OHRFC Geostatistical Simulation-based Short-term Ensemble Hydrologic Forecasting

Thomas Adams

thomas.adams@noaa.gov

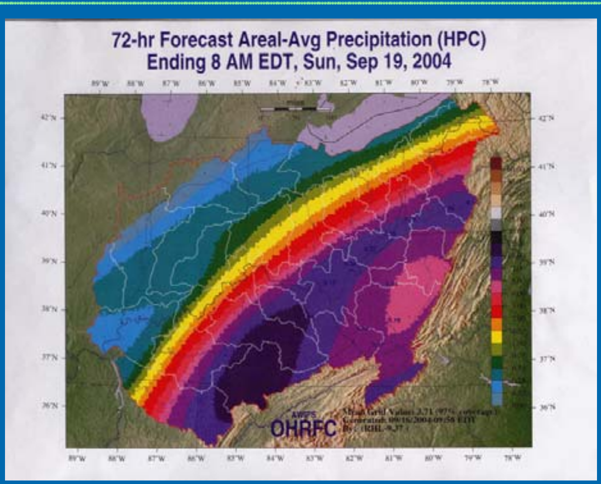
National Weather Service

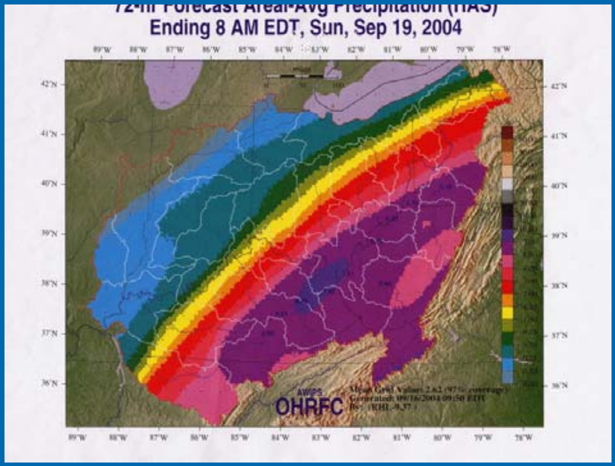
Ohio River Forecast Center, Wilmington, OH

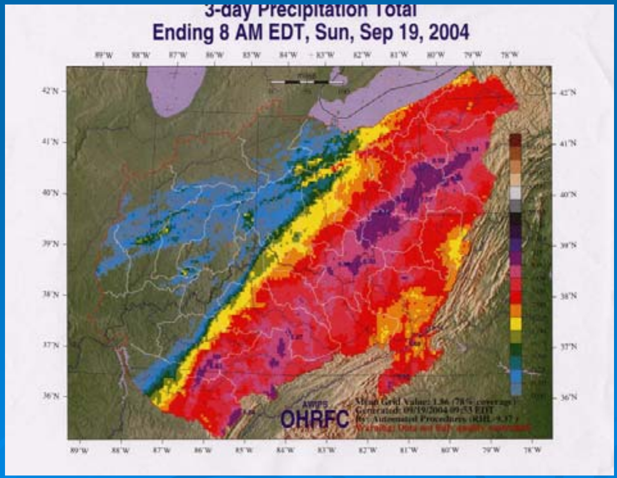
Topics

- Hydrologic Forecast Uncertainty
- Geostatistics
- Random Field Simulation
- Univ. of Washington — ProbForecast GOP method
- OHRFC MM5 mesoscale NWP modeling
- OHRFC proposed methodology
- Discussion & Questions

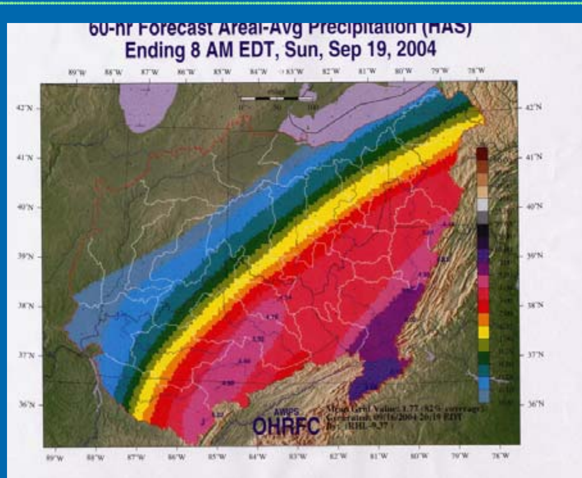
Storm-total QPF September 2004

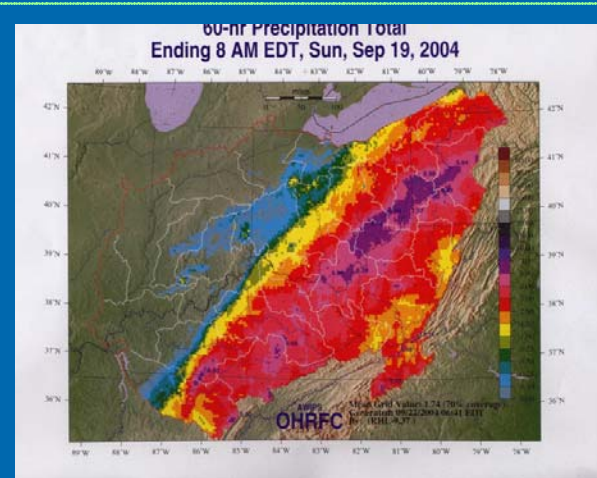


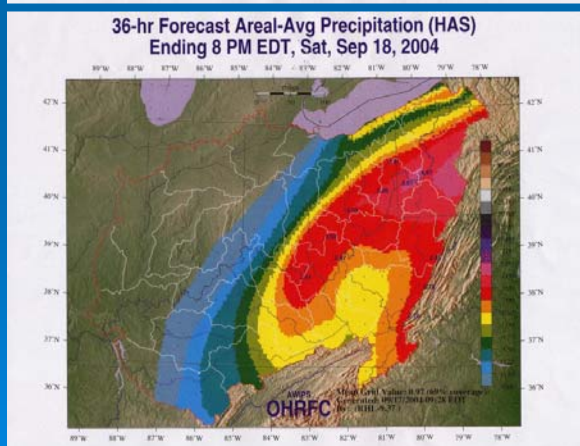


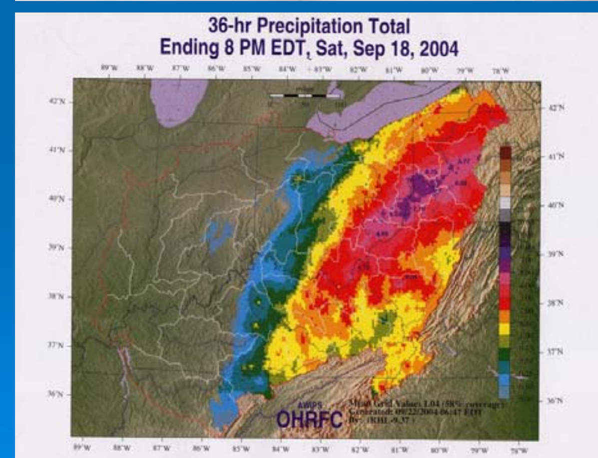


QPF — shorter lead-times









Hydrologic Forecast Uncertainty

QuickTime™ and a
Photo - JPEG decompressor
are needed to see this picture.

- Modeling error
- Error in input measurement
  - QPE
  - Temperature
  - Snowpack (SWE, depth, etc.)
  - Evapotranspiration
  - River stage/Gauge ratings
- Model states uncertainty

Geostatistics — Key concepts

- *Spatial dependence*: the value of a variable at a point in space is related to its value at nearby points; knowing the value of these points allows us to predict (with some degree of certainty) the value at the chosen point
- *Spatial structure*: the nature of the spatial relation: how far, and in what directions, is the spatial dependence? How does the dependence *var(y)* with distance and direction between points?
- *Support of a sample*: the physical dimensions it represents (n.b. may try to predict to coarser or finer resolutions)

Spatial Correlation

- Question: are nearby points in geographic space also *nearby* in feature space?
- That is, does knowing the value of some variable at some location give us information on the value at *nearby* locations?
- The concept of correlation between variables can be applied to correlation within a variable, using distance to model the relation.
- Auto-correlation
  - We want to apply the idea of correlation to one variable (auto-correlation); the prefix auto- means 'self ', here referring to the single variable. Here, the correlation is controlled by some other dimension:
    - time — if the variable is collected as a time-series
    - space — if the variable is collected at points in space
  - So we will get a measure of how much the variable is correlated to itself, considering the other factor (time or space) .
- Spatial auto-correlation
  - Two methods; the variable to be autocorrelated can be:
    - classified according to a stratification of space
    - considered as spatially continuous with the distance between point-pairs as the variable
  - The first is simpler to conceive, as it uses non-spatial correlation analysis on a spatially-classified variable. The second requires a new mathematical formulation and stronger assumptions.

