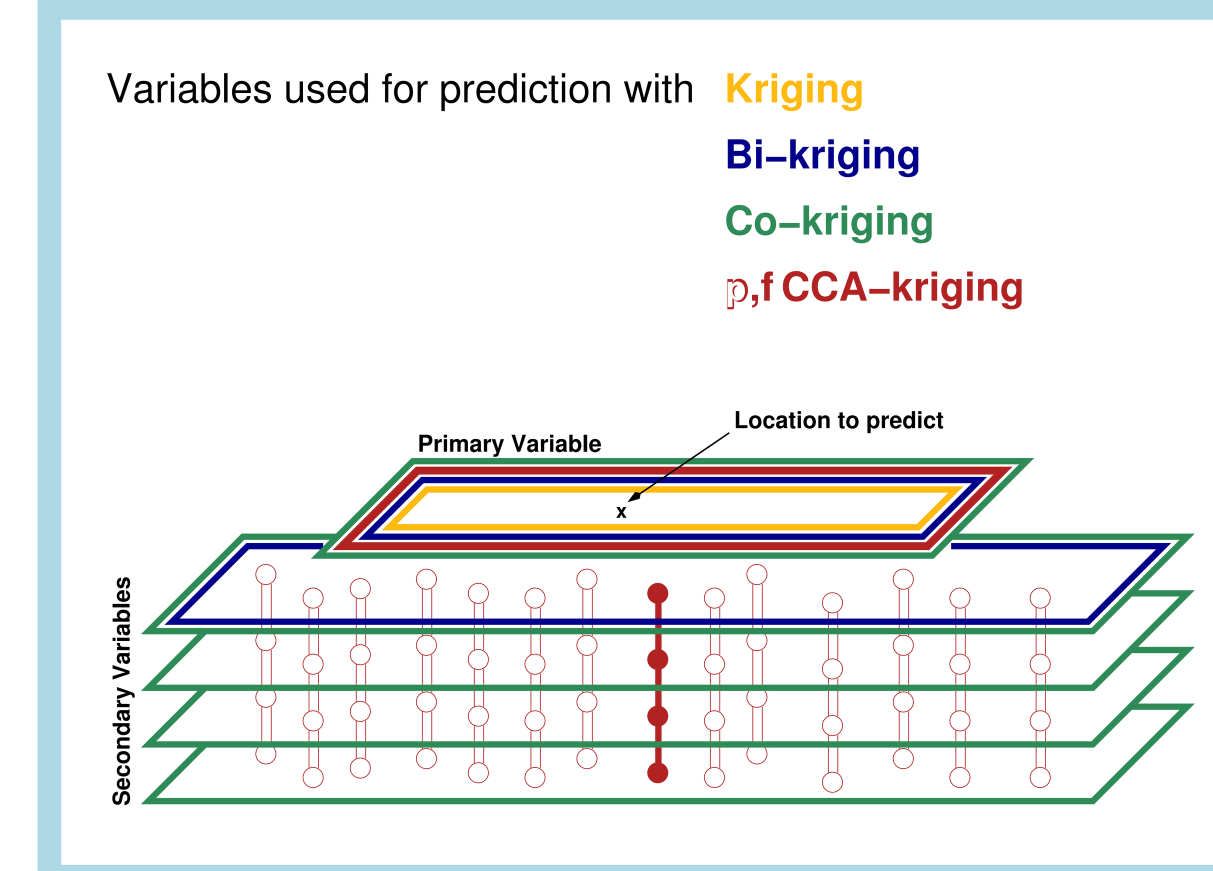
Efficient co-kriging approach based on canonical correlation analysis and tapering

REINHARD FURRER
Colorado School of Mines
rfurrer@mines.edu

MARC G. GENTON
University of Geneva
marc.genton@metri.unige.ch



Abstract: The best unbiased linear predictor (BLUP) of a spatially correlated random process, often called kriging in geostatistics, requires the solution of a large linear system based on the covariance matrix of the observations. Incorporating the presence of covariables leads to the so-called co-kriging approach which requires solving unreasonably large linear systems for virtually all problems interesting in practice. Starting from the mid eighties, several approximations and simplifications have been proposed to lessen the computational burden. Yet, these methods can not cope with the ever increasing data sizes. We propose an efficient suboptimal linear unbiased prediction based on canonical correlation analysis (CCA) and tapering. CCA is used to synthesize the covariables to a single meta-covariable. To perform the resulting "bi-kriging" efficiently, we taper the covariance matrices with an appropriate compactly supported covariance function creating a sparse linear system that can then be solved using sparse matrix algorithms. It can be shown that tapering reduces the computational burden significantly and the resulting predictor still has an asymptotic optimal mean squared error. We compare and contrast the proposed approach with other methods used in practice.

