MOAD'2024

The Sixth International Colloquium on Methods and Tools for Decision Support

October 20th–22nd 2024, Tizi-Ouzou, Algeria

Impact of missing data on prediction in random fields

Abdelghnai HAMAZ[1], Mohamed IBAZIZEN [2]

[1] Mouloud Mammeri University of Tizi-Ouzou, ALGERIA and [2] University of Poitiers, FRANCE.

ghani.hamaz@ummto.dz [1] and moad24@ummto.dz[2]

Abstract

The purpose of this paper is to treat the prediction problems where a number of observations are missing to the quarter-plane past of a stationary random field. Our aim is to quantify the influence of missing values on the prediction by giving the simple bounds for the prediction error variance. These bounds allow to characterize the random fields for which the missing observations do not affect the prediction. Simulation experiments and an application to real data are presented.

Introduction

The problem of linear prediction of stationary processes requires knowledge of the observed past and their covariance function. When the data is coming from physical and natural sciences it is common to have irregularities, missing or outlying observations. The problem of spatial prediction based on incomplete past was considered by Kohli and Pourahmadi [3] to provide estimates for the missing values as well as the predictors. The original impetus for their work came from the interpolation results in [2] and prediction based on incomplete past in [1] for a second order stationary time series. The key idea in both these methods is the appropriate orthogonalization of the "past" and "future" of the time series, where past corresponds to the infinite past of the first missing value and future to all the values observed between missing values and the time point at which we need to predict. In this paper, we investigate the problem of linear prediction of stationary random fields with non-symmetrical half-plane past. Our main contribution lies in finding an explicit formula of the mean square convergent autoregressive series representation for all (h_1, h_2)-step ahead linear predictors, $(h_1,\, h_2) \geq (0,\, 0)$. In order to calculate explicitly the prediction coefficients of our new expression

Main result:Autoregressive representation

The multi-step ahead prediction problem of stationary random fields has been studied by [3] when the third quadrant is used as the past. We extends their pioneer work to random fields with non-symmetrical half-plane past. Let $\{X(s,\, t); (s,\, t) \in \mathbb{Z}^2\}$ is a PND stationary random field. The procedure for solving the $(h_1,\, h_2)-$step ahead linear prediction problem with respect to the total order and nonsymmerical half-plane (NSHP) support defined by (3) involves the construction of predictor of future values as a linear combination of $\{X(k,\, l),\, (k,\, l) \in S\}$ which are close to $X(s + h_1,\, t + h_2)$, $(h_1,\, h_2) \geq (0,\, 0)$ in the sense of mean squared error. The representation (1) is inverted to give

$$\varepsilon(s,\, t) = \sum_{k=0}^{+\infty}\sum_{l=0}^{+\infty} a_{k,\, l} X(s-k,\, t-l). \tag{1}$$

		(n,m)=(100,150)		(n,m)=(150,200)		(n,m)=(250,250)	
a	b	ν_1	ν_2	ν_1	ν_2	ν_1	ν_2
0.05	0.557	0.652	0.650	0.647	0.644	0.641	0.639
0.10	0.569	0.691	0.687	0.638	0.640	0.631	0.627
0.15	0.582	0.697	0.692	0.621	0.622	0.615	0.611
0.20	0.594	0.701	0.698	0.620	0.616	0.608	0.602
0.25	0.603	0.709	0.707	0.621	0.619	0.609	0.613
0.30	0.608	0.661	0.669	0.592	0.582	0.581	0.576
0.35	0.607	0.603	0.609	0.607	0.598	0.594	0.591
0.40	0.601	0.599	0.603	0.589	0.596	0.576	0.573
0.45	0.587	0.579	0.588	0.580	0.573	0.567	0.561
0.50	0.562	00567	0.571	0.559	0.562	0.546	0.549
0.55	0.516	0.553	0.560	0.601	0.609	0.556	0.559
0.60	0.406	0.542	0.549	0.536	0.540	0.531	0.532

