Spatiotemporal sampling design adapted to heterogeneities and real-time observations

M.C. Bueso[1], J.M. Angulo[2], F.J. Alonso[2], M.D. Ruiz-Medina[2]

[1]Department of Applied Mathematics and Statistics, Technical University of Cartagena, Paseo Alfonso XIII 52, Cartagena, E-30203 Murcia, Spain
[2] Department of Statistics and Operations Research, University of Granada, Campus de Fuente Nueva s/n, E-18071 Granada, Spain

mcarmen.bueso@upct.es, jmangulo@ugr.es, falonso@ugr.es, mruiz@ugr.es

Abstract

Entropy-based criteria for spatiotemporal sampling design naturally incorporate prior knowledge on structural heterogeneities of processes involved in environmental applications, an important aspect of variation to be considered for risk assessment purposes. Whenever possible, real-time observations must be also integrated for dynamic adaptation of the spatial sampling configurations, eventually under certain restrictions, to account for the actual evolution of the system. In this paper, such information is exploited to redefine, at each time, the region of interest in terms of local density. Procedures are applied to simulated examples where different ranges of memory and spatial dependence, as well as different levels of local variability (fractality), are specified to study the structural influence of the model in the entropy-based spatiotemporal sampling design.

Key words: *Risk assessment; heterogeneity; Shannon's entropy; spatiotemporal sampling.*

1. Entropy-based adaptive network design

Let S_0 be the starting observation set, consisting of preliminary observations used for a first estimation of process X at the initial time. At each time $t \geq 1$, the adaptive design procedure consists of the following steps:

1. Update the region of interest for the unobservable process X at time $t+1$, Λ_{t+1}, as follows:
   (a) Specify a distribution F to select the sites to be included in the set Λ_{t+1}. In particular, a Beta distribution will be considered here to generate points of interest in the neighborhood of critical values observed at time t.
   (b) By using the observations available at time t from $S_1, \ldots S_t$, obtain estimated values for the process of interest X at time t on a regular mesh. These values are normalized into the interval $[0,1]$ by applying a nondecreasing transformation h; in the examples here, we use the function $h(x) = \frac{x-min}{max-min}$, where min and max respectively represent the minimum and the maximum taken over the estimates.
   (c) Generate random values from the distribution F and assign each generated value to the closest value among the transformed values on the grid. The sites corresponding to these values determine the set Λ_{t+1}.
2. Select the set of locations to be observed at time $t+1$, S_{t+1}, by maximizing the mutual information between $\mathbf{X}_{\Lambda_{t+1}}$ (the random vector whose co-ordinates represent the spatiotemporal process of interest at Λ_{t+1}) and $Y_{(S_1,1),\ldots,(S_t,t),(S,t+1)}$ (observable process at sets $S_1, \ldots, S_t, S$, at times $1, \ldots, t, t+1$, respectively):

$$S_{t+1} = \arg \max_{S \subset \boldsymbol{\mathcal{S}}_{t+1}} I\left(X_{\Lambda_{t+1}}; Y_{(S_1,1),\ldots,(S_t,t),(S,t+1)}\right),$$

   where $\boldsymbol{\mathcal{S}}_{t+1}$ denotes the class of admissible sets considered from where S will be selected.

2. A simulated example

Assume that X is modelled by convolution of the following spatiotemporal filter f with the innovation process ε, considered to be spatiotemporal Gaussian white noise with variance σ_ε^2:

$$f(s,t;\mathbf{x},\mathbf{y};\beta,\boldsymbol{\theta}) = \frac{\beta}{(1+|t-s|^2+\|\mathbf{x}-\mathbf{y}\|^2)^{h(\boldsymbol{\theta},\mathbf{x})}},$$

with $\beta > 0$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The covariance function of X is then defined by spatiotemporal self-convolution of f.
The observation process Y here consists of the process of interest affected by additive noise,

$$Y_t(\mathbf{s}) = X_t(\mathbf{s}) + \epsilon_t(\mathbf{s}), \quad \mathbf{s} \in R^2,\ t \in \{1,\ldots,T\},$$

where ϵ represents a spatiotemporal Gaussian white noise process, mutually uncorrelated with X, with mean 0 and variance σ_ϵ^2. All the parameter values involved are assumed to be known.
The method is illustrated considering weak and strong heterogeneous dependence modelled by an exponent function h, introducing heterogeneity in the filter f. In the weak-dependent case (Case 1), h is given by

$$h = g(\boldsymbol{\theta},\mathbf{x}) = \frac{\theta_1}{\theta_2 + \|\mathbf{x}\|^{2/\theta_3}},$$

