A nonparametric spatio-temporal approach for multiple CUSUM charts of evapotranspiration

D. Pellegrino[1,*], G. Giungato[1] and S. De Iaco[1]

[1] *University of Salento, Via per Monteroni, Complesso Ecotekne, Lecce, Italy; daniela.pellegrino@unisalento.it, giuseppina.giungato@unisalento.it, sandra.deiaco@unisalento.it*
**Corresponding author*

Abstract. *In recent years, the interest in natural resources management is increased. In this context, control charts, developed to monitor and maintain quality in industrial processes, are a useful monitoring and decision tool.*
In this paper, the behavior of an agro-metereological variable, named evapotranspiration, in an area of the northern part of Italy during a 25-year span (from 1992 to 2016) is studied through a nonparametric spatio-temporal geostatistical analysis and multiple CUSUM control charts. In particular, the probability that the variable registers an "out of control" is estimated over the area of interest, for three decades from the 10th to the 30th of January 2017.

Keywords. *Spatio-temporal geostatistics; Indicator kriging; Control charts.*

1 Introduction

The sustainable management of natural resources is an increasingly complex issue for environmental sciences. Hence, monitoring represents an important activity for decision-making procedures. In this context, the control charts might be useful for natural resources management, although they were developed as a tool in the Statistical Process Control (SPC) for improving industrial processes. On the basis of the classical approach, these SPC techniques are a representation of the quality characteristic measured in a sample or in several samples of an industrial process and allow pointing out if the process is "out of control" and it should be stopped (Montgomery, 2009).
The convenience of using the control charts approach in different fields such as environmental, economics, financial, social and healthcare sciences was discussed in several studies. In particular, the interest in SPC techniques to analyze environmental phenomena is increasing (Paroissin et al., 2016; Garthoff and Otto, 2016). On the other hand, few attempts to integrate the control charts with Geostatistics have been made, such as in Grimshaw et al. (2013). However, the geostatistical methods applied in the above mentioned papers were not used in a joint way in the space and in the time.
In this paper, the Cumulative Sum (CUSUM) charts, introduced by Page (1961), are used to study an agro-meteorological variable, i.e. evapotranspiration (ET_0), in 26 stations of Veneto region, in the period 1992-2016. In particular, the CUSUM charts technique has been integrated with nonparametric spatio-temporal geostatistical methods in order to predict the probability that the CUSUM chart signals that the variable is "out of control". These results could be useful to plan adequate water management strategies, since the ET_0 monitoring plays an important role in irrigation scheduling, watershed level budgeting, as well as climate and weather models.

2 CUSUM charts and geostatistical framework

In environmental field, the detection of changes in a phenomenon could represent a useful tool to define management and controlling plans. Hence, the CUSUM charts, based on the cumulative sums of deviations of the analyzed values from a target value (μ_0), might be a convenient technique for monitoring this variability. Neverthless, it is worth noting that the variables under study are usually nonstationary, so the residuals must be considered (Montgomery and Mastrangelo, 1991).
The residual values of a spatio-temporal environmental phenomenon, recorded at different time points and spatial locations, can be considered as a realization of a second-order stationary spatio-temporal random function ($STRF$), $\{Y(\mathbf{u}), \mathbf{u} = (\mathbf{s},t) \in D \times T\}$, where $D \subseteq \mathbb{R}^d$ and $T \subseteq \mathbb{R}$.
In particular, for each spatial location, the chart is obtained by plotting over the time the cumulative values $CS(\mathbf{s},t) = \sum_{j=1}^{t} [Y(\mathbf{s}, j) - \mu_0] = [Y(\mathbf{s},t) - \mu_0] + CS(\mathbf{s},t-1)$, where $CS(\mathbf{s},0) = 0$. On the other hand, the CUSUM could be expressed in the form of decision-interval, based on the cumulative sums of positive and negative deviations from the target value μ_0 that are greater than a reference value indicated with K, respectively:

$$CS^+(\mathbf{s},t) = max[0, Y(\mathbf{s},t) - (\mu_0 + K) + CS^+(\mathbf{s},t-1)],$$
$$CS^-(\mathbf{s},t) = max[0, (\mu_0 - K) - Y(\mathbf{s},t) + CS^-(\mathbf{s},t-1)],$$

with starting values $CS^+(\mathbf{s},0) = CS^-(\mathbf{s},0) = 0$. The CUSUM chart is obtained by plotting these statistics over the time. In particular, if measurements are above the reference value, the upper CUSUM CS^+ shows an upward trend; likewise, the lower CUSUM CS^- exhibits a downward trend if the phenomenon is consistently below the reference value.
Finally, the parameters K and H must be fixed. K is related to the size of the smallest shift in the level of the reference value that can be detected; while H is the threshold that CS^+ and CS^- should not exceed in order to consider the phenomenon "in-control".
In a nonparametric context, given the fixed threshold $z = H$ and the CS^+ and CS^- computed from residuals, a spatio-temporal indicator random field ($STIRF$),

