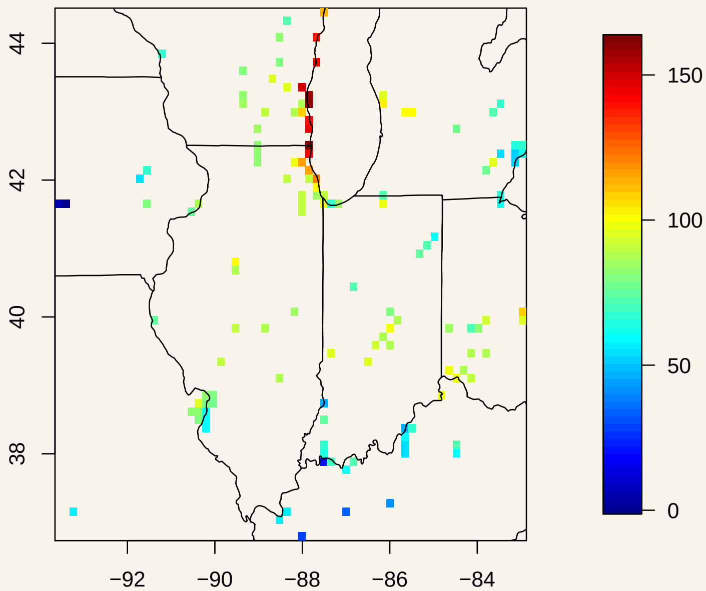
Kriging, the ensemble filter and Groundhog Day

Douglas Nychka
Geophysical Statistics Project
National Center for Atmospheric Research

Supported by the National Science Foundation DMS

Surface ozone data



Goal: Estimate the surface

Kriging

Given observations:

$$Y_k = g(u_k) + \epsilon_k$$

Find $g(u)$ at a grid of locations.

$\boldsymbol{g}$ a realization from $MN(\mu, \Sigma)$ ϵ_k are $N(0, R)$

The well known solution:

Kriging or the conditional distribution $\boldsymbol{g}$ has distribution :

$$MN(\quad cov(g, Y)cov(Y, Y)^{-1}(Y - \mu) + \mu, \quad \text{matrix stuff})$$

Dealing with one observation

The mean:

$$\hat{\boldsymbol{g}}^a = \mu + cov(g, Y_k) var(Y)^{-1}(Y_k - \mu_k)$$

A random draw:

$$\boldsymbol{g}^a = \boldsymbol{g}^f + cov(g, Y_k) var(Y)^{-1}(Y_k - \mu_k + e_k)$$

with $\boldsymbol{g}^f \sim MN(\mu, \Sigma)$ and $e_k \sim N(0, R)$

A cheating draw:

Replace $cov(g, Y_k)$ and $var(Y)$ based on sample quantities from an ensemble ... and localize/taper.

$\boldsymbol{g}^f$ is an ensemble member.

