**Remote sensing simulation model**

**6Feb21**

The goal of the simulation model is to include a collection of "difficulties" that will challenge PARTS and panel regression. This is a useful exercise in its own right. It might also be the basis for the demonstration part of a "software release" manuscript for Methods in Ecology and Evolution that presents remotePARTS. Since the utility of remotePARTS is really the spatial component, the MEE release of remotePARTS could be based around simulations with only a spatial dimension (which I did for the MethodsX supplement of the RSE manuscript). This would involve stripping out the temporal components of the spatiotemporal model developed here.

This is a community project: I've only started with the model below to try to show the type of model I envision for challenging PARTS and panel regression. At this stage, I think it would be interesting to put in as many complications we can think of as an exercise; we might want/need to reduce the complications for an actual simulation study.

For the simulation model, I have denoted random variables as Greek letters. Also, to make the simulation feasible as laid out below, we should probably use 100 x 100 maps. The computational limit is set by the simulation of spatially autocorrelated random variables. There are possible ways around this, such as using tapered distance weighting functions for spatial autocorrelation. However, I think the main challenge is the extent of spatial autocorrelation, not the size of the map per se, so 100 x 100 might be fine.

The basic model is

*xi*(*t*) = (S*k* u*k,i* *uk,i*) + w*i* *wi*(*t*) + (S*k* *k,i* *uk,i*) *t* + *i*(*t*) (1)

*i*(*t*) = 1,*i* *i*(*t*–1) + 2,*i* *i*(*t*–1) + *i*(*t*) + *i* *i*(*t*–1) (2)

where *xi*(*t*) is the value of interest (e.g., NDVI) in pixel (location) *i* in year *t* (*t* = 1, 2, ..., *T*). The model includes one spatiotemporal variable *wi*(*t*) such as temperature; there could be more than one, although at this stage let's not. The time trend (S*k* *k,i* *uk,i*) depends on *k* = 0, ..., *K* fixed spatial variables *uk,i* such as latitude and land-cover class. Here, I'm assuming that at *k* = 0, *u*0*,i* is the vector of ones. It is also possible to have nonlinear or polynomial trends in *t*, but let's not unless somebody really wants. Because spatial variables *uk,i* occur in the time trends, I've also put them in the intercept term (S*k* u*k,i* *uk,i*) with different coefficients.

The error term *i*(*t*) is an ARMA(2,1). We could increase the lags in the ARMA(p,q), but I don't think this is necessary; it might also be better to just keep the AR(1) structure (i.e., 2,*i*  = *i* = 0). The random variable *i*(*t*) that drives the variation is Gaussian with

E{*i*(*t*) *i*(*s*)} = 0 for *s* ≠ *t*. (3)

E{*i*(*t*) *j*(*s*)} = 0 for *s* ≠ *t*. (4)

cor{*i*(*t*), *j*(*t*)} = fd(*dij*; y*i,* r*ij*, a*ij*) (5)

fd(*dij*; y*i,* r*ij*, a*ij*) is some function that gives the spatial autocorrelation component of *i*(*t*) depending on distances *dij* between pixels *i* and *j*. In the RSE manuscript we used

fd(*dij*; y*i,* r*ij*, a*ij*) = y*i* + (1 - y*i*) exp(–(*dij*/*r*)*a*) (6)

where y*i* is the nugget at pixel *i*. The magnitude of y*i* can vary across space such that

y*i* = gy(*i*) (7)

I haven't put any parameters in the function gy(*i*), but gy(*i*) could be given by a gradient or some sinusoidal function. An alternative is to have y*i* as a random variable:

y*i* ~ N(*y*1, s2*y***V***y*) (8)

cor{y*i*, y*j*} = fy(*dij*; r*y*, a*y*) (9)

where equation (9) gives the correlations in the correlation matrix **V***y*. Rather than exp(–(*dij*/*r*)*a*), it might be better to use a formulation to allow very high correlations between adjacent pixels, like I found for Alaska. Finally, in equation (5) I've indexed r*ij* and a*ij* by the locations *i* and *j* so that the degree of spatial autocorrelation could potentially change through space (as Volker mentioned at our last meeting); however, it isn't straightforward how to implement this while making sure the correlation matrices remain positive definite. It might be better to just confine spatial variation in the spatial autocorrelation to y*i*.

The temporal autocorrelation coefficients 1,*i,* 2,*i*, and *i* could have fixed spatial variation. I found spatial differences in 1,*i* in the NDVI analyses in the RSE manuscript. Each of these terms could be modeled as (illustrated for 1,*i*)

1,*i* ~ N(*b*1, s2*b*1**V***b*1) (10)

cor{1,*i*, 1,*j*} = f1(*dij*; r1, a1) (11)

It would make sense for the correlations here to be pretty large, so that 1,*i* doesn't change rapidly across the map. An alternative would be to have values of 1,*i,* 2,*i*, and *i* vary across the map as a gradient or sinusoidal function.

Finally, all of the coefficients in equation (1) could have a spatial component: these parameters are u*k,i* (*k* = 0, ...., *K*)*,* w*i*, and*k,i* (*k* = 0, ...., *K*). Thus, these could be, for example,

u1,*i* ~ N(*c*1, s2*c*1**V***c*1) (12)

cor{u1,*i*, u1,*i*} = fu1(*dij*; ru1, au1) (13)

or u1,*i* could vary across the map as a gradient or sinusoidal function.