Auto-covariance & Semivariances

- Auto-covariance
  - The spatial auto-covariance is computed within the same variable, using pairs of observations.
  - Each pair of observations (x_i, x_j) has a covariance, showing how they jointly differ from the variable's mean
  - There are (n•(n − 1))/2 point pairs for which this can be calculated
  - This is a large number! For example, with 200 points this is 19,900 point pairs.
- Modelling the auto-covariance
  - By themselves the individual auto-covariances are not usefull; they just quantify the covariance of each point pair
  - We need to summarize the individual covariances as a covariance function of spatial separation
  - Theory: the covariance depends on the separation between points
  - We can then predict the covariance between any two locations in space.
- Semivariances
  - It is easier to model semivariances than with covariances
  - Each pair of observation points has a semivariance
  - Each point pair is separated by a known distance, so…
  - We can plot the semivariances against distance as a variogram 'cloud', with (n•(n − 1))/2 points in the graph
  - Can also summarize in a variogram
  - The 'semi' refers to the factor 1/2, because there are two ways to compute for the same point pair

Semivariogram

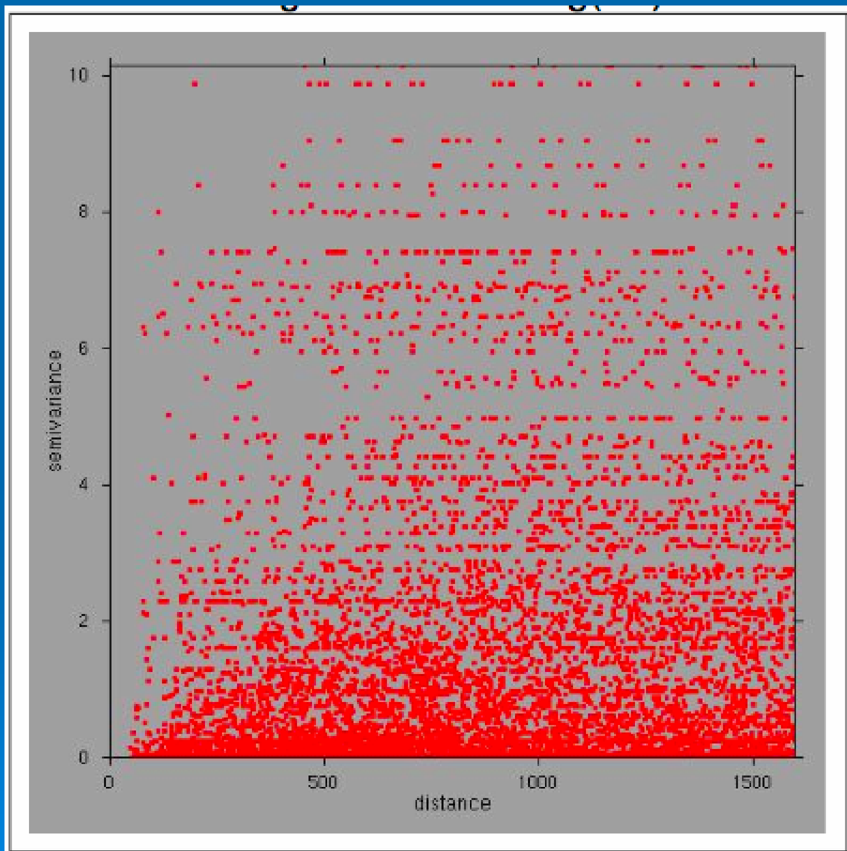
Semivariance:

$$\gamma(\vec{x}_i, \vec{x}_j) = \frac{1}{2}\{z(\vec{x}_i) - z(\vec{x}_j)\}^2$$

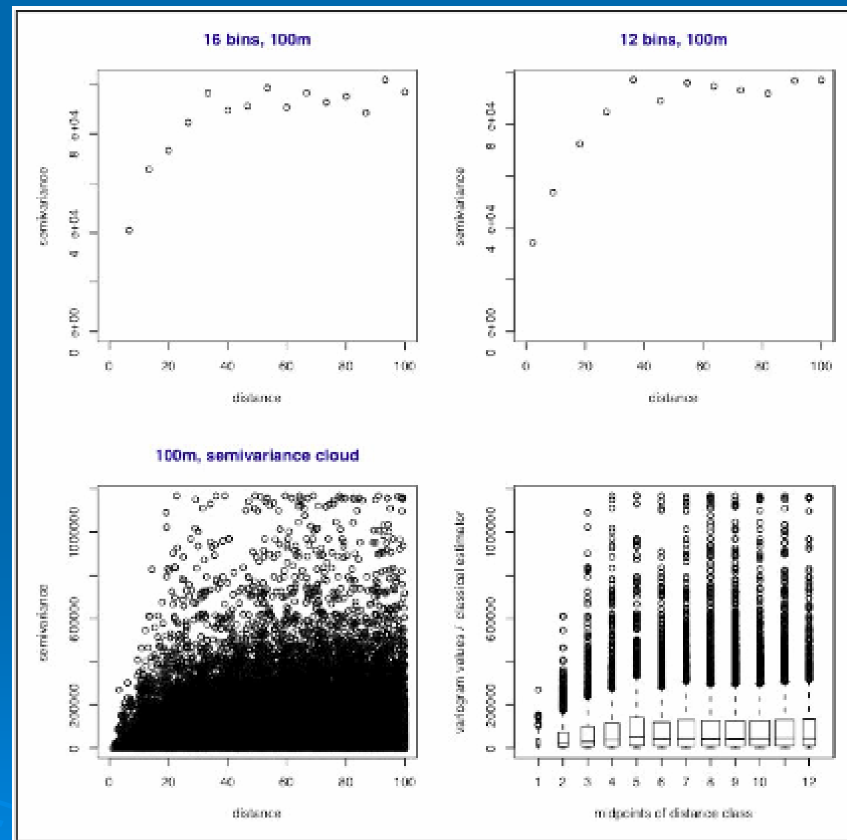
Empirical Semivariogram:

$$\bar{\gamma}(\vec{h}) = \frac{1}{2m(\vec{h})}\sum_{i=1}^{m(\vec{h})}\{z(\vec{x}_i) - z(\vec{x}_j)\}^2$$