Kriging and Groundhog day

Initialize

ensembles are draws from the stochastic model for $\boldsymbol{g}$

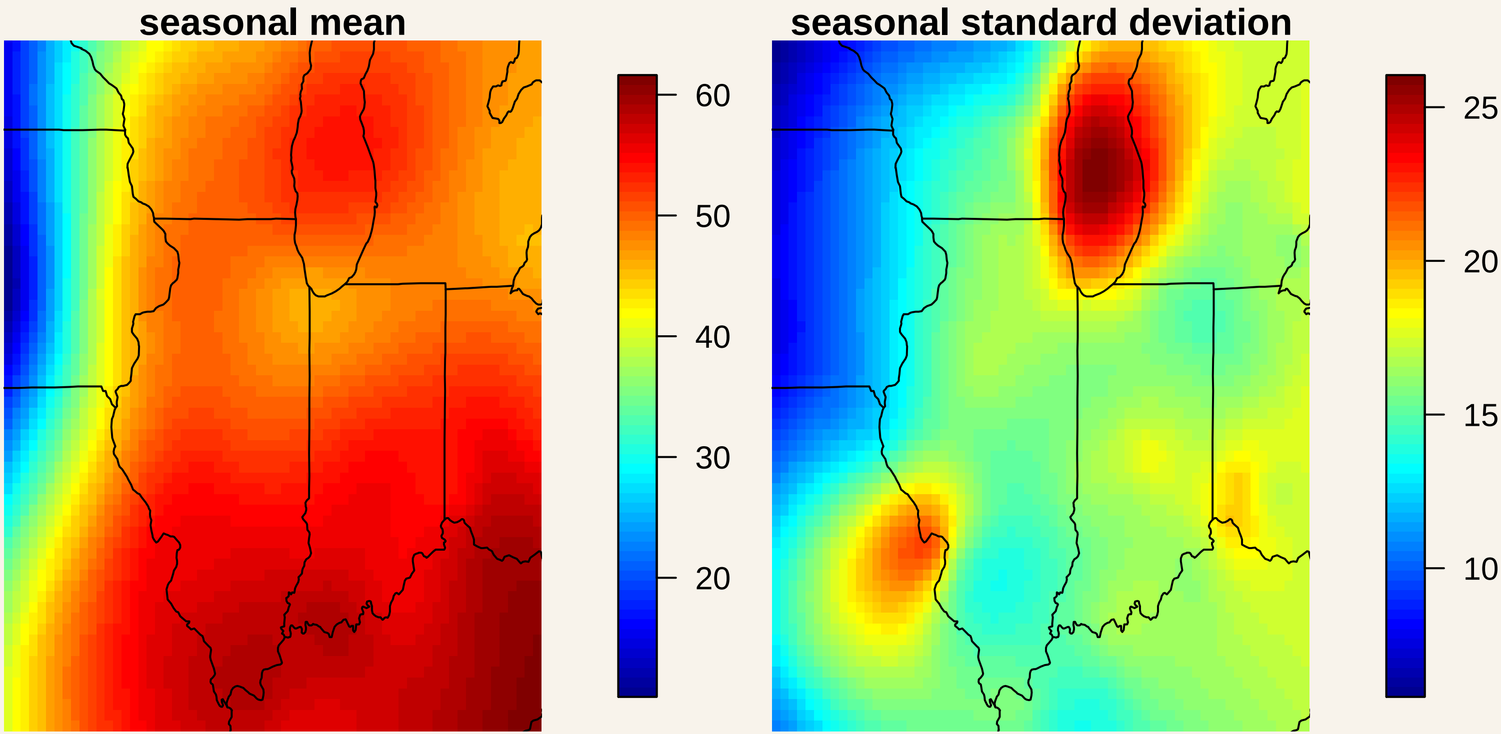
Sequentially update

ensembles using all the observations.