where for $\|\mathbf{x}\| > 0$, $\boldsymbol{\Theta}$ is designed with the restriction $h > 3/2$.
In the strong-dependent case (Case 2), the exponent function considered is $h = g^{-1}$. Note also that, in the strong-dependent case ($h^{-1} \in (0,3/2)$), the local regularity of the model increases with small $h-$values. Two particular cases are then considered corresponding to the parameter values $\beta = 25$, $\theta_1 = 6$, $\theta_2 = 0$, and $\theta_3 = 5$, and $\sigma_\varepsilon^2 = 3$, with $\sigma_\epsilon^2 = 0.0001$ in Case 1 and with $\sigma_\epsilon^2 = 0.001$ in Case 2. In both of them, with spatial domain $D = [0,10]^2 \subset \mathbb{R}^2$.

The set of potentially observable sites consists of 200 coordinates randomly generated from a regular 20×20 grid defined on D (see Figs. 1a) and 6a)). The set of locations of interest is defined by 16 fixed points, with coordinates $\{(2*i, 2*j), i = 1,\ldots,4, j = 1,\ldots,4\}$, plus 24 additional points to be selected, at each time, from the 20×20 grid.
For $t = 1,\ldots,6$, the design criteria are applied to sequentially select, at each time t, 50 observation sites from the set of candidates. The starting observation set S_0 is given by 50 sites randomly selected from that set, which are displayed in Figs. 1b) and 6b).
Realizations generated for process X on D and times $t = 1,\ldots,6$ are displayed in Figs. 2 and 7.
The locations included in the successive regions of interest obtained for $t = 1,\ldots,6$ are shown in Figs. 3 and 8, where we can observe relatively higher concentrations located in the areas with higher simulated values for the spatiotemporal process X. The resulting designs, displayed in Figs. 4 and 9, show the dynamical adaptation of the network, clearly influenced by the model dependence assumed in each case.
Finally, Figs. 5 and 10 show the results obtained when the first region of interest is maintained fixed for all subsequent times; the network minor changes over time are due only to structural variations, again reflecting the dependence effect.

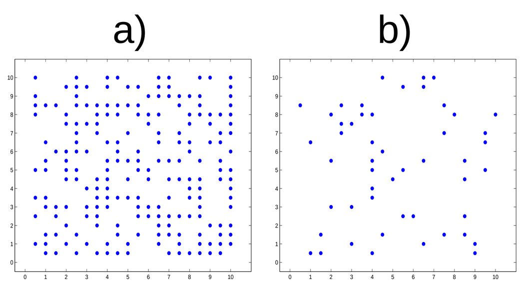

Figure 1: *Case 1. a) Set of candidate sites for observation, and b) starting observation set S_0.*

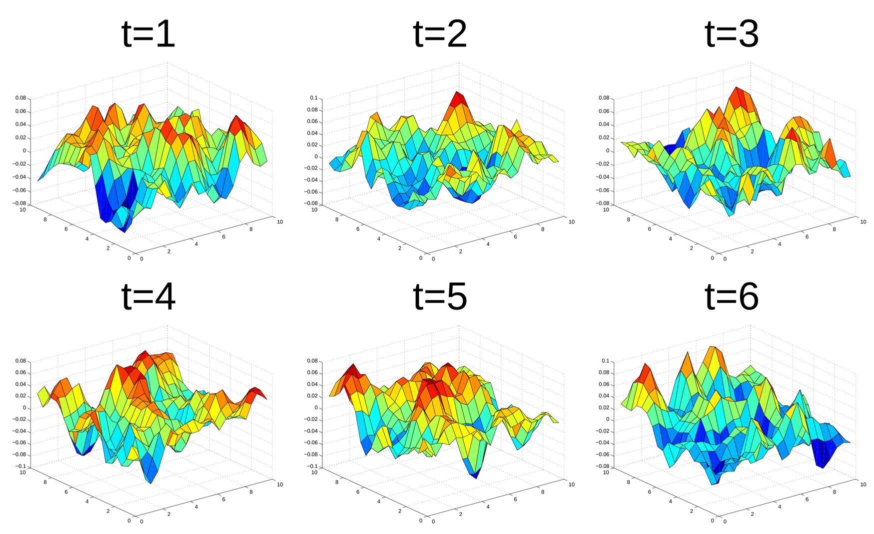

Figure 2: *Case 1. Simulated values of the spatiotemporal process of interest at times $t = 1,\ldots,6$.*