Variogram

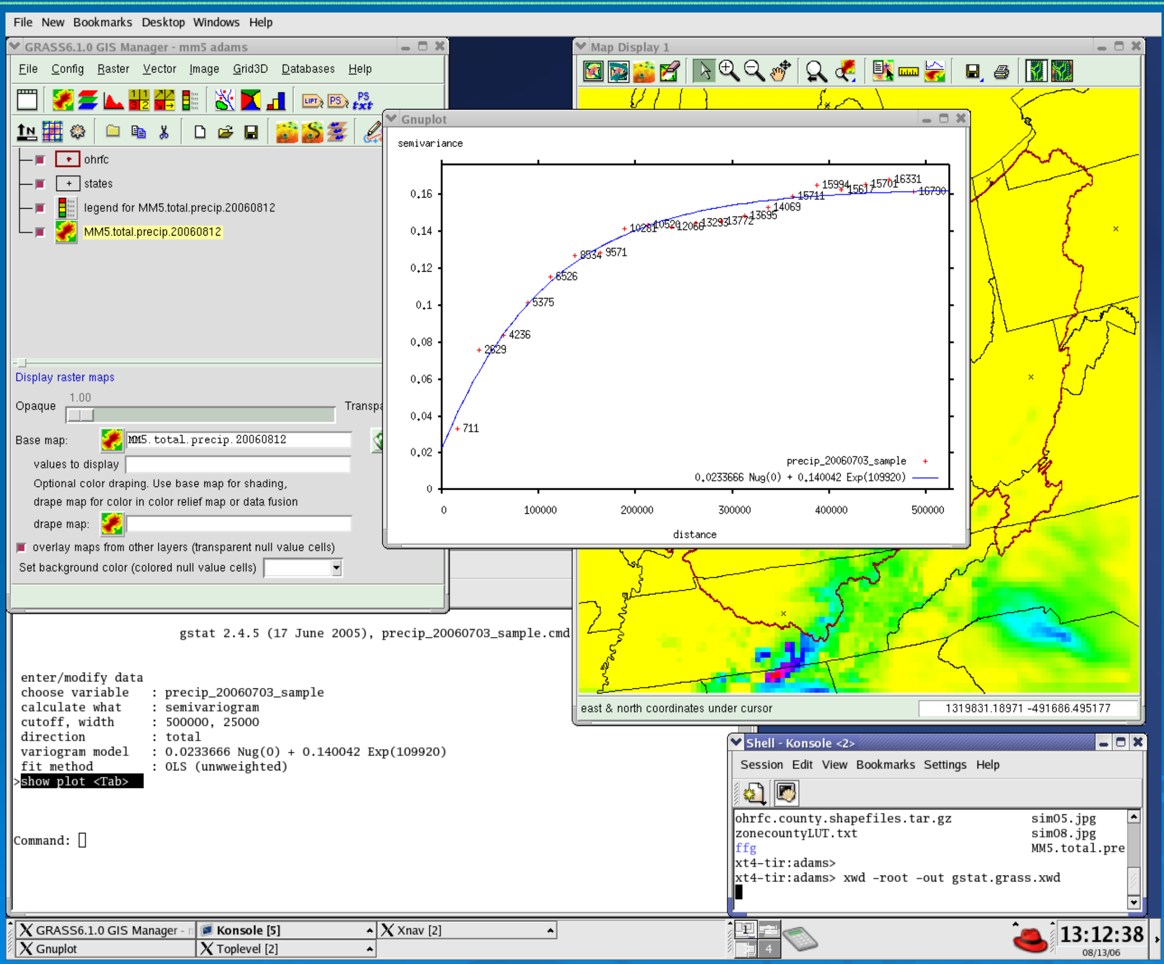


semivariance
distance
0
500
1000
1500
0
2
4
6
8
10



16 bins, 100m
12 bins, 100m
100m, semivariance cloud
semivariance
distance

GRASS 6.1 with GSTAT with experimental & fiited variogram



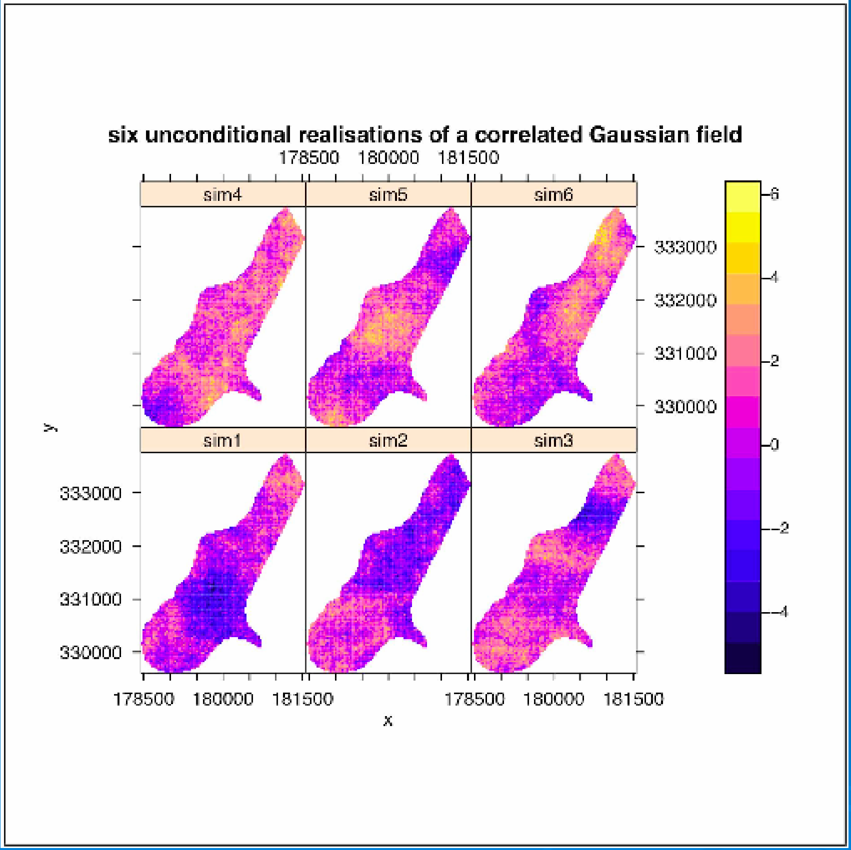
Random Field Simulation

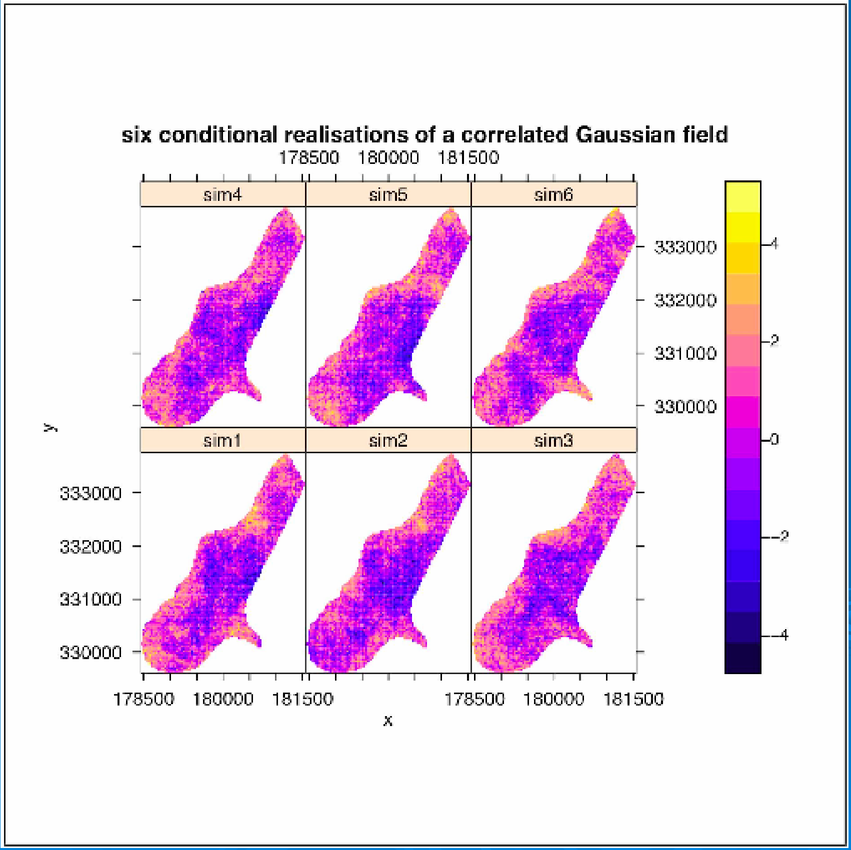
- *Simulation* is a general term for studying a system without physically implementing it.
- *Stochastic* simulation means that there is a random component to the simulation model: quantified uncertainty is included so that each simulation is different.
- Spatial uncertainty is a representation of the error over the entire field of prediction locations at the same time.
- Practical applications of spatial simulation
  - Procedure:
    - Simulate a large number of realizations of the spatial field
    - Run the model on each simulation
    - Summarize the output of the different model runs
  - The statistics of the output give a direct measure of the uncertainty of the model in the light of the sample and the model of spatial variability.