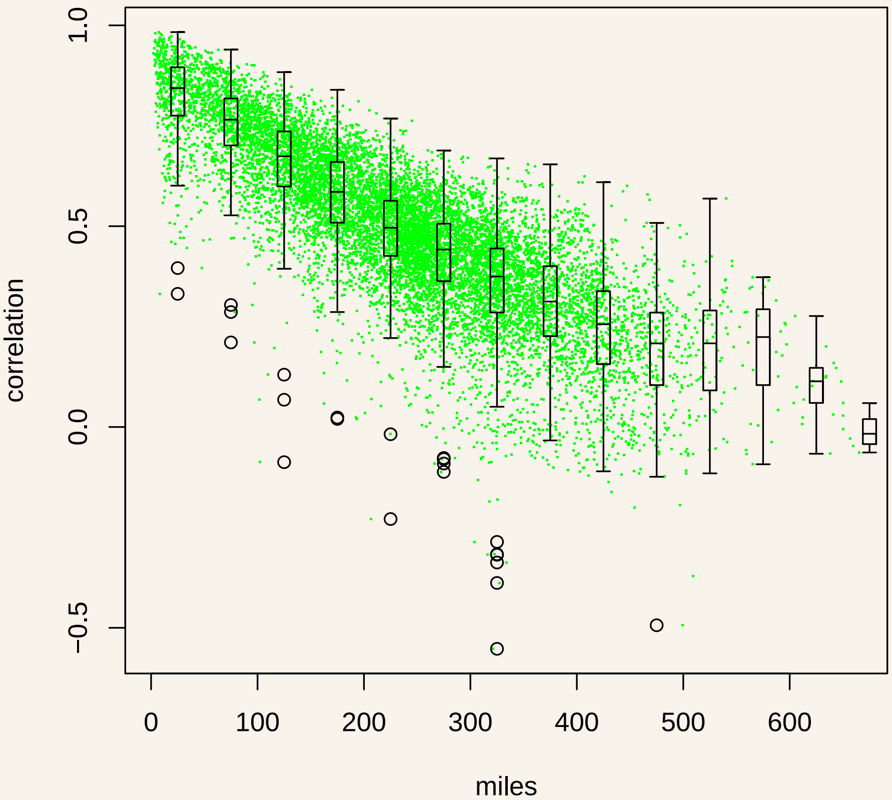
At the end of the day

ensembles will be approximate draws from a desirable distribution: The distribution of $\boldsymbol{g}$ *given* the observations.

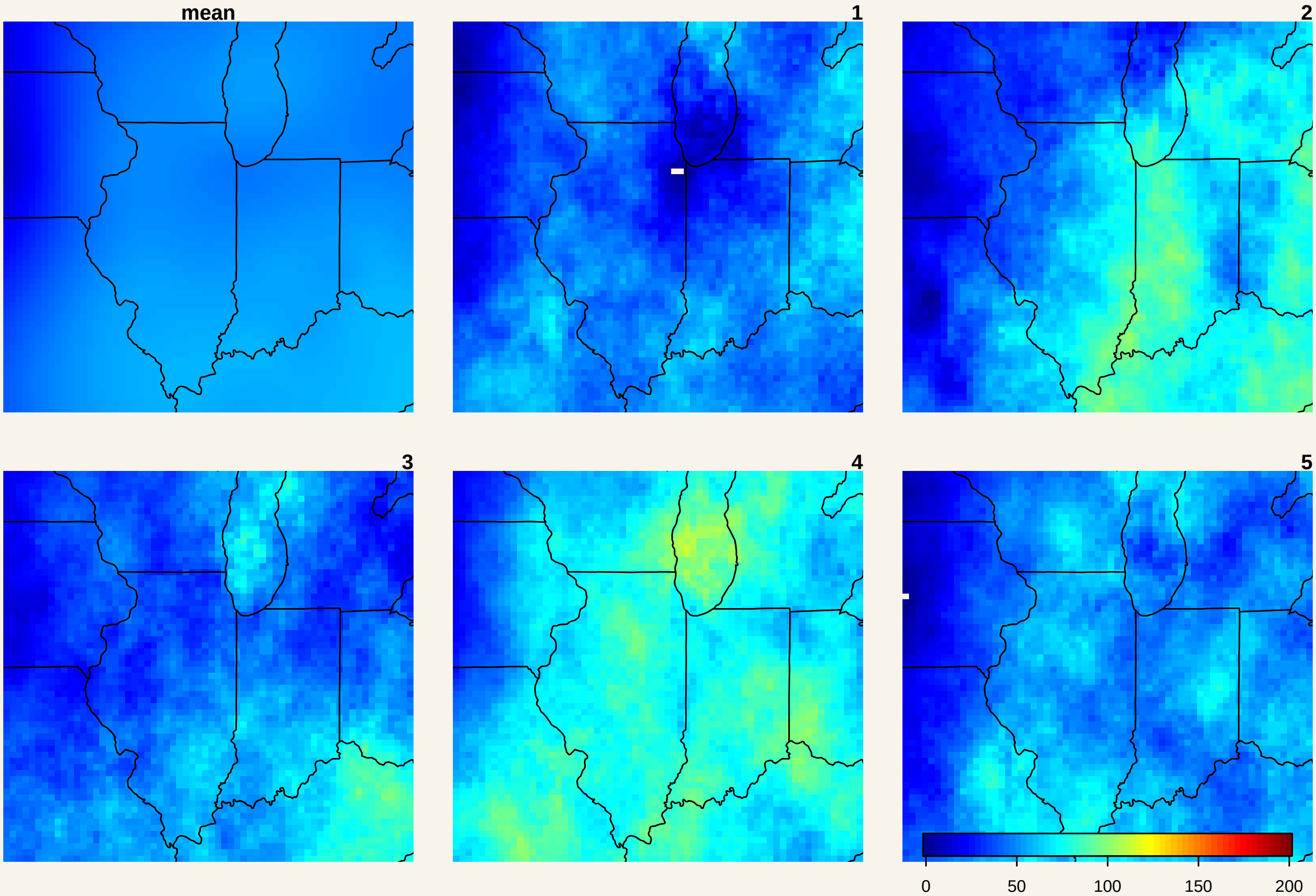
Some ozone statistics based on 79 days over the summer.



