

In analysis of spatial data, numerous attempts are made to explicitly include a spatial component into prediction models. In models with spatially correlated residuals and with auto-regressive disturbance (Pace & Gilley, 1998; Florax & Folmer, 1992) modeling consists of two steps. First, the response variable is treated as non-spatial and a linear model is applied. Then, the residuals of a linear model on training data are assumed spatially correlated and their dependence is modeled through a matrix W using an auto-regressive approach. The final model, using the notation from Table 1, is:

$$Y= X\beta+W(Y-X\beta)+ \varepsilon \qquad (1)$$

A variation of this model, when the response is assumed to be spatially-correlated, is proposed by (Burridge, 1981), where:

$$Y=X\beta+\rho WY+\varepsilon \qquad (2)$$

In a mixed regressive-spatial cross-regressive model (Florax & Folmer, 1992), the response linearly depends on driving attributes not only from the current but also from the neighboring points. This dependence is modeled using a vector γ of cross-correlation coefficients:

$$Y=X\beta+WX^*\gamma+\varepsilon \qquad (3)$$

Finally, the mixed regressive-spatial auto-regressive model (Florax & Folmer, 1992) is a generalization of models (2) and (3). This model assumes a spatially correlated response also dependent on attributes of the neighboring points:

$$Y=X\beta+WX^*\gamma+\rho WY +\varepsilon \qquad (4)$$

Models (1)-(4) have the following common properties:

- Models are used for interpolation of non-uniform event-driven samples;
- A time-independent response variable is assumed;
- The response variable is assumed to be linearly dependent on driving attributes.

Table 1. Notation for parameters and arguments of regression models

Spatial and spatial-temporal models:

n	number of patterns/data tuples
k	number of explanatory variables/observed attributes
Y	n*1 column vector of observed response variables on n patterns within the observed spatial region.
X	n*(k+1) matrix in which each row corresponds to *k* observed attributes for a pattern and an intercept
X*	n*k matrix of observed driving attributes without an intercept
β	(k+1)*1 vector of regression parameters
$x(\beta)$	non-linear function of matrix X with parameters β
W	sparse n*n matrix having zero diagonal elements
γ	k*1 column vector of cross-correlation coefficients
ρ	autocorrelation coefficient
ε	n*1 vector of independent identically distributed Gaussian disturbances
U	n*1 vector of correlated residuals

Index *t* denotes response variable/driving attributes at instant *t* for temporal models.

Non-spatial time models:

x_t, y_t, u_t observed attributes, response variable and residual in time *t*, respectively.

2.2 Temporal and Spatial-Temporal Auto-Regression

2.2.1 MODELING TEMPORAL DATA

Temporal data can be modeled using a serial-correlation model. The response variable is assumed to be a function of driving attributes, while residuals are assumed serially correlated, satisfying AR(1) model (Davidson & McKinnon, 1993):

$$y_t=x_t\beta+u_t \qquad (5)$$
$$u_t=\rho u_{t-1}+\varepsilon_t$$

To estimate a serial-correlation model, one can perform the following iterative procedure (Davidson & McKinnon, 1993):

1. Set ρ=1
2. Perform linear regression by estimating model
   $y_t-\rho y_{t-1}=(x_t-x_{t-1})\beta+\varepsilon_t$

3. Compute value of ρ in next iteration by estimating the regression model
   $u_t = \rho\, u_{t-1} + e_t$
   where e_t is an auxiliary residual.
4. Repeat steps 2. and 3 until a pre-specified convergence criterion is satisfied.

In this model, the stationarity of regression coefficients β is implicitly assumed, which is equivalent to imposing a modeling restriction $\beta_t = \beta_{t-1}$. Also, observe that the serial correlation model does not consider a spatial component of data. Therefore, a new class of spatial-temporal models has recently been developed.