Conditional vs Unconditional Geostatistical Simulation

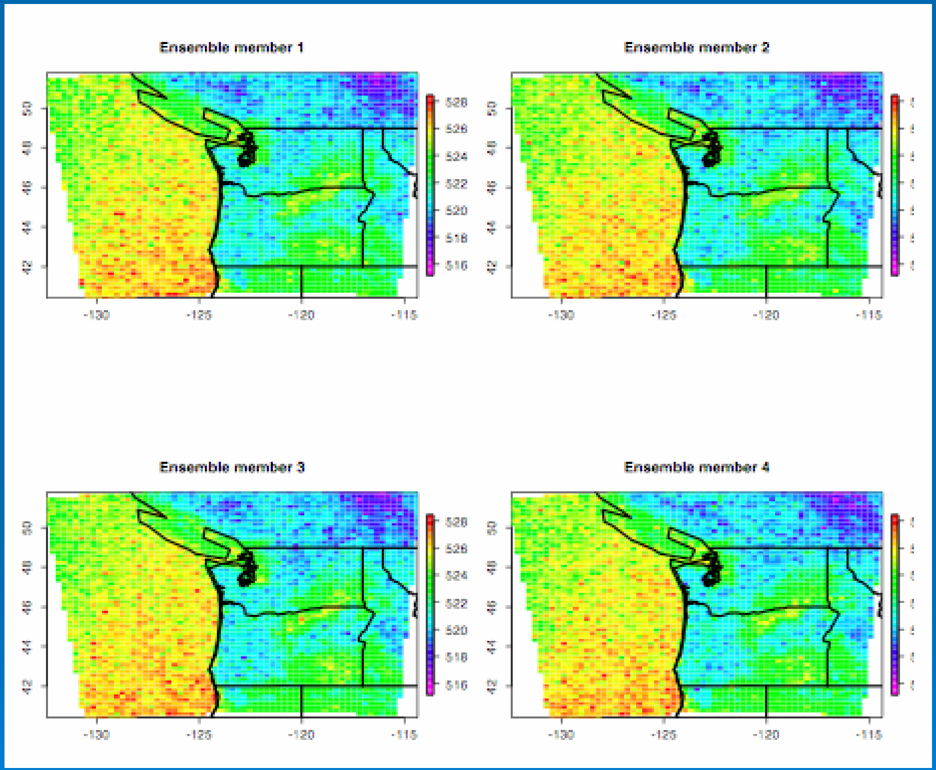
- Unconditional simulation
  - In unconditional simulation, we simulate the field with no reference to the actual sample, i.e. the data we have. (It's only one realization, no more valid than any other.) This is mainly to visualize a random field as modeled by a variogram, not for prediction.
  - What is preserved in unconditional simulation?
    - Mean over field
    - Covariance structure
    - Data points are not predicted exactly.
- Conditional simulation
  - This simulates the field, while respecting the sample. So the simulated maps look more like the best (kriging) prediction, but usually much more spatially-variable (depending on the magnitude of the nugget). These are inputs into spatially-explicit models, e.g. hydrology.
  - What is preserved in conditional simulation?
    - Mean over field
    - Covariance structure
    - Observed data (points are predicted exactly)