$$\{I(\mathbf{u},z), \mathbf{u} = (\mathbf{s},t) \in D \times T\}$$

can be defined as follows:

$$I(\mathbf{u},z) = \begin{cases} 1 & \text{if } CS^+(\mathbf{u}) \geq z \text{ or } CS^-(\mathbf{u}) \geq z \,, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Under the second-order stationarity, the spatio-temporal indicator variogram, which describes the correlation, depends on the threshold z and the lag vector $\mathbf{h}$, i.e. $2\gamma_I(\mathbf{h};z) = Var[I(\mathbf{u}+\mathbf{h};z) - I(\mathbf{u};z)]$, where $\mathbf{h} = (\mathbf{h}_s, h_t)$.
In this context, the empirical spatio-temporal indicator variogram can be modelled through the following generalized product-sum model (De Iaco et al., 2001), selected among different spatio-temporal models proposed in literature:

$$\gamma_I(\mathbf{h}_s, h_t; z) = \gamma_I(\mathbf{h}_s, 0; z) + \gamma_I(\mathbf{0}, h_t; z) - k\gamma_I(\mathbf{h}_s, 0; z)\gamma_I(\mathbf{0}, h_t; z), \tag{2}$$

where $\gamma_I(\mathbf{h}_s, 0; z)$ and $\gamma_I(\mathbf{0}, h_t; z)$ are, respectively, spatial and temporal valid bounded marginal variograms and $k \in]0, 1/max\{sill\gamma_I(\mathbf{h}_s, 0; z), sill\gamma_I(\mathbf{0}, h_t; z)\}]$ is the parameter of spatio-temporal interaction.
For a second-order stationary $STIRF$ I, a linear prediction of the probability that the CS^+ or CS^- is greater than the threshold z, that means that the phenomenon is "out of control", can be obtained by using a linear combination of neighbouring indicator variables, expressed by the spatio-temporal indicator kriging predictor $\widehat{I}(\mathbf{u}; z) = \sum_{\alpha=1}^{n} \lambda_\alpha(\mathbf{u}_\alpha; z) I(\mathbf{u}_\alpha; z)$, where $I(\mathbf{u}_\alpha; z)$, $\alpha = 1, 2, \ldots, n$ represent the indicator random variables at the sampled points $\mathbf{u}_\alpha \in D \times T$ and $\lambda_\alpha(\mathbf{u}_\alpha; z)$ are the kriging weights, determined by solving the indicator kriging system (Journel, 1983).

3 Case study

The control of ET_0 levels in a geographic area is an important tool for water management and planning, since this variable is a very crucial factor in river discharge, irrigation water requirement and soil moisture contents (Mohan and Arumugam, 1996).
In the present case study, the ET_0 levels (expressed in *mm*) provided by an Italian web system, named SCIA (Desiato et al., 2007), for 26 agro-meteorological stations located in the northeastern part of Italy (Veneto Region) have been analyzed. Note that selected data are averaged every ten days and refer to a 25-year span (from 1992 to 2016).
ET_0 is characterized by a periodic behavior: high levels registered in autumn and in winter are in contrast to low measurements in the other seasons. Hence, in order to remove the periodic component exhibited by the data, the FORTRAN program REMOVEMULT described in De Iaco et al. (2010) has been used; consequently the residual data have been used in the steps of the analysis.
From residuals, for each spatial location positive (CS^+) and negative (CS^-) CUSUM have been computed by fixed the parameters $K = 2\sigma$ and $H = 3\sigma$, where σ is the global standard deviation equals to 0.359. Hence, by considering the parameter $H = 1.077$ as the threshold z, a nonparametric analysis has been conducted on the indicator variable $I(\mathbf{u};z)$, in order to estimate the probability that the CUSUM CS^- exceeds the threshold z and to predict the probability that the ET_0 is "out of control" for three future time points, that is the 10th, 20th and 30th of January 2017.
After computing the sample spatio-temporal indicator variogram (Fig.1), the space-time correlation of the indicator variable has been modeled through the following product-sum model:

$$\gamma_I(\mathbf{h}_s,h_t;z) = N_s + c_s\,Exp(\|\mathbf{h}_s\|;a_s) + N_t + c_t\,Exp(h_t;a_t) - k\{[N_s + c_s\,Exp(\|\mathbf{h}_s\|;a_s)]\cdot[N_t + c_t\,Exp(h_t;a_t)]\}$$

where N_s and c_s are, respectively, the nugget and the sill contribution of the spatial marginal indicator variogram model which is $\gamma_I(\mathbf{h}_s,0;z) = 0.015 + 0.019Exp(\|\mathbf{h}_s\|;a_s)$, while N_t and c_t are, respectively, the nugget and the sill contribution of the temporal indicator variogram model which is $\gamma_I(\mathbf{0},h_t;z) = 0.012 + 0.097Exp(h_t;a_t)$, with spatial range a_s equals to 80 km and temporal range a_t equals to 55 days. Note that the parameter k, which is equals to 7.104, is such that the admissibility condition is satisfied and the global sill, equals to 0.117, is fitted.