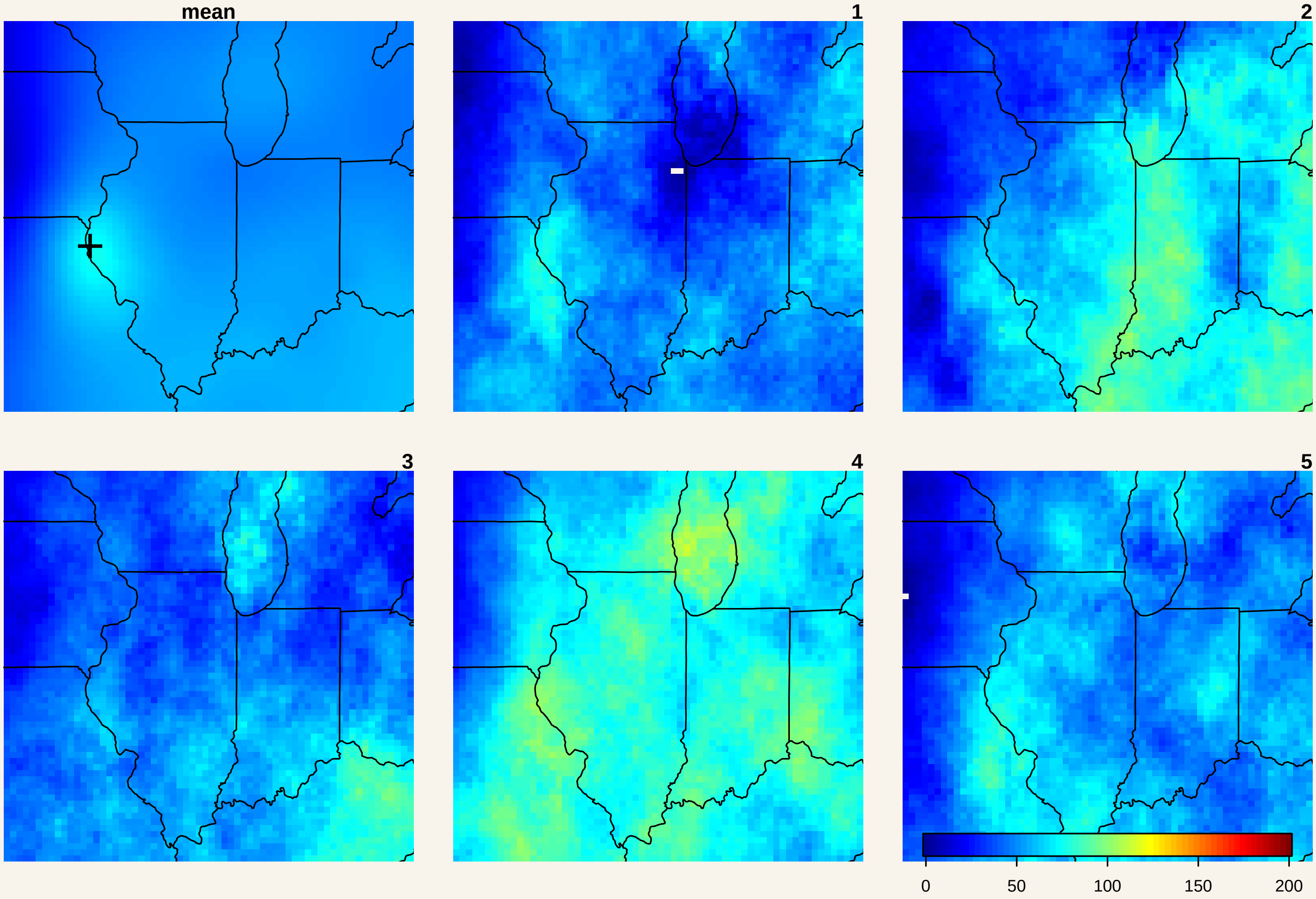
Correlations



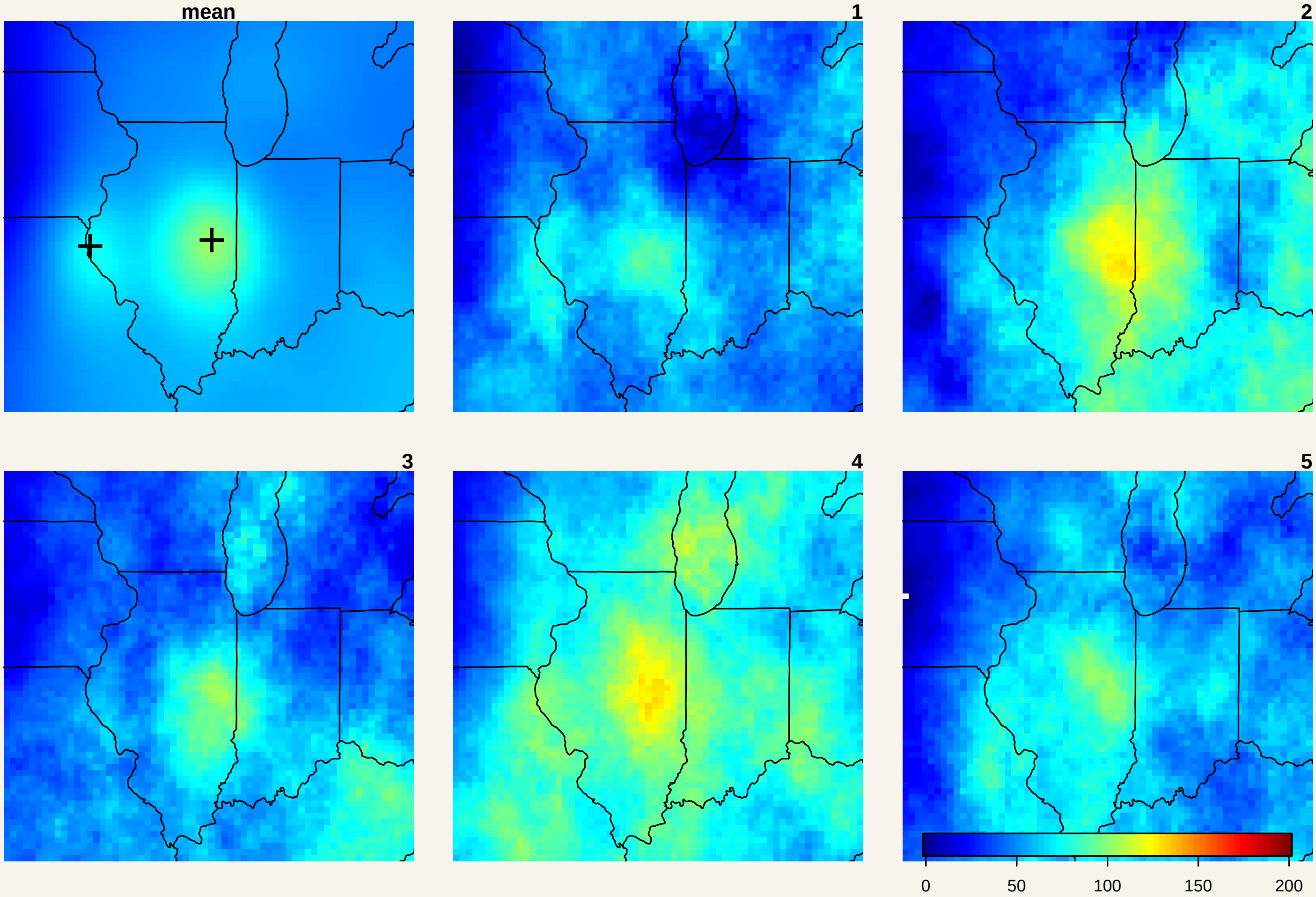
Mean and 5 members of initial fields