Conditional vs Unconditional Geostatistical Simulation — examples



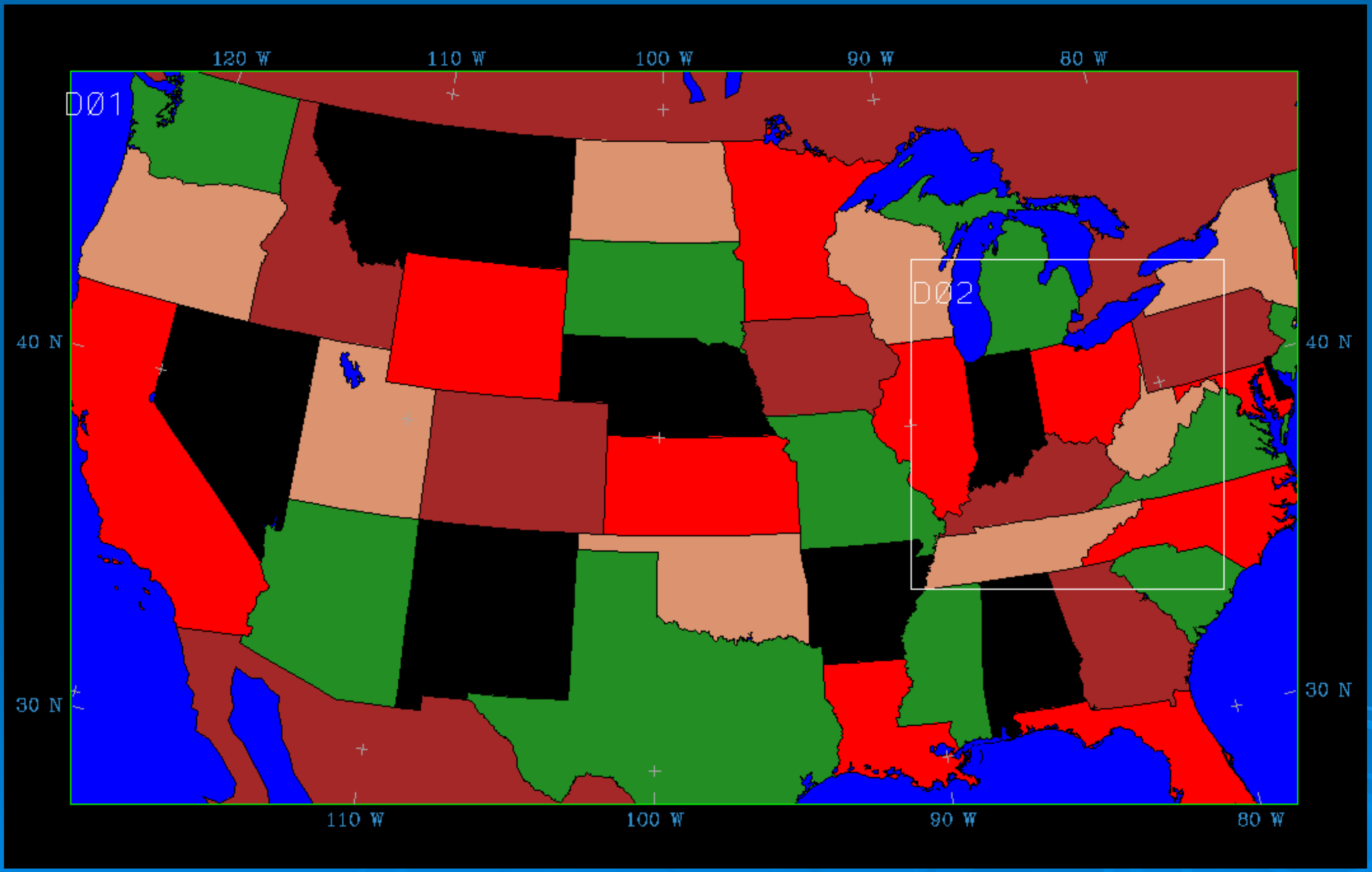


Univ. of Washington — generation of ensembles of temperature fields

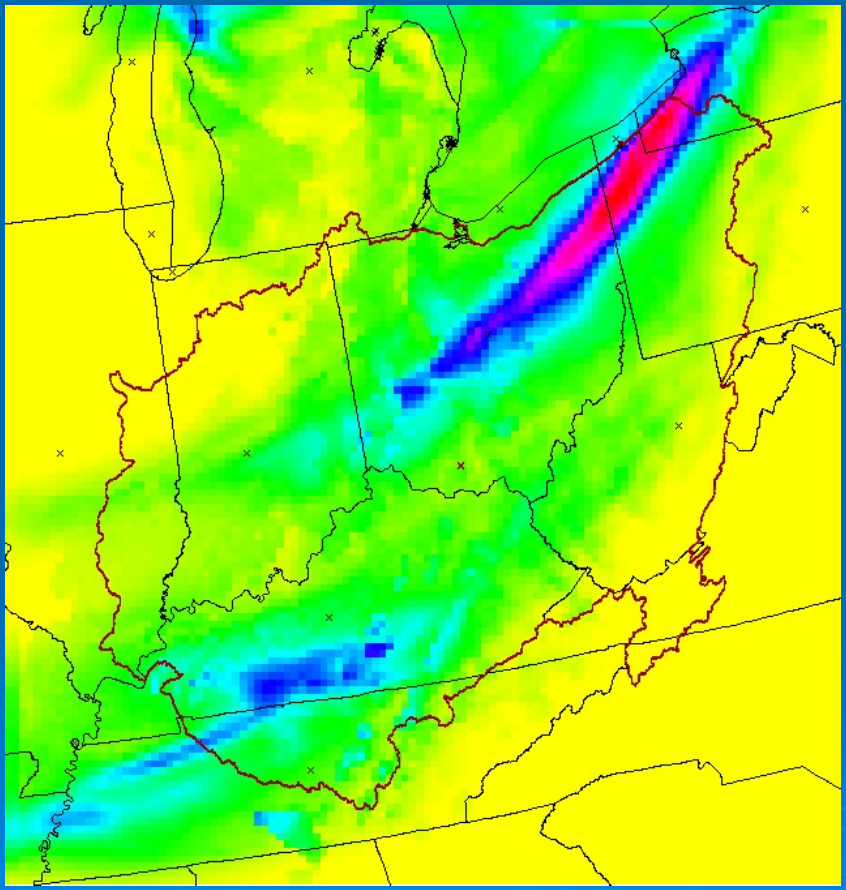


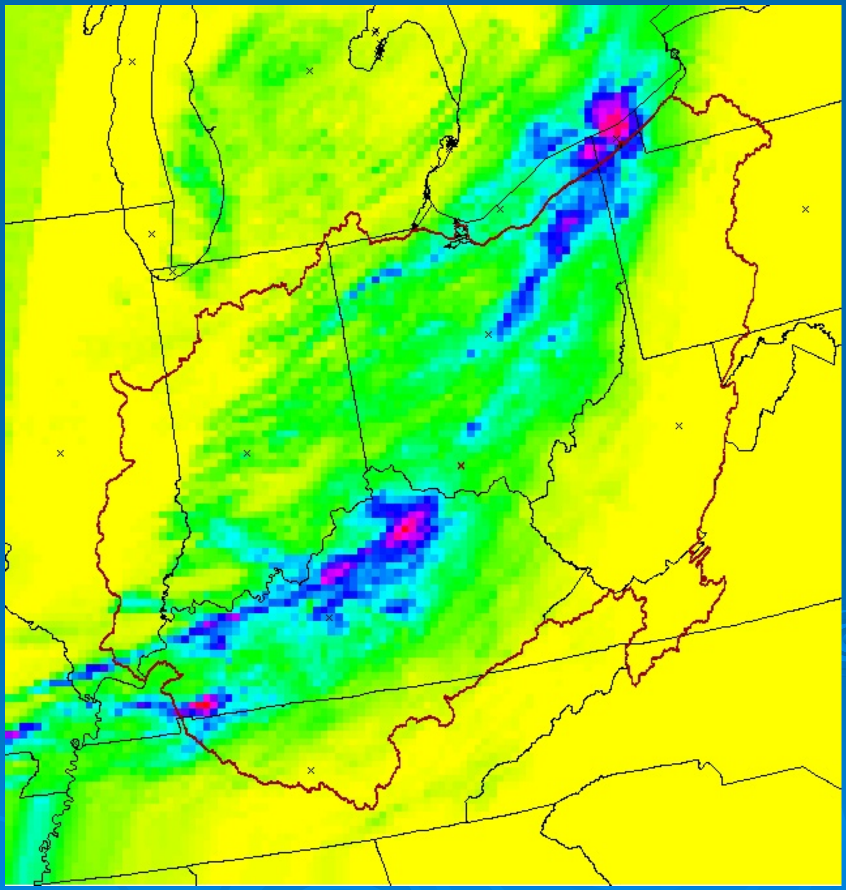
- Utilized MM5 NWP to make deterministic forecast runs
- Forecast uncertainty — random temperature fields — produced from geostatistical simulations
- *R* package: *ProbForecastGOP* method for generating ensembles of temperature fields
- GOP = Geostatistical Output Perturbation
- Gel, Y., A.E. Raftery, and T. Gneiting, 2004. "Calibrated Probabilistic Mesoscale Weather Field Forecasting: The Geostatistical Output Perturbation Method", J. Am. Stat. Assoc., 99(467).

OHRFC MM5 Mesoscale NWP Modeling

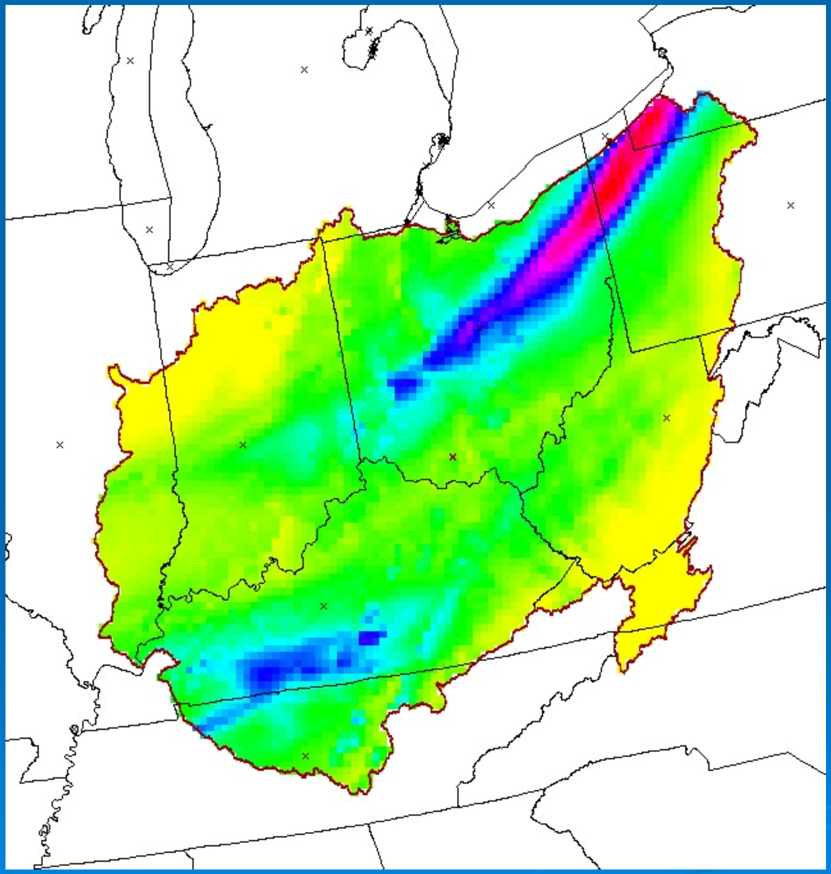


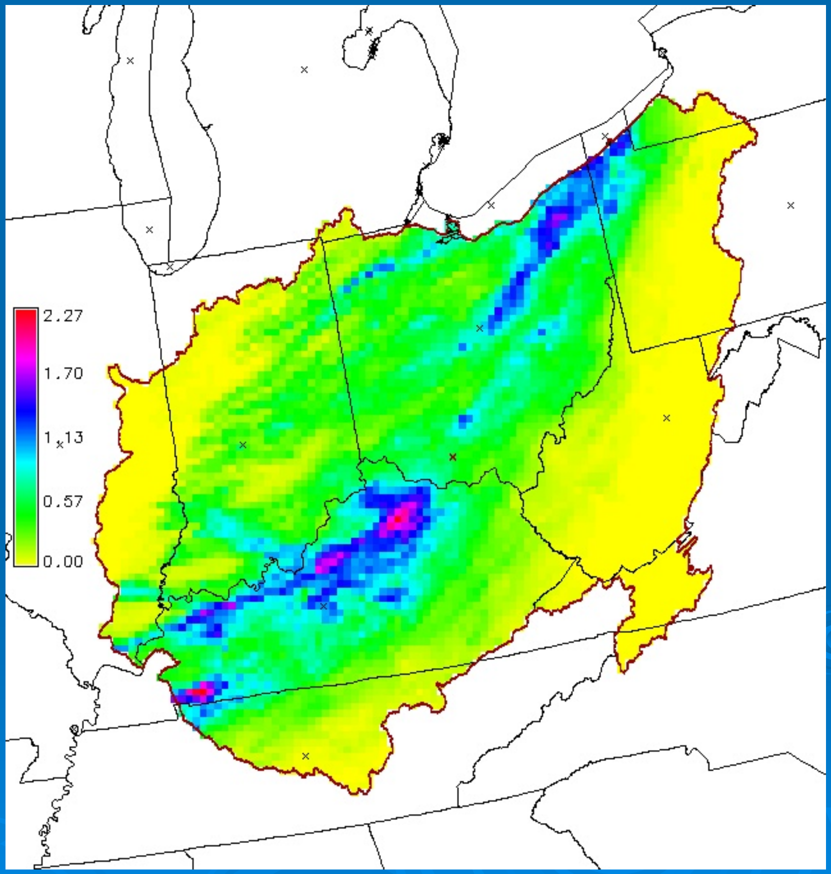
Example 24-hr rainfall ending 07/28/2006 - 12Z