Proof. We have (3) implies that

$$E(\varepsilon(s,\, t)X(s,\, t)) = E(\varepsilon(s,\, t))^2.$$

From (3) we deduce that

$$E(\varepsilon(s,\, t))^2 = a_{00} E(\varepsilon(s,\, t)X(s,\, t)),$$

and necessarily $a_{00} = 1$. Thus, (1) may be rewritten as

$$X(s,\, t) = \varepsilon(s,\, t) - \sum_{\substack{k=0 \\ (k,\, l)\neq(0,\, 0)}}^{+\infty}\sum_{l=0}^{+\infty} a_{k,\, l} X(s-k,\, t-l).$$

□

Lemma 3.1. Let $\{X(T);\, T \in \mathbb{Z}^2\}$ be a PND stationary random field, the MA and the AR parameters are $\{b_{k,\, l},\, (k,\, l) \in \mathbb{Z}^2\}$ and $\{a_{k,\, l},\, (k,\, l) \in \mathbb{Z}^2\}$, respectively, then the following equation is satisfied for all $(k,\, l) \geq (0,\, 1)$

$$\sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij} b_{k-i,\, l-j} = 0. \tag{2}$$

Proof. By substituting (1) we obtain for all $(k,\, l) \geq (0,\, 1)$

$$\begin{aligned} 0 = E(\varepsilon(s-k,\, t-l)\varepsilon(s,\, t)) &= E(\varepsilon(s-k,\, t-l)\sum_{i=0}^{+\infty}\sum_{j=0}^{+\infty} a_{ij} X(s-i,\, t-j)) \\ &= E(\varepsilon(s-k,\, t-l)\sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij} X(s-i,\, t-j)) \\ &= \sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij} E(\varepsilon(s-k,\, t-l)X(s-i,\, t-j)) \\ &= \sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij} b_{k-i,\, l-j} E(\varepsilon(s-k,\, t-l))^2. \end{aligned}$$

□

Simulation study

Consider the stationary first order multiplicative spatial autoregressive model (MSAR(1)) defined by

$$X(s,t) = aX(s-1,t) + bX(s,t-1) - a.bX(s-1,t-1) + \epsilon(s,t) \tag{3}$$

where $\{\epsilon(s,t); (s,t) \in \mathbb{Z}^2\}$ are independent random variables with $\mathbf{E}(\epsilon(s,t)) = 0$, $\mathbf{Var}(\epsilon(s,t)) = \sigma^2$, $|a| < 1$ and $|b| < 1$.

By using the recursions given by (4) and the fact that $a_{10} = a$, $a_{10} = b$, $a_{11} = -a.b$ and $a_{ij} = 0$ if $(i,j) \notin \{(1,0),(0,1),(1,1)\}$, it can be shown that the MA representation of the MSAR(1) model is

$$b_{k,l} = \begin{cases} a^k b^l, & \text{if } k \geq 0, l \geq 0 \\ 0 & \text{if } k < 0 \text{ or } l < 0. \end{cases} \tag{4}$$