CCA-Kriging

Objective: Construct weight vectors to form a linear combination of the secondary variables, which will be used to predict a single primary variable at arbitrary locations.

The MSPE of this approach lies between those of kriging and co-kriging. Optimally, it should be smaller than any bi-kriging MSPE (kriging with the primary and one secondary variable).
Besides calculating the weight vector(s), CCA-kriging is computationally identical to bi-kriging.

pCCA-kriging: $\max_{(a_1,\dots,a_\ell)} \operatorname{Corr}\Big(Y_0(\boldsymbol{s}_0), \sum_{i=1}^{\ell} a_i Y_i(\boldsymbol{s}_j)\Big)$

- Solution corresponds to ordinary CCA.
- The weight vectors depend on the point to predict $\boldsymbol{s}_0$ and on the location of the secondary variable $\boldsymbol{s}_j$ ("p" for point-wise).
- Resulting aggregation fields are difficult to interpret.
- Each vector can be arbitrarily scaled.

fCCA-kriging: $\max_{(a_1,\dots,a_\ell)} \Big\| \operatorname{Cov}\Big(Y_0(\boldsymbol{s}_0), \sum_{i=1}^{\ell} a_i \boldsymbol{Y}_i\Big)\Big\|$ s.t. $\boldsymbol{a}'\boldsymbol{a} = 1$

- Some norm is required because $\operatorname{Cov}(Y_0(\boldsymbol{s}_0), \boldsymbol{Y}_i) = \boldsymbol{C}_{0i}$ is not a square matrix.
- Solution is the eigenvector associated to the largest eigenvalue of $[\boldsymbol{T}]_{ij} = \operatorname{tr} \boldsymbol{C}_{i0}\boldsymbol{C}_{0j}$.
- The weight vector depends on the point to predict but not on $\boldsymbol{s}_j$ ("f" for the entire field).
- Is computationally more efficient compared to pCCA-kriging.

When all the secondary variables have the same covariance structure, optimality criteria have been derived.

Illustration: Gaussian process on the $[0,1]$ transect with equispaced observations (18 for Y_0 and 40 for Y_i) and spherical covariance structure (Fig. 1). For prediction at 0.5, we calculate:

(a) the weight vectors for pCCA- and fCCA-kriging (Fig. 2);
(b) the covariances of the weighted fields (Fig. 3);
(c) the MSPE of the different kriging methods (Table 1).

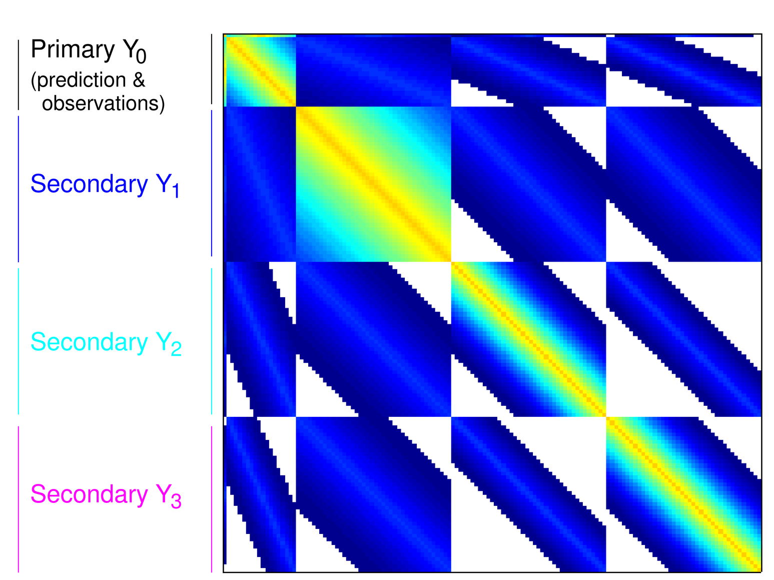


Figure 1: Covariance structure used in the illustration.

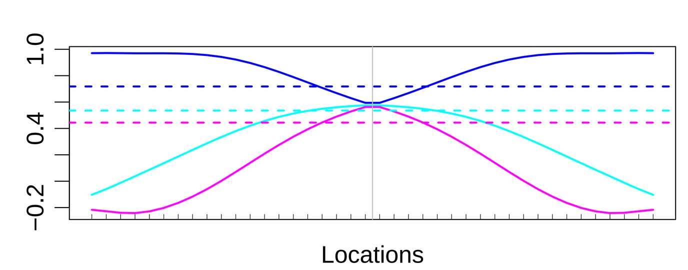


Figure 2: Weights for pCCA (solid) and fCCA (dashed).