2.2.2 Spatial-temporal Modeling

Spatial-temporal prediction can be performed using a generalization of the model with auto-regressive disturbance (1). Here, the correlation matrix W represents spatial-temporal correlation of residuals. This matrix is estimated assuming that the second-order statistics of the residuals satisfy theoretical spatial-temporal variograms (Cressie & Majure, 1997). Modeling includes estimation of linear regression coefficients β and computation of residuals, as well as the estimation of spatial-temporal variograms. When the model is estimated, the prediction is performed as a sum of regressors Xβ and residuals estimated using a spatial-temporal kriging.

Similar as the serial correlation model, this model implies the stationarity of regression coefficients is time. Also, to properly estimate spatial-temporal variograms, the existence of a relatively high number of data time layers is necessary. Recall that these are strong and often unattainable requirements for a number of spatial-temporal domains (e.g. in precision agriculture, due to a recent adoption of a global positioning system-based measurement technology, spatial data currently exists for about the last 5 years).

Another approach for regression of spatial-temporal data, proposed by (Pace at al, 1998), is a generalization of the mixed regressive-spatial auto-regressive model (4). Here, the correlation matrix is assumed to be a product of matrices T and S, related to the time and the space dependence, respectively. Each sample from training data is assumed dependent on a fixed number of its spatial neighbors (regardless the time) and a fixed number of its time neighbors (that immediately precedes the observed sample).

The main problem when applying this model is a correct estimation of T and S. In (Pace at al., 1998) the forms of matrices T and S are postulated and the maximal number of influential neighbors is pre-specified. The model is developed for real-estate data, where each sample occurs in a distinct time instant and the number of time neighbors considered is small. In contrast, in the case of a uniform grid there is a larger number of samples collected at each time moment. Hence, the resulting matrix T is huge and the application of this method can be prohibitively laborious.

3. Methodology

3.1 The Proposed Method

We propose a combined regression-auto-regression model, described by:

$$
\begin{aligned}
Y_t &= x_t(\beta_t) + U_t \\
Y_{t-1} &= x_{t-1}(\beta_{t-1}) + U_{t-1} \\
U_t &= W U_{t-1}
\end{aligned}
\qquad (6)
$$

After non-spatial regression models $x_t(\beta_t)$ and $x_{t-1}(\beta_{t-1})$ are trained on data from time layers *t*-1 and *t* and residuals U_{t-1} and U_t are computed, the spatial autoregression of U_t on U_{t-1} is performed such that residuals corresponding to neighbors of the same order are weighted with the same coefficient. Here, two samples are called the *l-th order neighbors* if maximal absolute difference of their spatial coordinates is $l\Delta$, where Δ is a sampling distance (see an example at Figure 1). The maximal order L of neighbors concerned is an input parameter of the algorithm. Also, both linear and non-linear non-spatial regression models $x_t(\beta_t)$ can be applied.

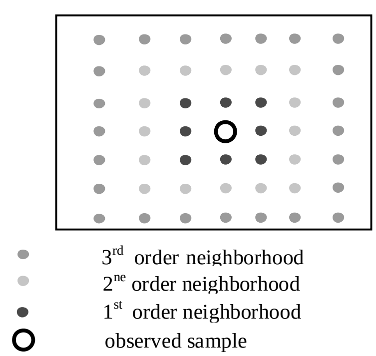

Figure 1. Definition of a neighbor order for spatial data with an uniform grid

Models are tested on data from time *t*+1 using driving attributes collected in time *t*+1 and residuals computed using the prediction and true response values at time *t*.

Observe that the proposed model exhibits certain similarities with the first iteration of Cochrane-Orcutt serial correlation algorithm. However, here regression coefficients do not obey a restriction $\beta_t = \beta_{t-1}$, hence the proposed procedure can be more suitable for non-stationary processes.

Similar to the spatial-temporal auto-regressive model (Pace at al, 1998), in the proposed model the influence of time and space neighbors is considered separately. Further, in both models the maximum size of a spatial neighborhood influence must be pre-specified. However, in contrast to the spatial-temporal auto-regressive model, the proposed model does not involve a spatial regression on attributes. Also, in the proposed model, the maximal time lag of considered residuals is one, which makes prediction less time-complex and potentially more resistant to data non-stationarity.