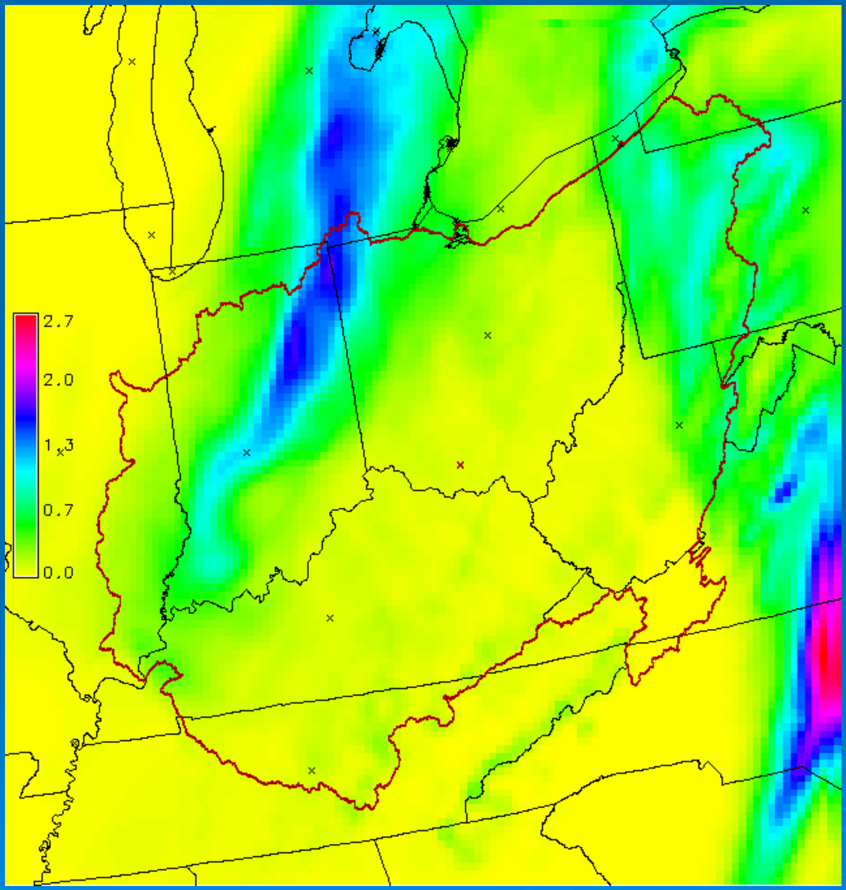


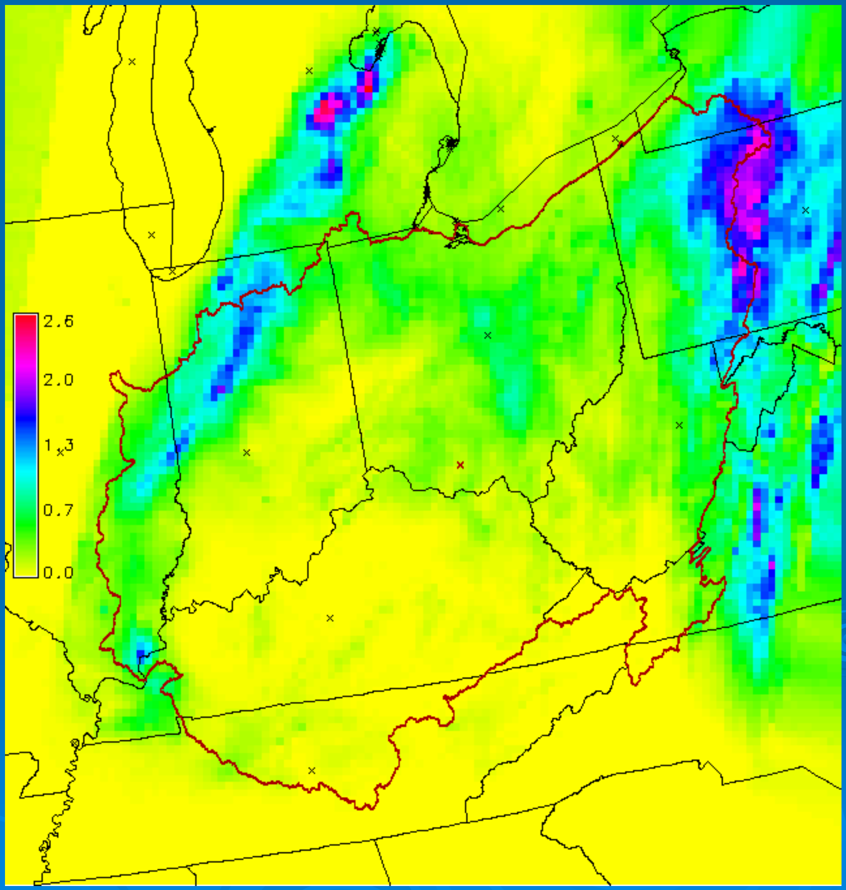
Example 24-hr rainfall ending 07/28/2006 - 12Z