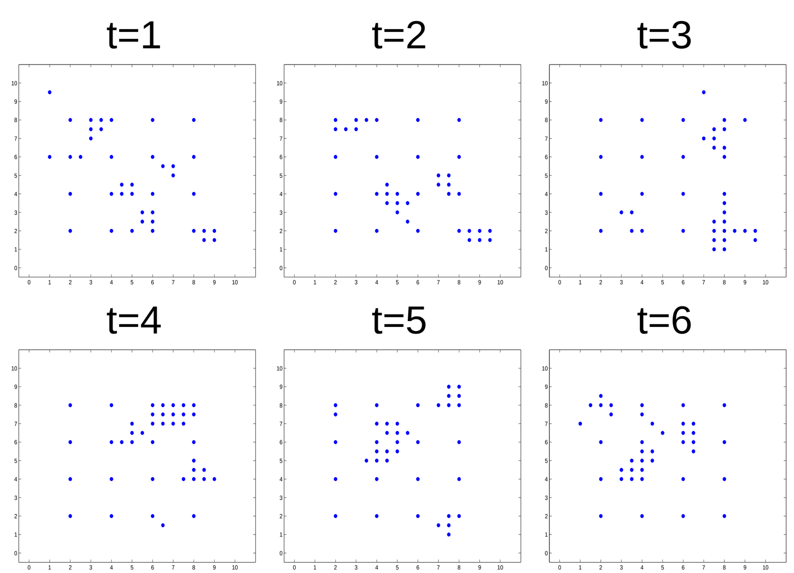

Figure 3: *Case 1. Locations of interest at times $t = 1,\ldots,6$.*

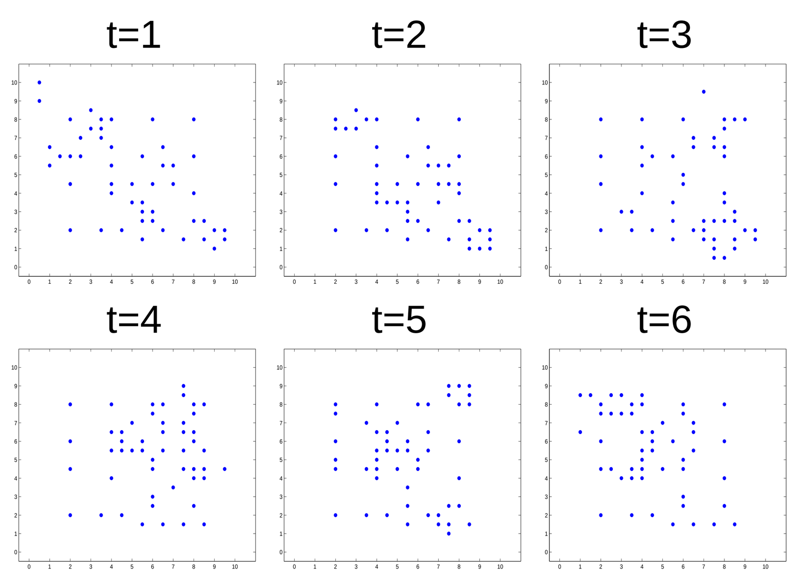

Figure 4: *Case 1. Sequentially selected network for times $t = 1,\ldots,6$, for the time-adaptive criterion.*

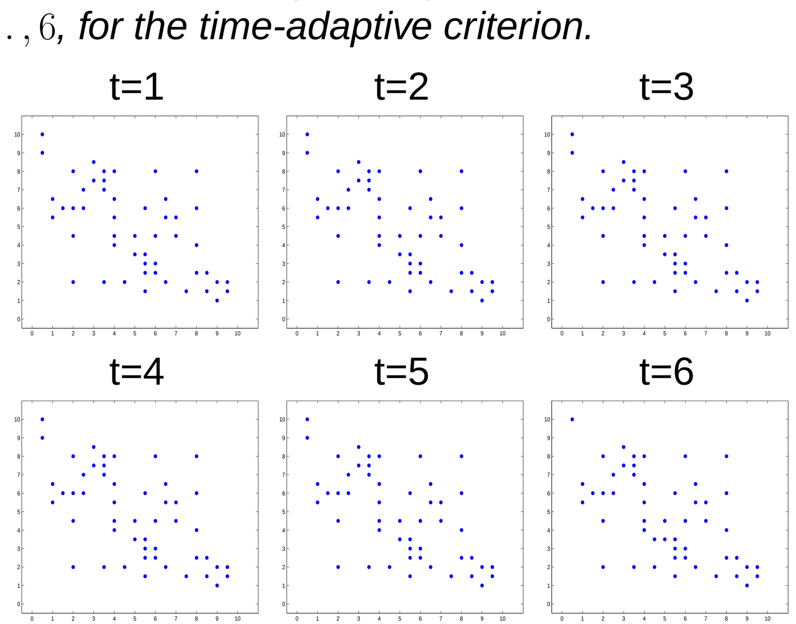

Figure 5: *Case 1. Sequentially selected network for times $t = 1,\ldots,6$, for fixed region of interest.*