Unlike spatial auto-regression models, the proposed model is aimed to the prediction of response values in future. Compared to generalization of the model with auto-regressive disturbance (Cressie & Majure, 1997), the proposed model has more degrees of freedom and therefore has a potentially higher explanatory power.

3.2 Generation of Experimental Data

To create data satisfying pre-specified spatial and temporal statistical properties and to control the data complexity and noise level, experiments were performed on data generated using our spatial-temporal data simulator (Pokrajac, Fiez & Obradovic, 2000).

The simulation process consisted of the following phases:

- Generation of driving attributes
- Generation of the response variable
- Simulation of sampling error, attribute sensor noise and unexplained variance

3.2.1 Simulation of Driving Attributes

Driving attributes are generated through a multistep process of grid determination, generation of spatially correlated attributes and cluster generation (Pokrajac, Fiez & Obradovic, 2000).

After determining a sampling distance and dimensions of a rectangular grid, for each attribute a specified number of time layers is generated using kriging of a random seed vector (Cressie, 1993). Using Cholesky decomposition, a seed vector is generated to satisfy specified spatial and temporal correlation (Pokrajac, Fiez & Obradovic, 2000; Pokrajac, Obradovic, unpublished results).

To generate the specified number of attribute clusters, corresponding cluster “seeds” are chosen and each data point is “moved“ towards the nearest cluster seed. The intensity of the shift is proportional to the distance to the seed point and can be adjusted to control cluster aggregation. To avoid unnaturally clear separation between clusters, “perturbation" noise with variance proportional to that of the attributes may be introduced.

3.2.2 The Simulation of Response Variable

For each attribute cluster and time layer, a distinct non-linear mapping of driving attributes into the response variable can be applied. The simulator supports several non-linear functions including multiplicative and plateau models, particularly suitable for agriculture applications. A temporal component of a response variability is simulated with AR(1) models (Davidson & McKinnon, 1993), independently assigned to each attribute cluster. Consequently, a real-life situation where particular zones of a considered spatial area behave differently in time can be modeled. The percent of the response variability due to an AR(1) process is also user-controllable.

3.2.3 The Simulation of Sampling Error, Sensor Noise and Unexplained Variance

Emulating the actual sampling process simulates the error introduced through the interpolation of sampled values. Data are sampled from the generated grids and used to interpolate values at un-sampled locations using kriging (Cressie 1993).

The effect of measurement error on driving attributes and the response variable is modeled as multiplicative Gaussian noise with a unit-mean and variance determined by a specified measurement error.

Finally, unexplained variance in the response variable is modeled by additive Gaussian zero-mean noise having a variance determined by the specified percentage of unexplained variance in resulting data.

4. Experiments

4.1 Properties of Experimental Data

Experiments were performed on simulated agricultural data consisting of five time layers. Data contained samples of 5 simulated driving attributes and the response variable. Each time layer consisted of 6561 samples from $800*800m^2$ rectangular field, on the sampling distance 10m.

Five simulated attributes had a spatial correlation similar to the following real-life agricultural variables: nitrogen (N), phosphorus (P), potassium (K), water content (W) and slope (S). Attributes W and S were assumed to be constant in time, while the other attributes were modeled as time-dependent. In the absence of real-life data temporal statistics, percents of total variability due to the temporal variance were varied in 10-80% range, while the auto-correlation of successive time layers was chosen according to an expert estimate (see Table 2). After the generation of correlated attributes, four clusters in the space of topographic attributes W and S were formed.

