Technical Vignette 3: Kriging, interpolation, and uncertainty

Christopher Paciorek, Department of Biostatistics, Harvard School of Public Health

Version 1.0: January 2008

There are two factors that determine the interpolation behavior of the kriging predictions in terms of whether the predictions go through the observations.

First, if the nugget is zero, then the kriging predictions go exactly through the observations. The kriging variance is zero at the observations, because with no residual, we know the value of the spatial process exactly. We see an example of this in the Fig. 1 (top), which shows prediction in one dimension.

Second, if the nugget is non-zero, whether kriging interpolates depends on whether the residual variation is considered to be instrument (measurement) error, which one would want to smooth over, or fine-scale (microscale) variability, in which case the prediction at an observation should be the observation itself, since it is observed without error. In the microscale case, we assume that slight changes in location result in residuals that are in practice independent spatially, because the scale of variation is smaller than the distances between observations. For atmospheric phenomena, measurement may be quite precise (pollution, temperature, etc.), but slight changes in location can cause big changes in the outcome. For example, if one moves to a different side of a building or from a ridge to a valley, pollution or temperature may change quickly and at a scale at which we cannot predict because of sparse observations. See Cressie (1993, p. 59) for more discussion.

Fig. 1 (middle) shows predictions and variances under microscale variability, showing that if one moves slightly away from the location of the observation, the prediction looks like a smoother and the variance increases very quickly. This is because we assume there is a component of variance that comes into play as soon as we are not exactly at the location of the observation. Kriging behaves very much like a smoother, except with jumps at the observations which produce plots that show these 'holes' at the observations. The predictions are discontinuous. Kriging 'honors the data' only in the limited sense of including a discontinuity at each observation.

In Fig. 1 (bottom), we see the predictions and variances under measurement error. Now the residual is treated as error, and our prediction at the observation is not the observation itself, because the observation is contaminated with error. We borrow strength from other observations to try to make a better prediction at any given observation, hoping to average over the error.

Note that in the latter two cases, if we seek to make predictions about the smooth latent process, rather than the observation process, then the predictions will be the same, as the prediction under microscale variation will predict the process without the fine-scale variation.

One way to think about the kriging model is to decompose the observations as $Z_i = \mu(s_i) + g(s_i) + w(s_i) + \epsilon_i$ where $\mu(\cdot)$ is large-scale trend, often taken to be a constant, $g(\cdot)$ is a smooth