Example 24-hr rainfall ending 11/17/2006 - 12Z

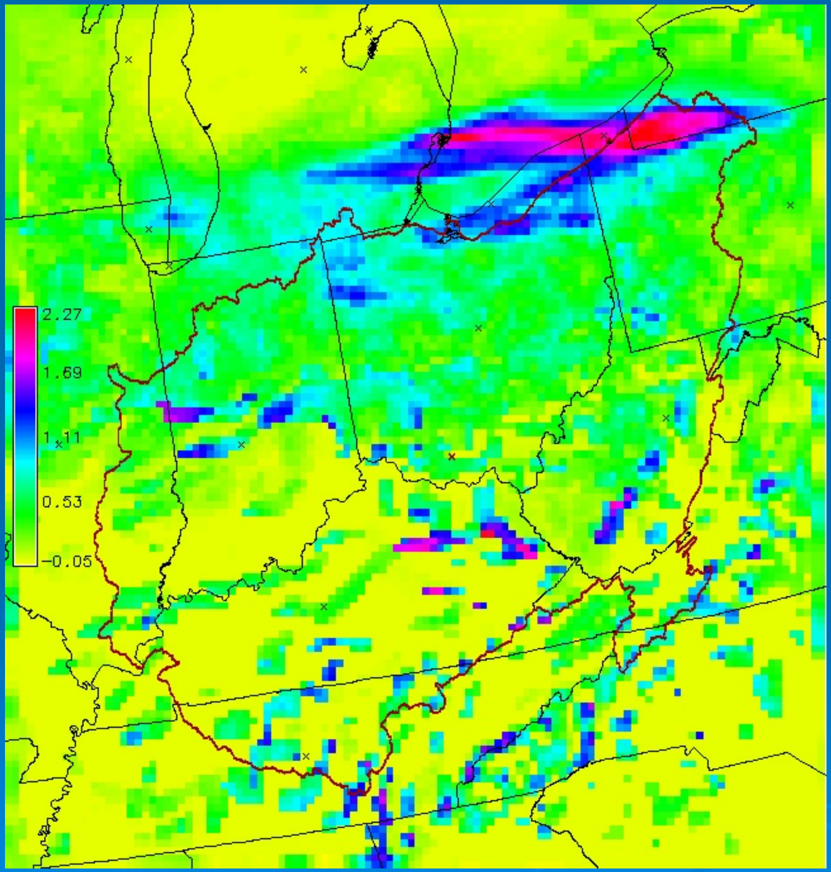


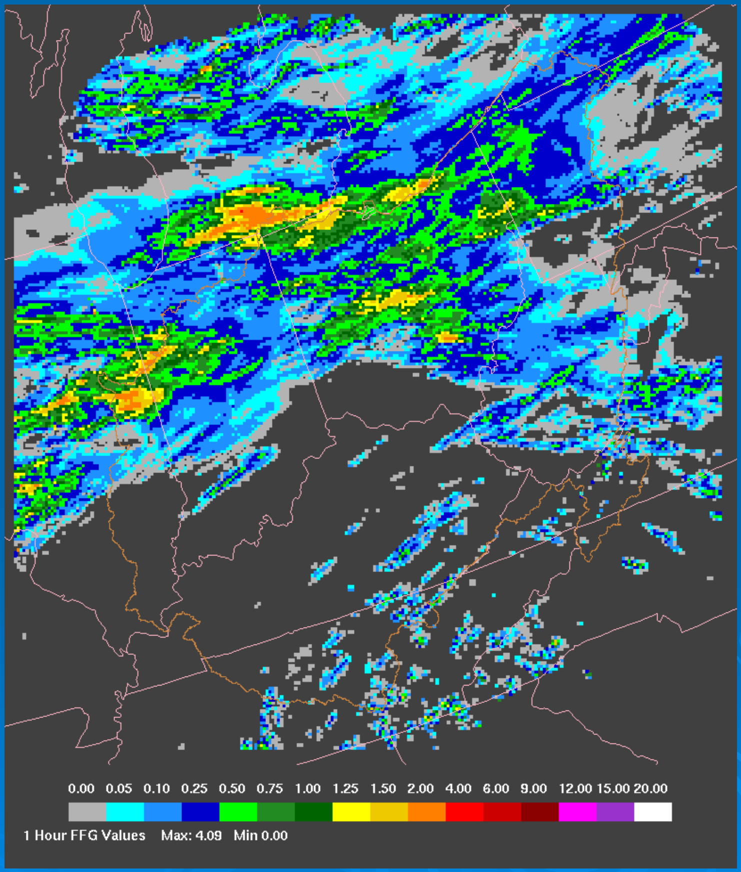


OHRFC proposed methodology

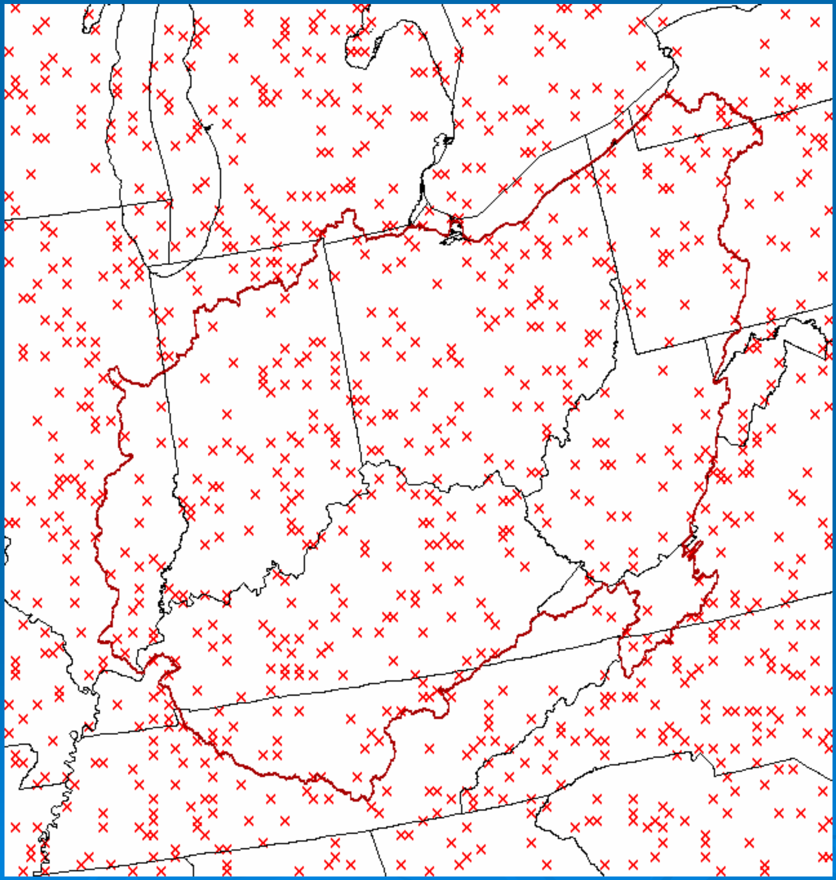
- Single 144-hr MM5 model run
- Random sample of MM5 forecasted hourly temperature & precipitation fields
- Normal score transformation (*GSLIB*) of temperature & precipitation fields
- Use *GRASS GIS* & *GSTAT* to make *conditional* gaussian geostatistical simulations
- Back-transformation (*GSLIB*) of normal score temperature & precipitation fields
- Generate basin average temperature & precipitation time series from the random field simulations
- Run *ESP*
- Run *ESPADP*
- Archive data
- Verification
  - versus deterministic forecast?
  - versus OHD methodology?

