Optimal interpolation of inhomogeneous fields using neural networks

Ph.L.Bykov
Hydrometeorological Research Center of Russian Federation, e-mail: bphilipp@inbox.ru.

Introduction. In this note we are concerned with the problem of interpolation of a random field $f(\vec{x}), \vec{x} \in \Omega$ from a scattered set of SYNOP stations $\vec{x}_j, j = 1...m$ onto another set of points $\vec{y}$ in a domain Ω on Earth surface. We will be using the moments of the field $f(\vec{x})$: the mean $\mu(\vec{x})$, the spread $\sigma(\vec{x})$, and the correlation function (CF) $K(\vec{x}, \vec{y})$. It is also natural to assume that the interpolation result $\hat{f}$ depends linearly on the interpolated data:

$$f(\vec{y}) \approx \hat{f}(\vec{y}) = w_x(\vec{y})^T [f_x - \mu_x] + \mu(\vec{y}), \tag{1}$$

where $w_x(\vec{y})$ is the (unknown) column of the interpolation weights we are looking to compute, f_x is column of known values of f. We can state the optimization problem as a problem of minimizing the mean interpolation loss *L*:

$$L(\mu, \sigma, K) = \sum_{j=1}^{m} e\left(f(\vec{x}_j), \hat{f}_j(\vec{x}_j)\right) \to \min_{\mu,\sigma,K}, \tag{2}$$

where e is the loss function and $\hat{f}_j$ is the interpolating operator computed from the same set of data without taking into account the value $f(\vec{x}_j)$. The classical optimal interpolation method [1] uses the mean squared error as loss function and computes the interpolation weights from the equation:

$$K_{xx}\left[\sigma_x \circ w_x(\vec{y})\right] = K_{xy} \circ \sigma(\vec{y}),$$

where K_{xx} – the positive definite $m \times m$ correlation matrix, $\circ$ stands for the Hadamard product.
The problem of estimating the correlation function *K* is easy to solve for homogeneous and isotropic fields *f*, i.e. in the case $K(\vec{x}, \vec{y}) = K(|\vec{x} - \vec{y}|)$. The case of inhomogeneous fields that we address here is more complicated. It follows from Mercer's theorem [2], [3] that the feature mapping $g : \Omega \to H$ exists for any field *f*, where H is Hilbert space. Furthermore, the feature mapping *g* turns inhomogeneous anisotropic field *f* into a homogeneous isotropic field in a Hilbert space *H*:

$$K(\vec{x}, \vec{y}) = K_g\left(\left\|g(\vec{x}) - g(\vec{y})\right\|_H\right).$$

Several papers [4], [5], [6] suggest methods for approximation of the feature mapping *g* and maximization of the logarithm of the likehood in Gaussian covariance model:

$$L_{Gauss}(g, K_g) = -\frac{1}{2}\left[\bar{f}_x^{\,T} K_{xx}^{-1} \bar{f}_x + \ln\det K_{xx}\right] \to \max_{g, K_g}, \tag{3}$$

where $\bar{f} = (f - \mu)/\sigma$ is the normalized field. However, this choice is questionable, since the interpolation function (1) and the functional (2) depend only on the local oscillations of the means $\mu(\vec{x}) - \mu(\vec{y})$ and the spreads $\sigma(\vec{x}) / \sigma(\vec{y})$, while the functional (3) depends on their absolute values.

Approach. Let $\Theta(\vec{x})$ be a set of predictors of inhomogeneity of the interpolated field *f*. For minimizing the interpolation loss (2) we shall apply the backpropagation method [7] to the following functions, described by the neural networks and parameters:

1. The feature mapping *g* to the 4-dimensional space *H* as a graphic of the neural network $\tilde{g}$:
$$g(\vec{x}) = (\vec{x}, \tilde{g}(\Theta));$$
2. The mean $\mu(\vec{x}) = \mu(\Theta(\vec{x}))$ and the spread $\sigma(\vec{x}) = \sigma(\Theta(\vec{x}))$ as the neural networks;
3. The parameters $\vec{\upsilon} = (\varepsilon, \beta, R_1, R_2)$ of family of CFs:
$$K_g(r, \vec{\upsilon}) = \varepsilon I_0(r) + (1 - \varepsilon)\left[(1 - \beta)(1 + r/R_1)\exp(-r/R_1) + \beta \exp\left(-r^2/2R_2^2\right)\right],$$
where $r = |g(\vec{x}) - g(\vec{y})| = \sqrt{|\vec{x} - \vec{y}|^2 + |\tilde{g}(\vec{x}) - \tilde{g}(\vec{y})|^2}$ – the distance in extended space *H*.

All neural networks we consider are two layers perceptrons with the ReLU activation function and 32 neurons at hidden layer. The Huber loss function [8] is chosen.