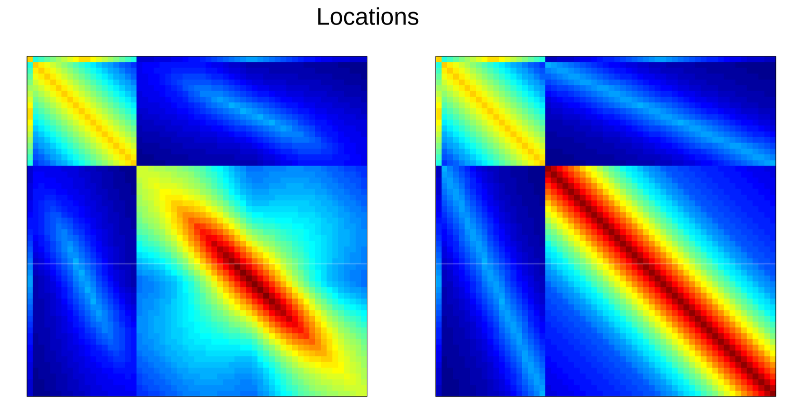


Figure 3: Induced covariance matrices by pCCA- and fCCA-kriging (same scale as Figure 1).

Table 1: MSPE ($\times 10^2$) for predicting at 0.5.

Kriging	simple	bi (Y_1)	bi (Y_2)	bi (Y_3)	pCCA	fCCA	co
MSPE	2.4514	2.3244	2.2809	2.1342	2.1337	2.1437	2.0965

Tapering

Objective: Reduce the computational burden of prediction while maintaining an asymptotically optimal mean squared error.

The covariances are multiplied with a correlation function with finite support (the taper function). This operation is called tapering.

Optimality is based on Stein's "misspecified covariance" idea. We need to impose differentiability conditions of the taper function at the origin.

It is also possible to use tapering within the estimation procedure.

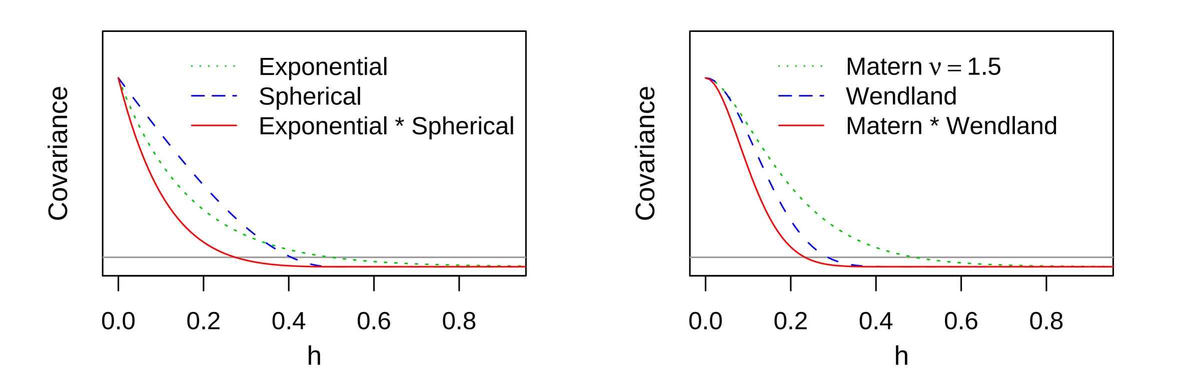


Figure 4: The effect of tapering is to create a sparse approximate covariance matrix for use of sparse matrix algorithms.

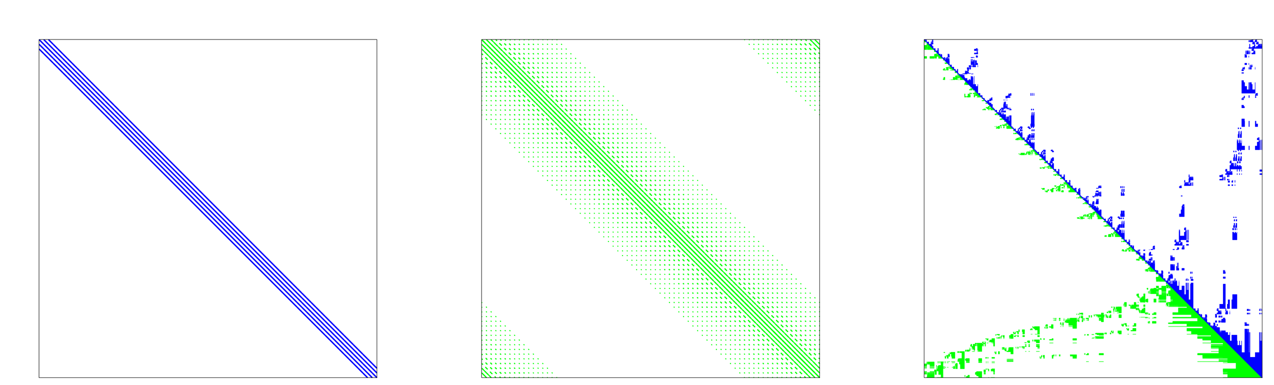


Figure 5: Sparseness structure of the tapered covariance matrix for a climatological field with $5^\circ \times 5^\circ$ lat-lon resolution (left with Euclidian distance, middle with great circle distance; taper ranges are 15° and 10°). Right: corresponding Cholesky factors with a min-degree ordering.