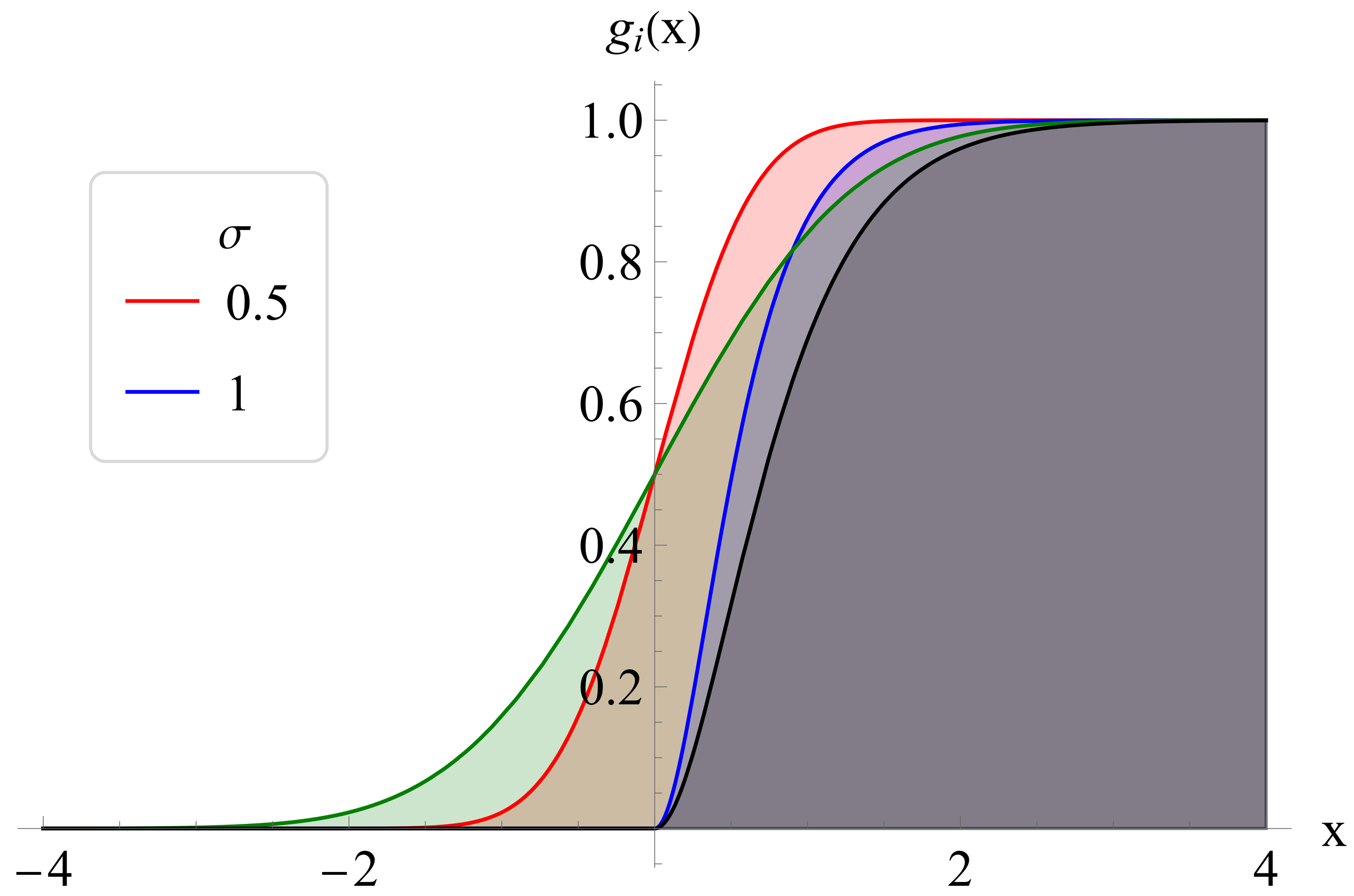
In the same way, the best linear predictor of $X(0,0)$ based on future observations is

$$\hat{X}_{h_1,h_2}(0,0) = \frac{1}{a.b}\left(aX(0,1) + bX(1,0) - X(1,1)\right),$$

and then the suboptimal predictor is

$$\tilde{X}(0,0) = \alpha\hat{X}(0,0) + \beta\hat{X}_{h_1,h_2}(0,0). \tag{5}$$

The best linear interpolator of $X(0,0)$ by performing the extension of the prediction.



Conclusion

The purpose of this paper is to treat the prediction problems where a number of observations are missing to the quarter-plane past of a stationary random field. Our aim is to quantify the influence of missing values on the prediction by giving the simple bounds for the prediction error variance. These bounds allow to characterize the random fields for which the missing observations do not affect the prediction. Simulation experiments and an application to real data are presented.

References

[1] Cheng, R. (2015). Prediction of stationary Gaussian random fields with incomplete quarterplane past. Journal of Multivariate Analysis, 139, 245-258.

[2] Hamaz, A., Arezki, O., & Achemine, F. (2020). Impact of missing data on the prediction of random fields. Journal of Applied Statistics, 47(1), 132-149.

[3] Kohli, P., & Pourahmadi, M. (2014). Some prediction problems for stationary random fields with quarter-plane past. Journal of Multivariate Analysis, 127, 112-125.