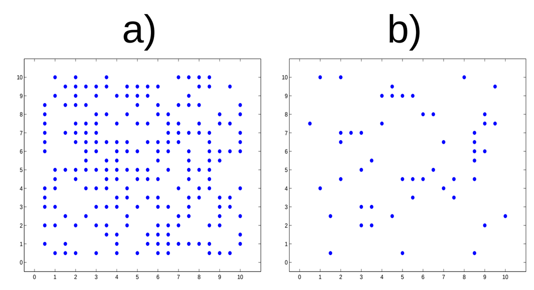

Figure 6: *Case 2. a) Set of candidate sites for observation, and b) starting observation set S_0.*

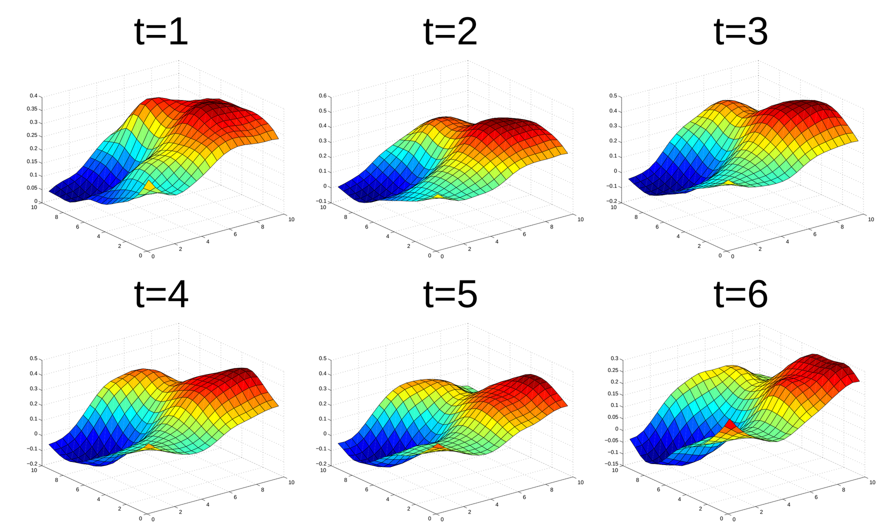

Figure 7: *Case 2. Simulated values of the spatiotemporal process of interest at times $t = 1,\ldots,6$.*