Application

Objective: Construct monthly temperature fields for pre-September 1957 using HadCRUT3 gridded temperature data (primary variable) and model temperature data from 20 climate models (secondary variable).

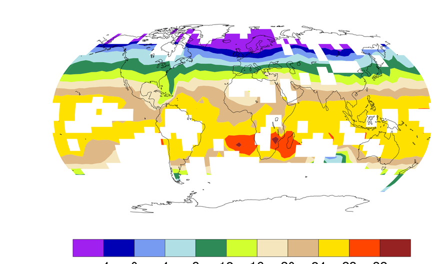
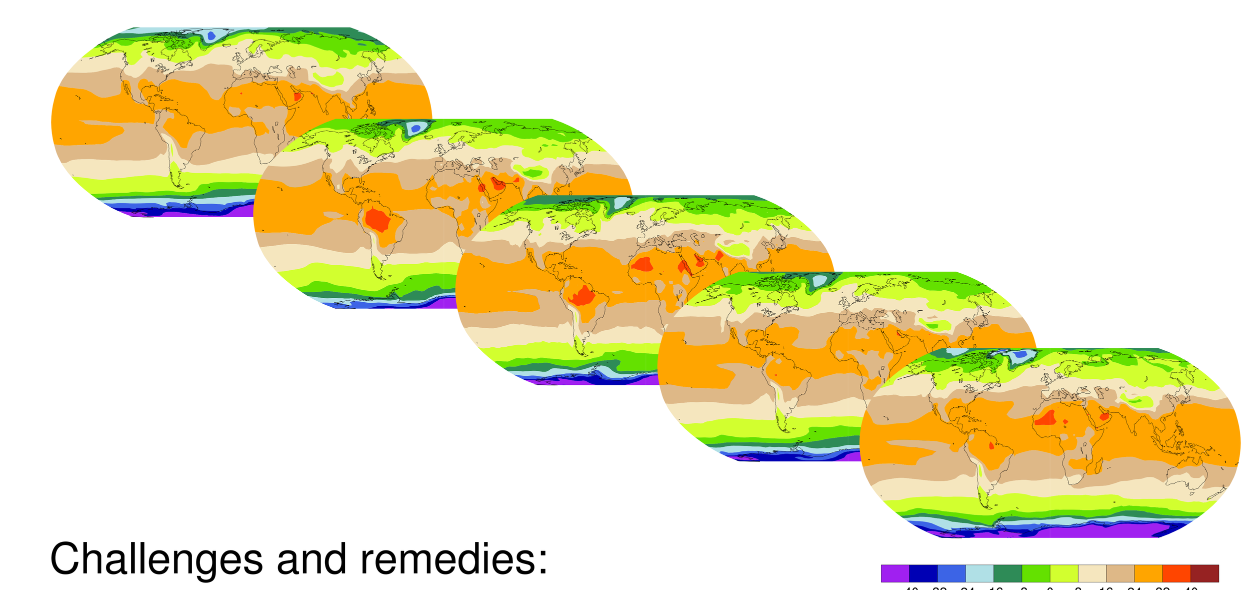


Figure 6: Left: Observations for Sept. 1957 ($5^\circ \times 5^\circ$). Below: Sept. 1957 model output from five GCMs ($2.8^\circ \times 2.8^\circ$ lat-lon resolution).



Challenges and remedies:

- Highly non-stationary fields:
  ⇝ "Moving window" type covariance estimator.
- Sparse primary variable and extrapolation may be critical:
  ⇝ Mean functions have to be well chosen.
  ⇝ Use post-1957 NCEP fields for training and verifcation.
- 7936 observation for each of the 20 secondary variables:
  ⇝ We only use a subset of secondary variables: CCCMA-CGCM3.1, GFDL-CM2.1, MPI-ECHAM5, NCAR-CCSM3.0, UKMO-HadCM3
  ⇝ With appropriate tapering, say 25 observations in the taper range, we are able to work with the matrices.

Typeset with pdfLaTeX, Reinhard Furrer, 2006-10-20