Table 2: Spatial and temporal statistic parameters of simulated driving attributes

Attribute name		N	P	K	W	S
Spatial parameters	Range(m)	200	300	400	100	200
	Nugget(%)	0	0	0	0	0
Temporal parameters	% temporal variability	80	20	10	Attributes do not change over time	
	Correlation	0.9	0.9	0.7		

Crop yield, the response variable, was generated using linear plateau models. For each cluster, the relative influence of particular attributes on the simulated response and a shape (slope and thresholds) of linear plateau functions were varied according to an expert knowledge. The mean and standard deviation of simulated crop yield were similar to that of real-life data. Finally, an unexplained variance in range of 5-35% and an attribute sensor error in range 5-15% were introduced.

4.2 Evaluation of Method Accuracy

A linear regression was performed using the OLS method (Devore 1995). A non-linear modeling was performed using the sigmoidal perceptrons trained with the Levenberg-Marquardt algorithm (Haykin, 1999), with 1 hidden layer having 4 neurons. Experiments with non-linear models were repeated 10 times each.

The prediction accuracy was measured using the coefficient of determination R^2. R^2 is a measure of the explained response variability. In the case of useful prediction models it ranges from 0 to 1, where 0 results from using a trivial mean predictor and 1 represents the ideal case of no prediction error. A one-sided t-test was used to compare the accuracy of linear and non-linear models (Devore 1995).

Results of the proposed method were compared to the results of ordinary linear and non-linear regression models. Five time layers were generated and since the training of the proposed method required data from two successive layers, trained models were tested on time layers 3,4 and 5.

4.3 Results

The results of comparing the proposed and ordinary regression methods are presented in Table 3, where the set of attributes used for training a non-spatial model x(β) in (6) is varied. The accuracy of the proposed method is shown both for non-spatial (L=0) and the spatial auto-regression with first-order neighbors (L=1). Further increasing of L did not result with significant accuracy

Table 3: Comparison of the proposed method and ordinary linear and non-linear regression

a) all driving attributes are used for model training.

Regression model	Time layer	Mean accuracy (R^2)		
		Ordinary regression	**Proposed method**	
			L=0	**L=1**
Linear	**3**	0.11	0.63	0.65
	4	0.19	0.80	0.81
	5	0.18	0.73	0.75
Non-linear	**3**	0.46**	0.64	0.70**
	4	0.70**	0.83**	0.84
	5	0.67**	0.83**	0.85**

b) 3 time-dependent features (N,P,K) were used for modeling

Regression model	Time layer	Mean accuracy (R^2)		
		Ordinary regression	**Proposed method**	
			L=0	**L=1**
Linear	**3**	0.13**	0.59**	0.59**
	4	0.12	0.80**	0.82**
	5	0.15	0.72	0.73
Non-linear	**3**	0.02	0.41	0.44
	4	0.19**	0.77	0.79
	5	0.20**	0.70	0.71

c) only spatial coordinates (x,y) were used for modeling

Regression model	Time layer	Mean accuracy (R^2)		
		Ordinary regression	**Proposed method**	
			L=0	**L=1**
Linear	**3**	-0.48	-0.06	-0.05
	4	0.16	0.76	0.78
	5	0.19	0.73**	0.74*
Non-linear	**3**	-0.63	-0.05	-0.03**
	4	-0.00	0.77	0.79**
	5	0.01	0.71	0.72

improvements, and so the corresponding results are not presented. Our hypothesis is that when the higher order neighbors are considered (corresponding to larger L), the auto-regression models become too complex and thus gradually loose their generalization capabilities.

When models x(β) were trained on all driving attributes, results suggest that the accuracy of both linear and non-linear regression models can be significantly improved using the proposed method, as shown in Table 3a. This is particularly true for linear models (e.g. for time layer 4, R^2 has an improvement of 61%). Observe that in this case non-linear models steadily outperformed linear ones with

99% significance of a t-test (denoted by ** on Tables 3 and 4).

With models x(β) trained only on time-dependent attributes (N,P,K), a significant improvement of accuracy using the proposed method was again achieved (Table 3b). This is in accordance with a claim that using auto-regressive models can improve prediction accuracy when some driving attributes are missing from the model (Colwell, Cannaday & Wu, 1983). Observe that in this case linear models frequently outperformed the non-linear ones when used within the proposed method.

