Technical Vignette 4: When can we ignore temporal correlation in space-time data?

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

Version 1.0: January 2008

There has been a lot of research recently on space-time modeling, including fitting space-time covariance models. For prediction, such models smooth in both space and time, borrowing strength to make predictions. While fitting a full space-time model is an appropriate goal, as statisticians, part of our job is to focus on the key features of the data. I believe that in certain circumstances, even if the data are temporally correlated, one can focus simply on modeling the spatial structure at each discrete point in time with little loss in predictive ability.

First, I'll describe the technical conditions needed, and then I will discuss the practical situations in which these conditions might be expected to hold approximately. Assume the following conditions: (1) a space-time covariance model with separable structure, (2) the same locations sampled at every time point, (3) all error assumed to be local heterogeneity rather than instrument error, and (4) normal residuals (i.e., the variance of these residuals is the nugget). If these assumptions hold, the best prediction for a new set of locations at a given time is based only on measurements from that same time, with the predictions conditionally independent of data at other times given the data at the time of interest. For simplicity assume the overall mean is known to be zero. The space-time kriging prediction under those assumptions is

$$(C_t \otimes C_{21})(C_t \otimes C_{11})^{-1}Y = (C_t C_t^{-1}) \otimes (C_{21}C_{11}^{-1})Y = I_T \otimes (C_{21}C_{11}^{-1})Y,$$

,which does not involve the temporal covariance, C_t, but only the spatial covariance between prediction and observed sites, C_{21}, and the spatial covariance amongst observed sites, C_{11}, which can include a nugget for local heterogeneity. See Cressie (1993, p. 59) for a discussion of the distinction. However, the model cannot include instrument error and preserve this property. Similar calculations demonstrate that the kriging variances (but not the covariances) are based only on the spatial covariance structure: $C_t \otimes (C_{22} - C_{21}C_{11}^{-1}C_{21}^T$, where C_{22} is the spatial covariance amongst the prediction sites. Also note that if one wants to average over multiple months, one will not get the right variances for the averages, unless one accounts for the temporal autocorrelation.

Now let's dicuss the practical implications of these assumptions. Assumption (1) is of a separable covariance. Separability means that the space-time covariance is the product of a spatial covariance and a temporal covariance and is equivalent to the following Markov property (O'Hagan, A., A Markov Property for Covariance Structures, unpublished but easily findable through Google). An observation is independent of observations at other times and locations given observations co-occurring in time and observations co-occurring in space. If there is no transport or dynamics in the phenomenon, then we might expect that all the relevant information for prediction at a given space-time point comes from other points at the same time and the same point at other times. However, if there is transport then the observation at a given space-time point might be highly correlated with an observation at another location at a different time. For example, wind would blow pollution from one location to another over an appropriate amount of time. In general, for many natural phenomena, separability may be a reasonable assumption provided one has averaged over a long enough period of time such that short-term dynamics do not come into play. For example for atmospheric processes, separability may not be reasonable for hourly or daily data, but may well be reasonable for monthly or yearly data. My heuristic for thinking about separability is to think about whether such dynamics may be important. A special case of this is if the temporal covariance is weak relative to the spatial covariance, which would suggest we can rely on spatial smoothing and ignore temporal structure. The separability assumption is equivalent to assuming a Markov property

Assumption (2) is that the same locations are sampled at all time points. This assumption means that there are no locations that are sampled at one time and not at another. If there were, we would want to borrow strength across time to fill in the missing time points. But if there are not, then there is no interpolation in time, only in space. We have complete coverage in time, so the information available at a sampled location is the best information we have about that location. Note that if assumption (3) of local heterogneity did not hold, then we would want to borrow strength across time to smooth over measurement error. Also note that if observations are available for a block of time, and then unavailable for a block of time, there is little information for smoothing over time, except at the time points immediately after sampling ceases, assuming the temporal correlation declines quickly in time.

Assumption (3) is that residual variability is local heterogeneity and not instrument error. This is important, because it means we have an exact measurement at any given location. In this case, for sampled locations, there is no reason to borrow strength over time, because the measurement we have is exact. To predict in space, we need to borrow strength from other locations at the same time, but since an unsampled location is unsampled at all time points, the only relevant information (provided the other assumptions are satisfied) is from observations at the same time.

Assumption (4) of normality is needed to do the conditional normal derivations of the kriging predictions above, but with a bit more work, one could probably show that the space-time kriging predictions without normality are the same as I give above, in which case this assumption is not necessary. In any event, in many situations, results that hold exactly under normality hold reasonably well as long as one is not too far from normality.

So intuitively, what do we need to just model observations at each time point separately? Basically we need that all the relevant information for prediction comes only from other locations, not from other times. Separability ensures that we do not have dynamics that induce space-time dependence from a point in space to a different point in space in the past or future (except dependence coming from the spatial and temporal covariances). Complete sampling ensures that we are never in the situation of borrowing strength over time for a location that is partially sampled in time. Local heterogeneity ensures that we do not need to smooth over time to smooth over measurement error.

In many natural phenomena, such as atmospheric processes, for long averaging times and precise instruments, with measurements from regular sampling networks, these assumptions are likely to be roughly satisfied. I expect that slight departures from the assumptions, such as small amounts of instrument error and occasional missing samples, would not justify fitting a full space-time model.

Basically, the intuition is to think about where the information is coming from for making a space-time prediction. If the information is coming from other spatial locations, then accounting for temporal correlation is not important.

Also, if one is doing smoothing via spline models or other approaches, I expect that these results will be relevant, as most modeling basically amounts to weighted averaging of the data and space-time prediction amounts to averaging with respect to nearby observations in space and time. Under the conditions above, observations at other times provide no additional information relative to observations at the same point in time.

The implication is that one can fit separate models for each point in time. However, one may want to assume common hyperparameters for the different time points (particularly with sparse data), in which case one may want to use the product of the likelihoods over the different time points, treating the observations as independent. I believe this will give reasonable hyperparameter estimates, but have not looked into this formally.

References

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