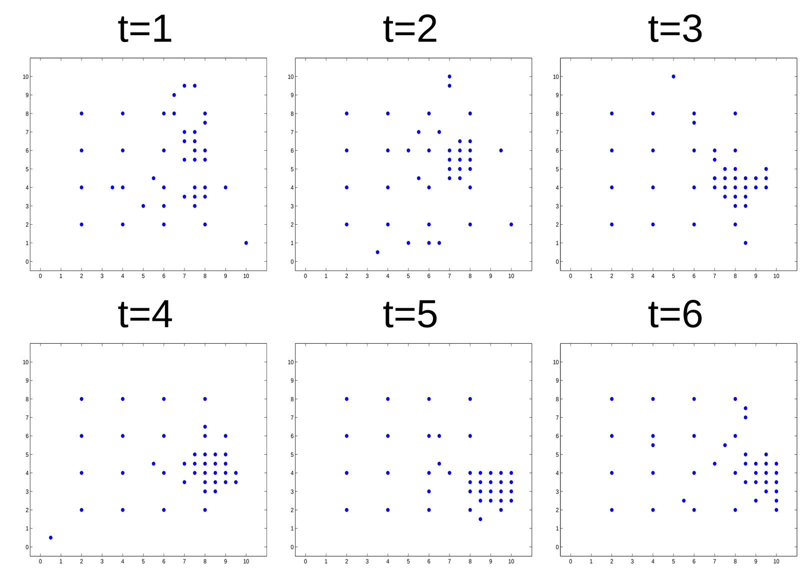

Figure 8: *Case 2. Locations of interest at times $t = 1,\ldots,6$.*

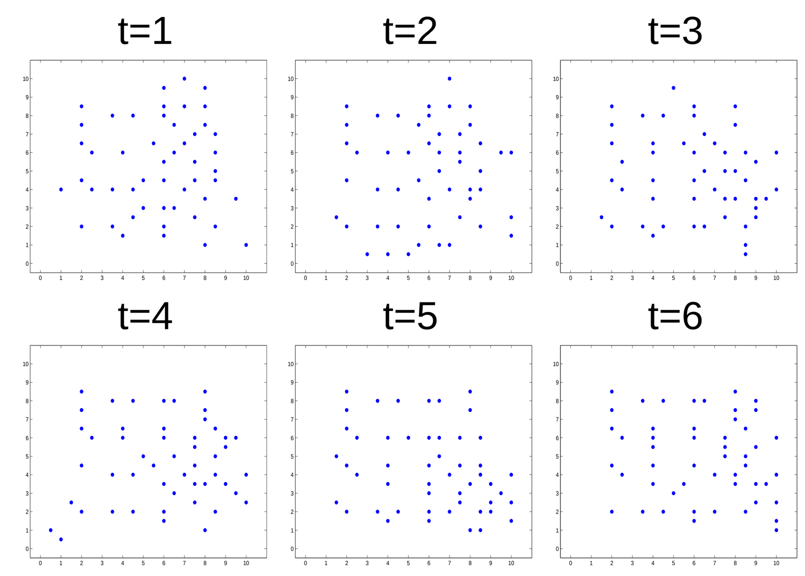

Figure 9: *Case 2. Sequentially selected network for times $t = 1,\ldots,6$, for the time-adaptive criterion.*

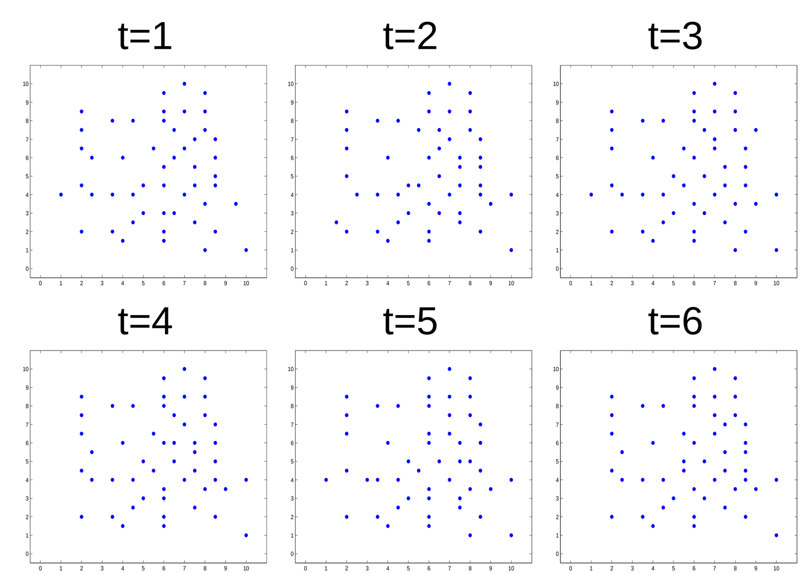

Figure 10: *Case 2. Sequentially selected network for times $t = 1,\ldots,6$, for fixed region of interest.*

3. References

Angulo, J.M., and Bueso, M.C. (2001). Random perturbation methods applied to multivariate spatial sampling design. *Environmetrics*, **12**, 631-646.

Angulo, J.M., Bueso, M.C., and Alonso F.J. (2000). A study on sampling design for optimal prediction of space-time stochastic processes. *Stochastic Environmental Research and Risk Assessment*, **14**, 412-427.

Angulo, J.M., Ruiz-Medina, M.D., Alonso F.J., and Bueso, M.C. (2005). Generalized approaches to spatial sampling design. *Environmetrics*, **16**, 523-534.

Bueso, M.C., Angulo, J.M., and Alonso, F.J. (1998). A state-space model approach to optimum spatial sampling design based on entropy. *Environmental and Ecological Statistics*, **5**, 29-44.

Le, N.D., and Zidek, J.V. (2006). *Statistical Analysis of Environmental Space-Time Processes*. New York: Springer.

Ruiz-Medina, M.D., and Angulo, J.M. (2007). Functional estimation of spatiotemporal heterogeneities. *Environmetrics*, **18**, 775-792.

Wikle, C.K., and Royle, J.A. (2005). Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics*, **16**, 507-522.

This work has been supported in part by projects P05-FQM-00990 of the Andalusian CICYE and MTM2005-08597 of the DGI, Spain.