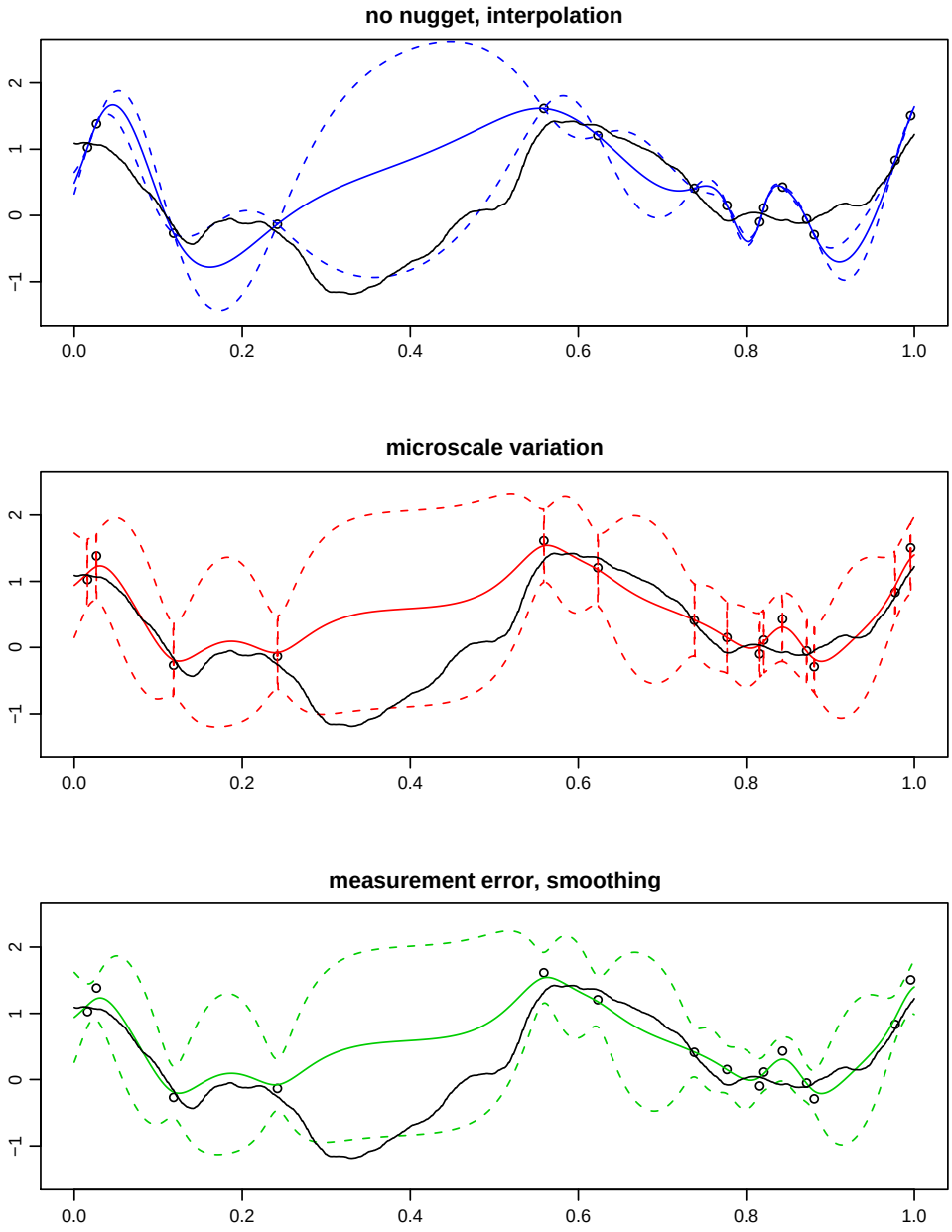

Figure 1. Kriging predictions under three scenarios. In each scenario, the black line is the true underlying smooth process, generated from the sum of a very smooth process and a more local, less smooth process, while the observations (the dots) also include a component of independent errors. The solid colored lines are the kriging predictions and the dashed lines pointwise confidence intervals. In (top), we assume a nugget of zero, which fits an interpolator. In (middle) we assume the residual is microscale variation, which causes the predictions to have holes where the predictor jumps from smoothing to the actual observation. In (bottom), the residual is assumed to be instrument error, so kriging serves as a smoother.

Table 1. Kriging predictions and variances under the three situations..

	$\tau^2 = 0$	Microscale variation	Instrument error
$E(Z_2\|Z_1)$	$\mu 1 + C_{21}C_{11}^{-1}(Z_1 - \mu 1)$	$\mu 1 + C_{21}(C_{11} + \tau^2 I)^{-1}(Z_1 - \mu 1)$	same as microscale
$E(Z_1^*\|Z_1)$	$Z_1 = \mu 1 + C_{11}C_{11}^{-1}(Z_1 - \mu 1)$	$Z_1 = \mu 1 + (C_{11} + \tau^2 I)\cdot$ $(C_{11} + \tau^2 I)^{-1}(Z_1 - \mu 1)$	$\mu 1 + C_{11}(C_{11} + \tau^2 I)^{-1}(Z_1 - \mu 1)$
$E(f_2\|Z_1)$	$\mu 1 + C_{21}C_{11}^{-1}(Z_1 - \mu 1)$	$\mu 1 + C_{21}(C_{11} + \tau^2 I)^{-1}(Z_1 - \mu 1)$	same as microscale
$E(f_1\|Z_1)$	$Z_1 = \mu 1 + C_{11}C_{11}^{-1}(Z_1 - \mu 1)$	$\mu 1 + C_{11}(C_{11} + \tau^2 I)^{-1}(Z_1 - \mu 1)$	same as microscale
$\text{Var}(Z_2\|Z_1)$	$C_{22} - C_{21}C_{11}^{-1}C_{12}$	$C_{22} + \tau^2 I - C_{21}(C_{11} + \tau^2 I)^{-1}C_{12}$	same as microscale
$\text{Var}(Z_1^*\|Z_1)$	$0 = C_{11} - C_{11}C_{11}^{-1}C_{11}$	$0 = C_{11} + \tau^2 I - (C_{11} + \tau^2 I)\cdot$ $(C_{11} + \tau^2 I)^{-1}(C_{11} + \tau^2 I)$	$C_{11} + \tau^2 I - C_{11}(C_{11} + \tau^2 I)^{-1}C_{11}$
$\text{Var}(f_2\|Z_1)$	$C_{22} - C_{21}C_{11}^{-1}C_{12}$	$C_{22} - C_{21}(C_{11} + \tau^2 I)^{-1}C_{12}$	same as microscale
$\text{Var}(f_1\|Z_1)$	$0 = C_{11} - C_{11}C_{11}^{-1}C_{11}$	$C_{11} - C_{11}(C_{11} + \tau^2 I)^{-1}C_{11}$	same as microscale

process that can be estimated from the data and has variance σ^2, $w(\cdot)$ is the fine-scale process with variance τ^2_{MS}, and ϵ_i is measurement error with variance τ^2_{ME}. In this model, the nugget is $\tau^2 = \tau^2_{\text{MS}} + \tau^2_{\text{ME}}$ and cannot be decomposed without further information from replication or prior belief. The decomposition of μ and g is not unique without replication, and all we are doing is decomposing the variation at different scales. My personal belief is there is often no real need to do anything with μ other than have it be a constant mean that is estimated from the data. By definition from having $w(\cdot)$ be independent in space, we cannot predict w, except at the observed locations without more dense data, just as the residuals, ϵ, are unknowable at new locations.