Examples. We shall demonstrate our approach for two fields of particular interest: 1) the 2-meter temperature *T2m* model biases; 2) the snow depth *SD*. We consider the following set of predictors of inhomogeneity $\vec{\ }$:

A. The first guess field (for *T2m* only; the COSMO-Ru model [9] forecast with lead time 6 hours);
B. The Earth surface altitude;
C. The sine and cosine of the Julian day $2\pi t / T_y$, where t is time, $T_y = 1$ year;
D. The elevation angle of the Sun.

The feature mapping g has to be injection and should take into account all predictors of inhomogeneity Θ.

Results. Adding additional dimensions to the feature mapping allows us to explain most part of the variance of the fields *T2m* and *SD*: the $\lim_{\vec{y}\to\vec{x}} K(\vec{x},\vec{y})$ is significantly larger (see figure). Moreover, the presented method gives more accurate and more detailed fields: since due to our choice of the feature mapping the CFs decrease much faster. Finally, the backpropagation for interpolation losses allows us to avoid extra assumptions on the field *f*.

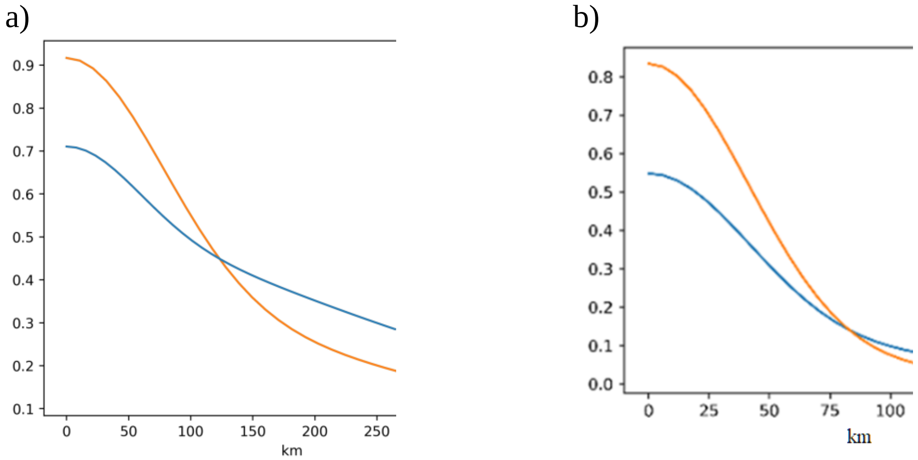

Figure. The CFs for a) *T2m* model biases and b) snow depth *SD* as the functions of the distance (x-axis). Orange curves correspond to the extended 4-dimensional space *H* and the blue curves correspond to the real space with extra assumptions of homogeneity and isotropy.

This work is done as part of the PP MILEPOST project (COSMO WG4). This research is funded by the Federal Service for Hydrometeorology and Environmental Monitoring of Russia (topic AAAA-A20-120021490079-3).

References

[1] Gandin L.S., Objective Analysis of Meteorological Fields, Jerusalem, 1965, p. 896.

[2] Mercer J., "Functions of positive and negative type and their connection with the theory of integral equations," *Philosophical Transactions of the Royal Society A*, vol. 209, no. 441-458, pp. 415-446, 1909.

[3] Aronszajn N., "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, no. 3, pp. 337-404, 1950.

[4] Sampson P., Guttorp P., "Nonparametric estimation of nonstationary spatial covariance structure," *J. Amer. Statist. Assoc.*, vol. 87, p. 108–119, 1992.

[5] Bornn L., Shaddick G., Zidek J.V., "Modeling Nonstationary Processes Through Dimension Expansion," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 281-289, 2012.

[6] Zammit-Mangion A., Tin L. J. Ng, Vu Q., Filippone M., "Deep Compositional Spatial Models," 2019. https://arxiv.org/pdf/1906.02840.pdf.

[7] Rumelhart D.E., Hinton G.E., Williams R.J., "Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing," vol. 1, Cambridge, MIT Press, 1986, pp. 318-362.

[8] Huber P.J., "Robust Estimation of a Location Parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73-101, 1964.

[9] Rivin G.S., Rozinkina I.A., Vil'fand R.M., Alferov D.Yu., Astakhova E.D., Blinov D.V., Bundel' A.Yu., Kazakova E.V., Kirsanov A.A., Nikitin M.A., Perov V.L., Surkova G.V., Revokatova A.P., Shatunova M.V., Chumakov M.M., "The COSMO-Ru system of nonhydrostatic mesoscale short-range weather forecasting of the Hydrometcenter of Russia: The second stage of implementation and development" *Russ. Meteorol. Hydrol.*, vol. 40, no. 6, pp. 400-410, 2015.