Our hypothesis was that due to the spatial correlation of driving attributes an introduction of any spatial information into regression models could have a positive impact on the prediction accuracy. To illustrate this, we repeated previous experiments training regression models x(β) on spatial coordinates x and y only. As expected, a significant improvement of accuracy using the proposed method was achieved again (Table 3c).

As can be seen from Table 3, the introduction of the first neighbor residuals in prediction models resulted in a rather small increase of modeling accuracy. Similar results were obtained when attribute noise was added. However, when the proposed method was applied on data with a high unexplained variance (noise) in the response variable, an improvement due to a spatial neighborhood consideration (L=1) was significant (Table 4). An analog behavior was evident for unobserved variance levels in the range (5-35%)

Table 4: Comparison of the proposed method and ordinary linear and non-linear regression when 35% unexplained variance added to the response variable and all driving attributes used for model training

Regression model	Time layer	Mean accuracy (R^2)		
		Ordinary regression	**Proposed method**	
			L=0	**L=1**
Linear	**3**	-0.37	-0.17	-0.05
	4	-0.01	0.32	0.45**
	5	0.00	0.34	0.43
Non-linear	**3**	-0.28	-0.14**	-0.05
	4	0.11	0.34**	0.44
	5	0.13	0.37**	0.44**

5. Conclusions and Further Research

In this paper, a spatial-temporal data prediction technique based on the combination of non-spatial regression and spatial-temporal auto-regression of residuals, is proposed. Using simulated spatial-temporal data, the proposed method was compared to the ordinary linear and non-linear models. Experimental results suggest that the proposed method can significantly improve the prediction accuracy of linear regression models. Also, an accuracy improvement can be achieved when driving attributes are missing and when attribute/response variable noise is present.

The research in progress includes theoretical verifications of the proposed method and the comparison with other known spatial-temporal regression methods on simulated and real-life data of various types.

References

Colwell P., Cannaday, R., & Wu, C. (1983). The Analytical Foundations of Adjustment Grid Methods. *Journal of the American Real Estate and Urban Economics Association, 11,* 11-29.

Cressie, N. (1993). *Spatial statistics*. New York: Willey.

Cressie, N., & Majure, J. J. (1997). Spatio-temporal statistical modeling of livestock waste in streams. *Journal of Agricultural, Biological, and Environmental Statistics,* 2, 24-47.

Davidson R., & Mc Kinon, J. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Devore, J.L. (1995). *Probability and Statistics for Engineering and the Sciences. 4th ed*. Belmont: Int'l Thomson Publishing Company.

Florax, R., & Folmer, H. (1992). Specification and Estimation of Spatial Linear Regression Models-Monte Carlo Evaluation of Pre-test Parameters. *Regional Science and Urban Economics, 22*, 405-432.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation, 2nd ed.* New Jersey: Prentice Hall.

Pace Kelly R.,& Gilley, O. (1998). Generalizing the OLS and Grid Estimators. *Real Estate Economics, 26,* 331-347.

Pace Kelly R., Barry, R., Clapp, J.& Rodriquez, M. (1998). Spatiotemporal Auto-regressive Models of Neighborhood effects. *Journal of Real Estate Economics, 17*, 15-33.

Pierce, F. (1997). *The state of Site-Specific Management for Agriculture American Society for agronomy* (pp. 283-300).Crop Science Society of America, Soil Science Society of America.

Pokrajac, D., Fiez, T., & Obradovic, Z. (2000). A Tool for Controlled Knowledge Discovery in Spatial Domains. *Proc. 14th European Simulation Multiconference (ESM),* (in press).

Pokrajac, D., Obradovic, Z. & Fiez, T. (2000). Understanding the Influence of Noise, Sampling Density and Data Distribution on Spatial Prediction Quality Through the Use of Simulated Data. *Proc. 14th European Simulation Multiconference (ESM),* (in press).