

## GLMM's on large spatial grids

by Kasper Kristensen, Hans J. Skaug and David Fournier

It has for a long time been possible to fit GLMMs (Generalized Linear Mixed Models) in ADMB-RE. A typical example is correlated count data with Poisson distribution. However, when the observations are located on a spatial grid the number of latent variables (random effects in the ADMB-RE terminology) grows quadratically in the number of grid points in each geographical direction. The large number of random effects causes a computational challenge.

It has recently been shown (see link below) that a particular class of spatial correlation functions gives rise to a particularly efficient implementation in ADMB. The correlation between two points in the latent Gaussian random field, separated by a longitudinal distance  $\Delta x$  and a latitudinal distance  $\Delta y$ , has the separable form,  $\text{cor}(\Delta x, \Delta y) = \text{cor}_1(\Delta x)\text{cor}_2(\Delta y)$  where  $\text{cor}_1$  and  $\text{cor}_2$  are longitudinal and latitudinal correlation functions, respectively. Now, if an autoregressive model is used to model the dependency in each direction, i.e.  $\text{cor}_1(\Delta x) = \exp(-\alpha_1 \cdot \Delta x)$  and  $\text{cor}_2(\Delta y) = \exp(-\alpha_2 \cdot \Delta y)$ , it can be shown that the inverse correlation matrix (precision matrix) over all points in the grid is a sparse matrix. Hence the sparse matrix feature of ADMB-RE can be exploited to speed up calculations. An example with 10,000 grid points can easily be fit on a computer with 8Gb of memory.

### Example: Poisson counts with spatial dependence

Consider a situation where we want to model the spatial variation in animal density. The coloring in the upper left corner of Figure 1 shows the  $\log(\text{density})$  of animals. However, the density surface is unknown to us. We only observe the animal density at certain locations. The number of animals present in a small region (not necessarily the same as the grid points below) around the locations is counted. The locations, and the number of animals present, are also indicated in the plot (the numbers). Our task is to estimate the underlying density field from these observations.

We start by constructing a fine grid (50 times 50) such that each grid cell contains at most one observation. Note that most grid cells are empty (no observation is made in that cell). The animal density is assumed to be generated by a Gaussian spatial field with the correlation structure described earlier. Conditionally on the value of this field (the coloring in the figure) the observation follows a Poisson distribution with expected value proportional to the density field. Figure 1 shows both estimated correlation functions and the estimated (predicted) log-density field.

**More details** and ADMB files can be found at

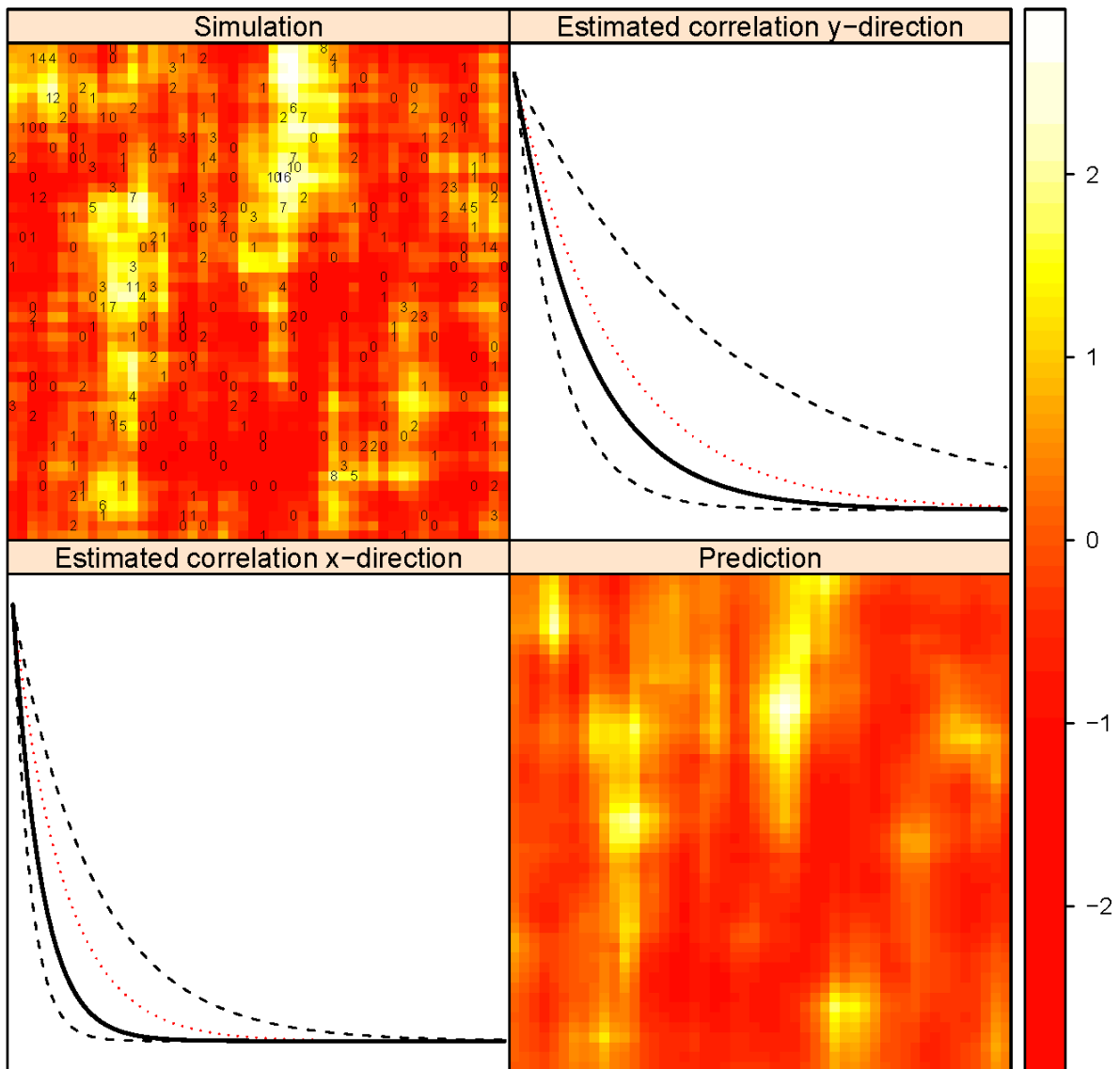


Figure 1: Results from running the model on  $50 \times 50$  grid with simulated data. Image of simulated  $\log(\text{density})$  in each lattice point with observed Poisson counts corresponding to  $\approx 90\%$  missing data (top-left). Estimated exponential correlation functions with 95% -confidence intervals (black lines) and true correlation functions (red dotted line) (top-right and bottom-left). Prediction of un-observed  $\log(\text{density})$  (bottom-right).