Discussion of the presentation:
Comparing and Selecting Spatial Predictors Using Local Criteria

by Bradley, Cressie, and Shi

Monica Musio

University of Cagliari

Workshop
March 21-23 2013, Padova
Observation process given by

\[ Z(s) = Y(s) + \epsilon(s); \quad s \in D \]

where \( D = \{u_1, \cdots, u_N\} \subset \mathbb{R}^d \) is a spatial lattice, \( Y(\cdot) \) is a hidden process and \( \epsilon(\cdot) \) is a measurement-error process independent of \( Y(\cdot) \).

- Start from an arbitrary set of predictors: Traditional stationary kriging (TSK), smoothing splines (SSP), negative-exponential distance weighting (EDW), Fixed Rank Kriging (FRK), a modified predictive processes approach (MPP), a stochastic partial differential equation approach (SPD) and lattice kriging (LTK).

- Choose between the various predictors separately at each site, using possibly different predictors in different places. At each site some measure of predictive performance is optimised.
Local squared validation error:

\[
\text{LSVE}^{(k)}(u_i, Z^{\text{trn}}, H^{\text{val}}) = \frac{1}{|H^{\text{val}}(u_i)|} \sum_{s \in H^{\text{val}}(u_i)} (Z(s) - \hat{Y}^{(k)}(s, Z^{\text{trn}}))^2
\]

\[
k^*(u_i, Z^{\text{trn}}, H^{\text{val}}) = \arg\min \{\text{LSVE}^{(k)}(u_i, Z^{\text{trn}}, H^{\text{val}}), k = 1, \ldots, K\}
\]

\[
\hat{Y}^{\text{LSP}} = (\hat{Y}^{(k^*)}, i = 1, \ldots, N)'
\]

Locally selected predictor

Empirical deviance Information criterion (Bradley, Cressie and Shi (2013))
Very pragmatic approach, aiming to improve predictive performance

- These spatial predictors are often derived from different assumptions on the stochastic process, or may be deterministic.

- They can be regarded as approximations of the underlying random field.

- All these methods sacrifice some important information in the data in order to gain computational efficiency.
Selecting a single predictor from some class of predictors does not take prediction uncertainty into account.

Natural extension: instead of selecting a single predictor at each site, assign a weight to each predictor at each site and consider a new hybrid local predictor given by the weighted average of all local predictors:

$$\tilde{Y} = \sum_{j=1}^{K} \omega_j \tilde{Y}(j)$$

e.g. using Buckland et al. (1997) measure of predictive performance, $I_j$:

$$\omega_j = \frac{\exp(-I_j/2)}{\sum_{k=1}^{K} \exp(-I_k/2)}.$$
Another idea:

- in a Bayesian framework: use strictly proper scoring rules to evaluate probabilistic predictions in the form of a predictive distribution.

- Scoring rules provide a summary measure for evaluating a probabilistic prediction given the predictive distribution and the observed outcome.

- Can use different scoring rules for different purposes.

- Several different scoring rules can be used to compare locally the ability of candidate predictors.
Example:

- $Z(\cdot)$ is a discrete process, $\pi^{(k)}(s) = \{\pi_1(s), \ldots, \pi_J(s)\}$ is the posterior predictive distribution of predictor $k$ at $s$, $k = 1, \ldots, K$

- we can compare the predictors locally in terms of mean score

\[
\hat{S} = \sum_{s \in H_{val}(u_i)} \frac{S(z(s), \pi(s))}{|H_{val}(u_i)|}
\]

- Quadratic: $S(z, \pi) = \sum_{j=1}^{J} \pi_j^2 - 2\pi z$

- Spherical: $S(z, \pi) = -\frac{\pi_z}{(\sum_{j=1}^{J} \pi_j^2)^{1/2}}$

- Logarithmic: $S(z, \pi) = -\ln \pi_z$

(Dawid and Musio (2013); Finley, Banerjee and McRoberts (2009), Gneiting and Raftery (2007))
Continuous case: consider a local scoring rule

- **log score**

  \[ S(z, \pi) = -\ln \pi(z) \]

- **Hyvärinen score (Parry et al. (2012))**

  \[ \nabla = (\partial/\partial z^j), \ \Delta = \sum_{j=1}^k \partial^2 / (\partial z^j)^2 \]

  \[ S(z, \pi) = \Delta \ln \pi(z) + \frac{1}{2} |\nabla \ln \pi(z)|^2 = \frac{\Delta \sqrt{\pi(z)}}{\sqrt{\pi(z)}} \]

you can compare the predictors locally in terms of mean score

\[ \hat{S} = \sum_{s \in H^{\text{val}}(u_i)} \frac{S(z(s), \pi(s))}{|H^{\text{val}}(u_i)|} \]
The proposed method used only estimated means, not variances:

\[
\text{LSVE}^{(k)}(u_i, Z^{\text{trn}}, H^{\text{val}}) \propto \sum_{s \in H^{\text{val}}(u_i)} (Z(s) - \hat{Y}^{(k)}(s, Z^{\text{trn}}))^2
\]

If we have an estimate \( v(s) \) of the variance of \( Z \) at each validation site \( s \), we could use it to weight the terms:

\[
\text{LSVE}^{(k)}(u_i, Z^{\text{trn}}, H^{\text{val}}) \propto \sum_{s \in H^{\text{val}}(u_i)} \frac{(Z(s) - \hat{Y}^{(k)}(s, Z^{\text{trn}}))^2}{v(s)}
\]

— but this is not a proper scoring rule!

Instead, use

\[
\sum_{s \in H^{\text{val}}(u_i)} \frac{(Z(s) - \hat{Y}^{(k)}(s, Z^{\text{trn}}))^2}{v(s)} + \ln v(s)
\]

(Dawid and Sebastiani, 1999)

which is!
Bibliography