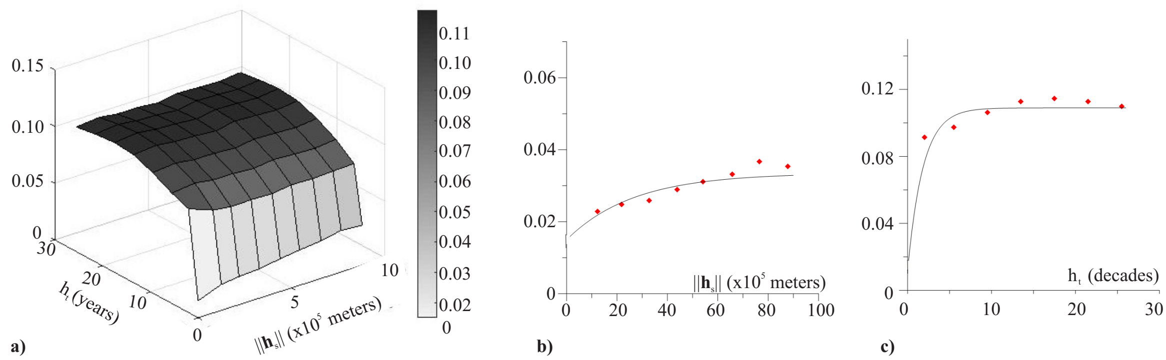

Figure 1: a) Sample indicator spatio-temporal variogram surface, b) marginal spatial variogram and fitted model, c) marginal temporal variogram and fitted model.

The reliability of the fitted spatio-temporal model is evaluated through cross-validation technique and some fitting indexes. The linear correlation coefficient between the observed values and the estimates

from cross-validation, equals to 0.8, confirms the goodness of the fitted model. Moreover, the Mean Error (ME) and the Root Mean Square Error (RMSE) computed on the fitting errors between the empirical surface and the model, equal to 0.007 and 0.005, respectively, confirm the accuracy of the fitted spatio-temporal model.
Finally, these models have been applied in order to obtain spatio-temporal indicator kriging predictions over the area of interest for three decades, from the 10th to the 30th of January 2017, through a modified *GsLib* routine (De Iaco et al., 2011). Then, the probability maps of the negative CUSUM CS^{-} exceeding the fixed threshold have been obtained. The results highlight that, in the analyzed area, there are low probabilities that the CUSUM exceeds the fixed threshold. Hence, the ET_0 behavior will be "in-control" in these three decades.

References

[1] De Iaco, S., Myers, D.E., Posa, D. (2001). Space-time analysis using a general product-sum model. *Statistics and Probability Letters* **52**(1), 21–28.

[2] De Iaco, S., Myers, D.E., Palma, M., Posa, D. (2010). FORTRAN programs for space-time multivariate modeling and prediction. *Computer & Geosciences* **36**(5), 636–646.

[3] De Iaco, S., Posa, D. (2011). Predicting spatio-temporal random fields: some computational aspects. Computer & Geosciences, doi: 10.1016/j.cageo.2011.11.014

[4] Desiato, F., Lena, F., Toreti, A. (2007). SCIA: a system for a better knowledge of the Italian climate. *Bollettino di Geofisica Teorica ed Applicata* **48**(3), 351–358.

[5] Garthoff, R., Otto, P., (2017). Control charts for multivariate spatial autoregressive models. *AStA Adv Stat Anal* **101**, 67–94.

[6] Grimshaw, S. D., Blades, N. J., Miles, M. P. (2013). Spatial Control Charts for the Mean. *Journal of Quality Technology* **45**(2), 130–148.

[7] Journel, A.G. (1983). Nonparametric estimation of spatial distributions. *Mathematical Geology* **15**(3), 445–468.

[8] Mohan, S., Arumugam, N., 1996. Relative importance of meteorological variables in evapotranspiration: Factor analysis approach. *Water Resource Management* **10**(1), 1–20.

[9] Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*. Sixth Edition. John Wiley & Sons, Inc.

[10] Montgomery, D. C., Mastrangelo, C. M. (1991). Some Statistical Process Control Methods for Autocorrelated Data. *Journal of Quality Technology* **23**(3), 179–193.

[11] Page, E. S. (1961). Cumulative sum charts. *Technometrics* **3**(1), 1–9.

[12] Paroissin, C., Penalva, L., Pétrau, A., Verdier, G. (2016). New control chart for monitoring and classification of environmental data. *Environmetrics* **27**, 182–193.