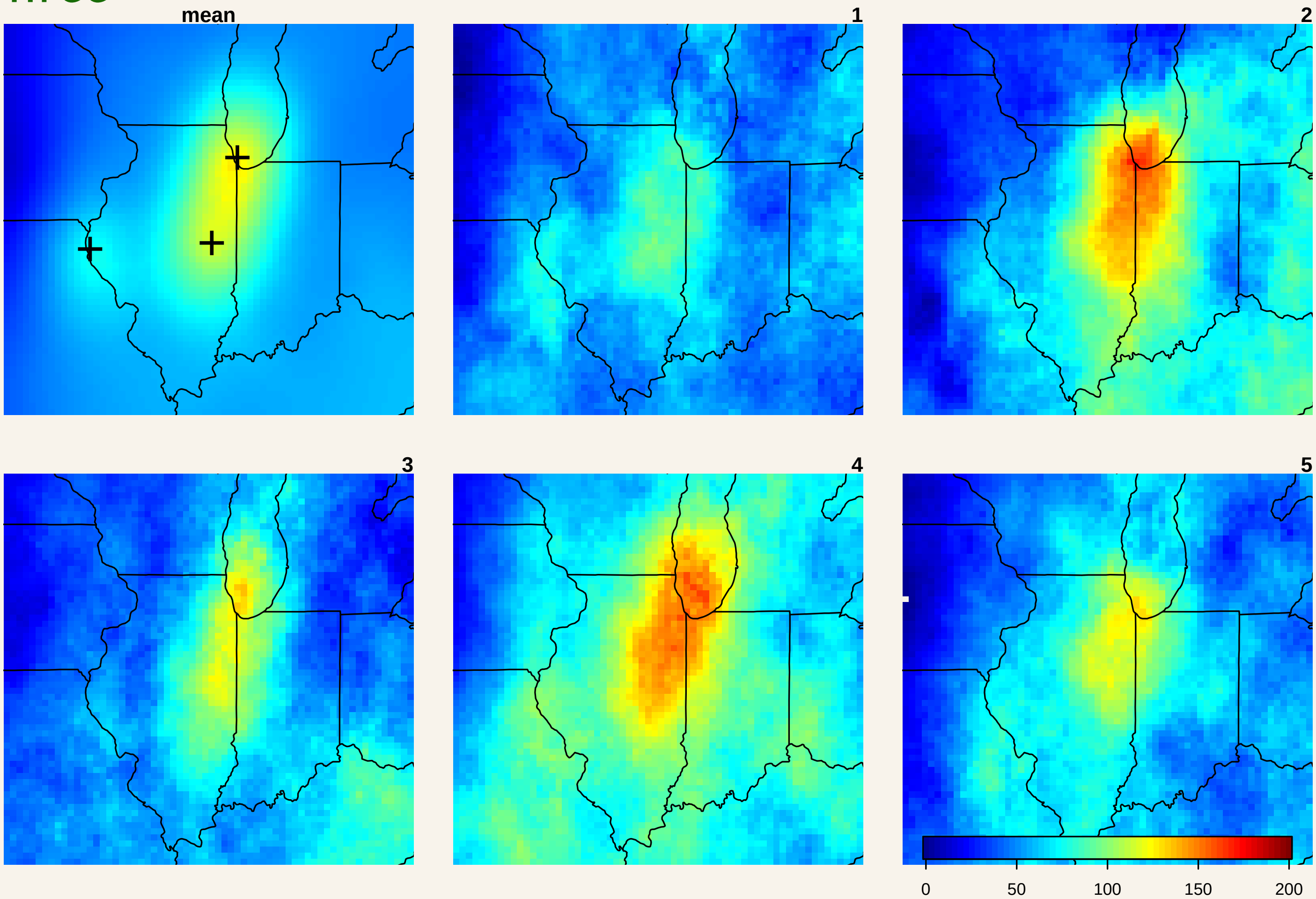
Update the fields with one observation



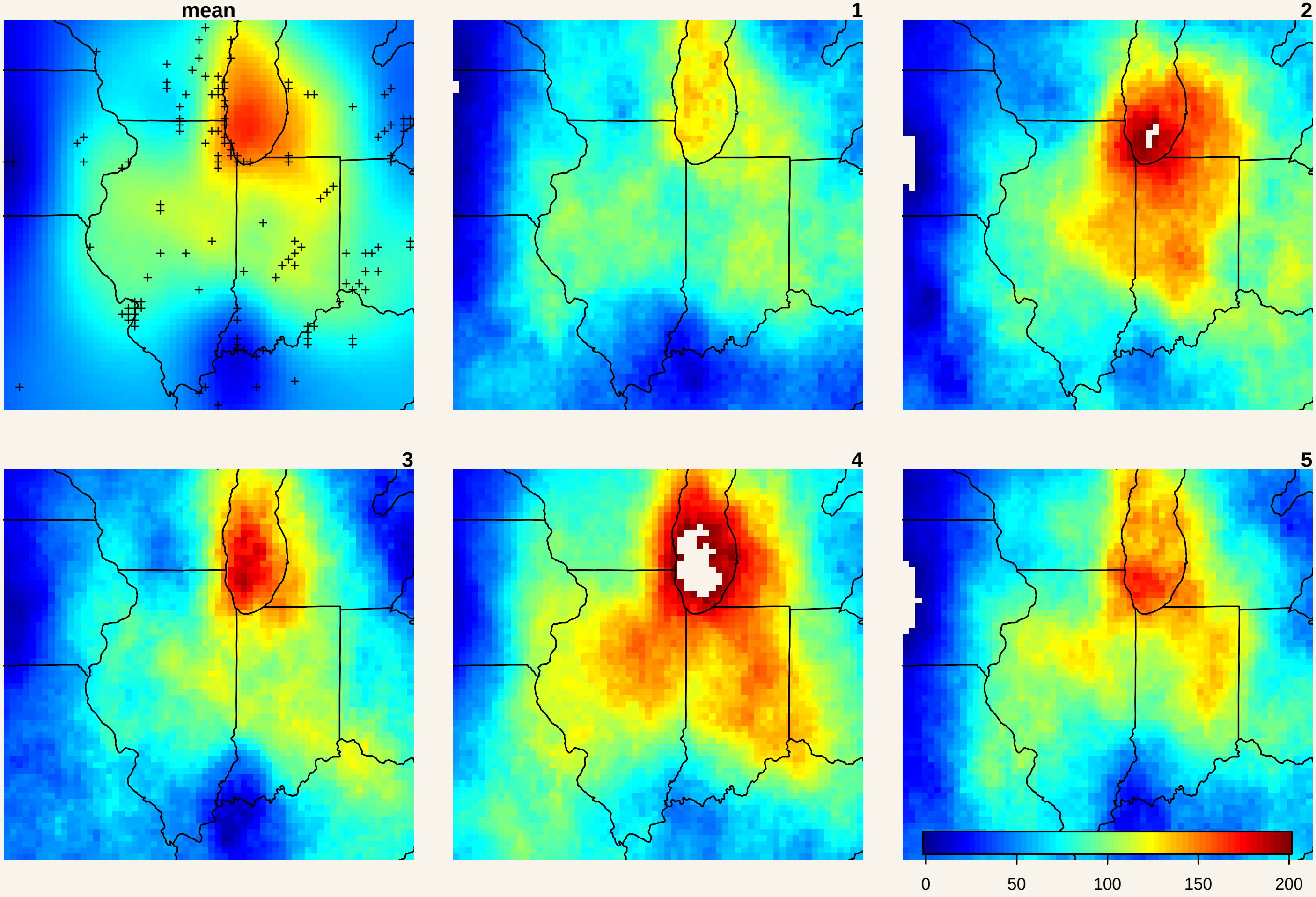
Update the fields with another observation



Three



Using all the data



Summary

DART provides a basic calculation that, by the choice of ensembles, to handle large classical spatial statistics problems.

Although the computations are approximate it has the advantage that one obtains an ensemble of possible fields (i.e. draws from an approximate posterior).