If we assume $\tau^2 = \tau^2_{MS}$ then ideally we wish to predict $\mu + g + w$, whereas if we assume τ^2_{ME} then we are trying to predict $\mu + g$. Of course even if we believe $\tau^2 = \tau^2_{\text{MS}}$ we can always choose to predict $\mu + g$ and acknowledge that we are just trying to predict the potentially predictable part of the variation, ignoring w as being too fine-scale to resolve with the density of our data.

Note that for prediction uncertainty, whether we include τ^2 in the prediction variance depends on whether we want uncertainty about a new observation, in which case it should always be included, or prediction about the smooth process, in which case if we assume measurement error it should not. If we assume microscale variation and want prediction about the full process then we would include τ^2.

In the Table 1, I lay out the kriging predictions and variances under various scenarios. Define Z_1 as the observations, Z_1^* as new observations at the same location, and Z_2 as new observations at new locations. $f_1 = \mu 1 + g_1$ are the smooth process values at the observed locations and $f_2 = \mu 1 + g_2$ are the smooth process values at new locations. C_{11} is the covariance of g_1 and C_{22} the covariance of g_2, while $C_{12} = C_{21}^T$ is the covariance of g_1 and g_2. The kriging prediction is always $\mu 1 + \Sigma_{21}\Sigma_{11}^{-1}(Z_1 - \mu 1)$ and kriging variance, $\Sigma_{22} + \Sigma_{21}^{-1}\Sigma_{12}$. Our challenge here is to determine in which cases $\tau^2 I$ is added to C to calculate Σ for the various covariance matrices and in which case $\Sigma = C$.

The only difference between assuming microscale variation and measurement error is that for $E(Z_1^*|Z_1)$ under measurement error, we have $\mu 1 + C_{11}(C_{11} + \tau^2 I)^{-1}(Z_1 - \mu 1)$ because $\Sigma_{21} = Cov(Z_1^*, Z_1) = C_{11}$, since new observations will not be the same as the old observations and will not share the τ^2 variance component. In contrast under microscale variation, $\Sigma_{21} = Cov(Z_1^*, Z_1) = C_{11} + \tau^2 I$. Similarly for the prediction variance, under measurement error,

we have $C_{11} + \tau^2 I - C_{11}(C_{11} + \tau^2 I)^{-1} C_{11}$ because of the different construction of $\Sigma_{21} = \Sigma_{11}$ under the two scenarios.

If $\tau^2 = 0$ then we assume that there is no non-spatial variability in the process, so uncertainty is by definition about the smooth process, which is known exactly at the observed locations.

In geoR, using krige.conv(), one can specify that you want to predict the smooth signal only using the argument `output=output.control(signal=TRUE)`. This will give prediction variances without the component $\tau^2 I$, i.e., the last two rows of Table 1. I believe the default in geoR is `signal=FALSE`, so the prediction variance includes τ^2. This is appropriate under microscale variation if one wants the full process uncertainty, or under measurement error, if one wants prediction variance for new observations. If one wants the smooth process uncertainty under either assumption, choose `signal`=TRUE.

In ArcGIS, you can select error modeling in the variogram step and choose 100% measurement error to get the measurement error scenario. However, I'm not sure if this results in the component $\tau^2 I$ being included in the variance. I believe choosing the default of 100% microscale includes $\tau^2 I$ in the variance, while 100% measurement error omits this component. If you're using software and don't know what assumption it is making in doing the predictions and the variances, make a prediction (and get the variance) at an observation and very close to it and find out if the prediction at the observation is the observation itself and if the prediction close to the observation behaves like a smoothed prediction. Is the variance zero at the observation?

If one gets prediction variances that do not include τ^2, which are just uncertainty about the smooth process, one can always add τ^2 to the output from the software to get prediction variances.

So what should one do in practice? First, I would generally estimate a nugget. If estimated to be zero, then one naturally gets an interpolator. If the nugget is estimated to be non-zero, one has a choice to make. If one can get data to estimate what portion of the nugget is microscale and what portion error, one could be sophisticated about it. Otherwise, since the results are very similar, one might just assume error and take the smoothing approach. Then if one does want to make predictions at the observations, one can just substitute the observations and take the variance to be zero. However, one shouldn't deceive oneself and think that by assuming microscale variation that one is able to predict variation at fine scale. Our predictions at ANY other locations, regardless of how close they are to an observation, are smoothed predictions because τ^2 does not contribute to any covariance between the observation and other locations. This inability is indicated in Fig. 1 (middle). In applications I generally will report the smoothed predictions in all cases, even at the observations. My reporting of uncertainty varies with the application and my assumptions about measurement error or microscale variation.

References

Cressie, N. (1993), *Statistics for Spatial Data* (Rev. ed.), New York: Wiley-Interscience.