24-hr Precipitation ending 07/04/2006 12-Z

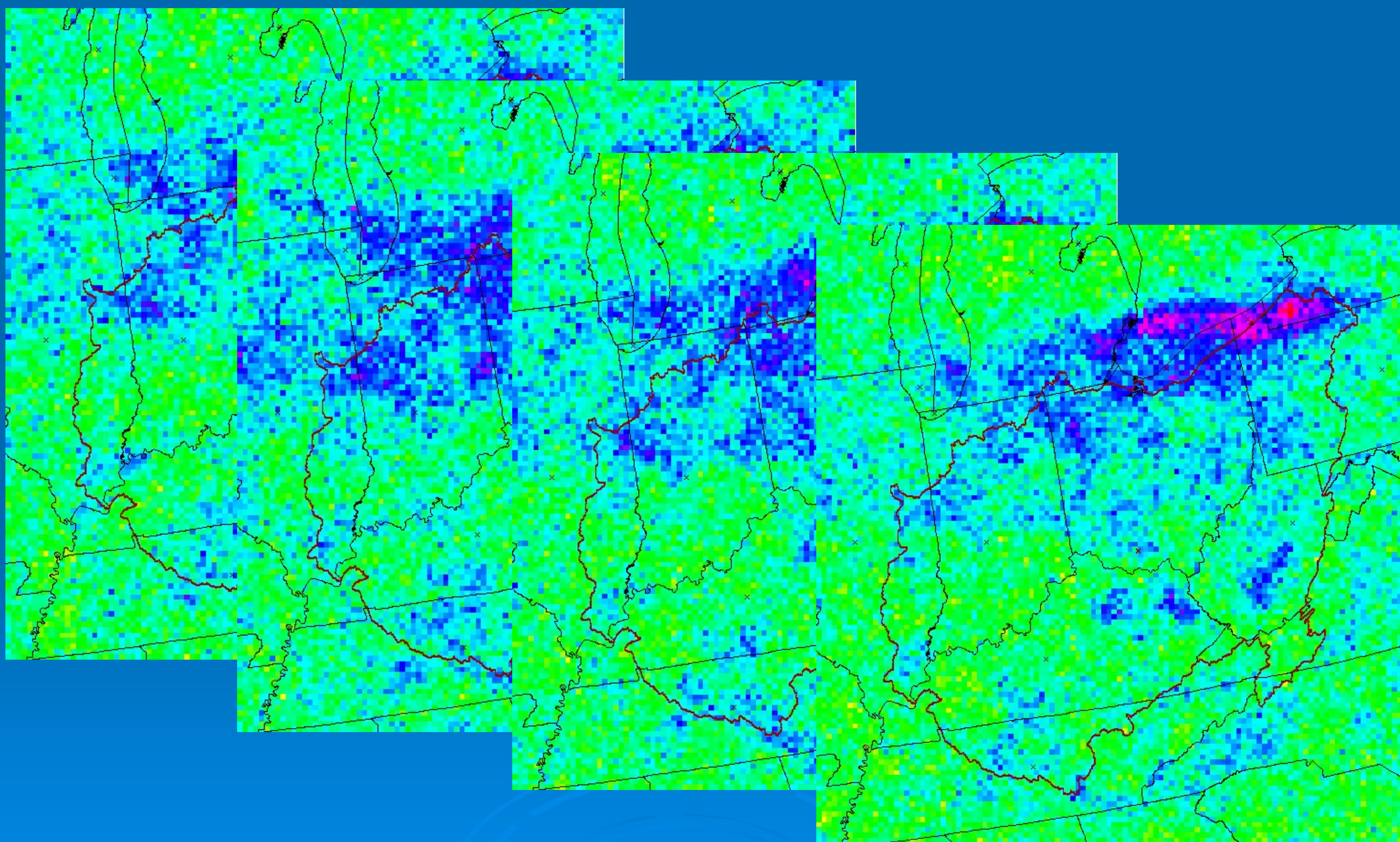




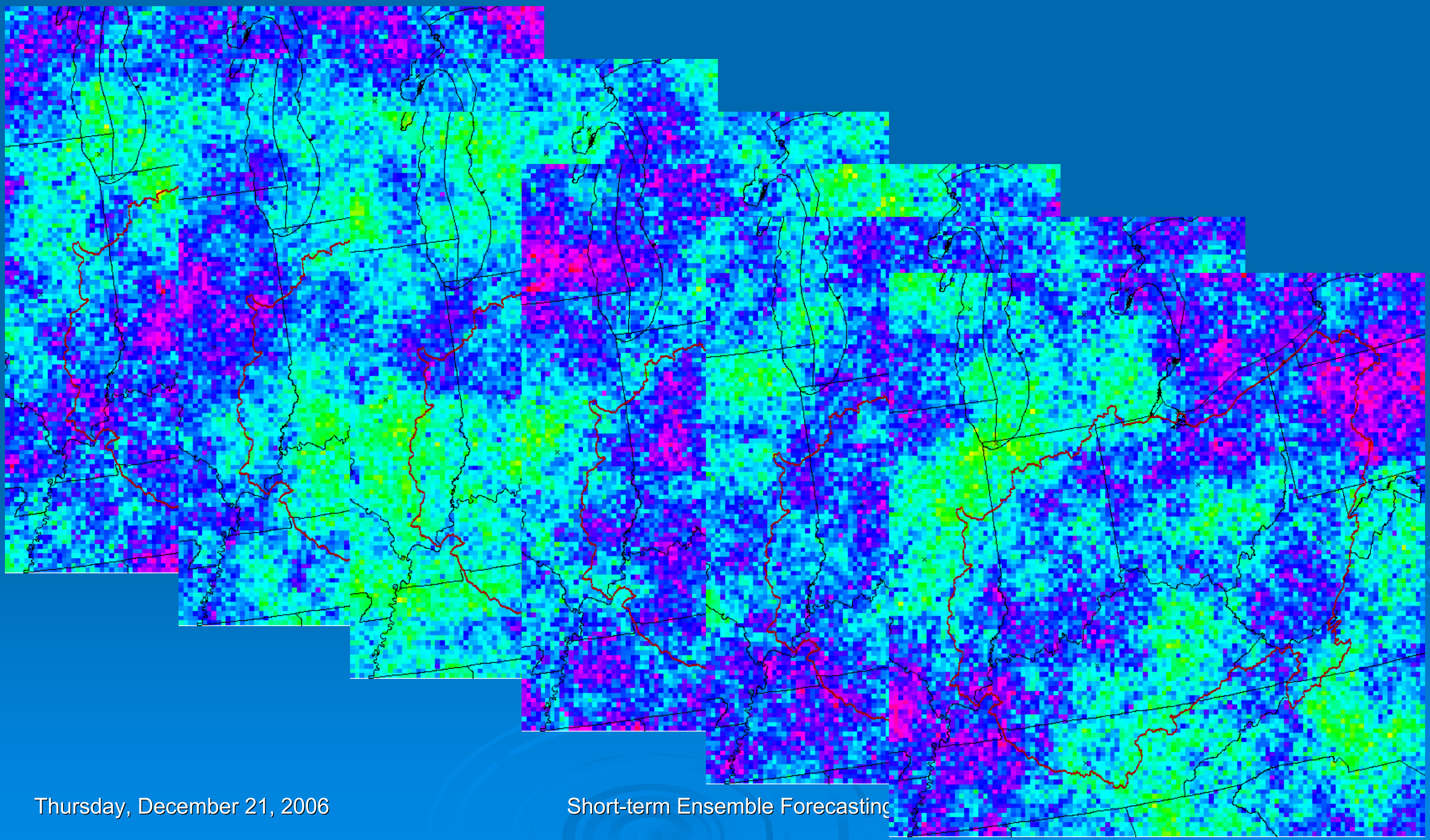
Random Sampling of MM5 Temperature & Precipitation Fields



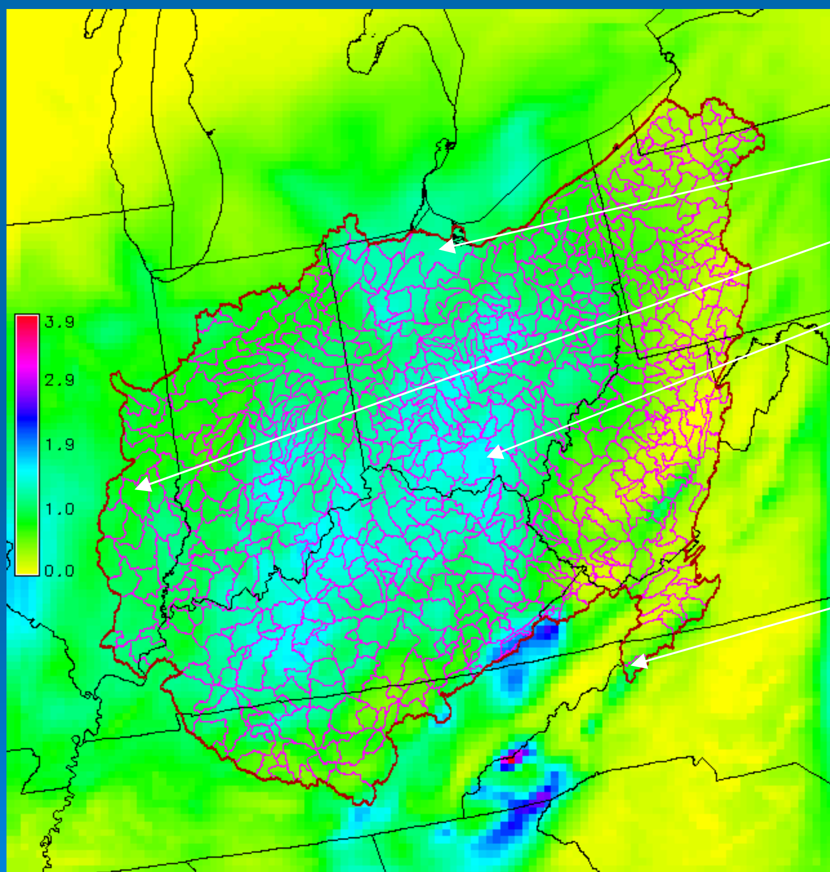
4 Realizations of Conditional Guassian Geostatistical Simulation using GSTAT & GRASS GIS



Unconditional Gaussian Simulation



MM5 grids to ESP time series



	T_1	T_2	T_3	. . .	T_n
$BASIN_1$	(t_{11}, p_{11})	(t_{12}, p_{12})	(t_{13}, p_{13})	. . .	(t_{1n}, p_{1n})
$BASIN_2$	(t_{21}, p_{21})	(t_{22}, p_{22})	(t_{23}, p_{23})	. . .	(t_{2n}, p_{2n})
$BASIN_3$	(t_{31}, p_{31})	(t_{32}, p_{32})	(t_{33}, p_{33})	. . .	(t_{3n}, p_{3n})
.					
.					
.					
$BASIN_m$	(t_{m1}, p_{m1})	(t_{m2}, p_{m2})	(t_{m3}, p_{m3})	. . .	(t_{mn}, p_{mn})

T_n : Time for period, n

$BASIN_m$: BASIN, m

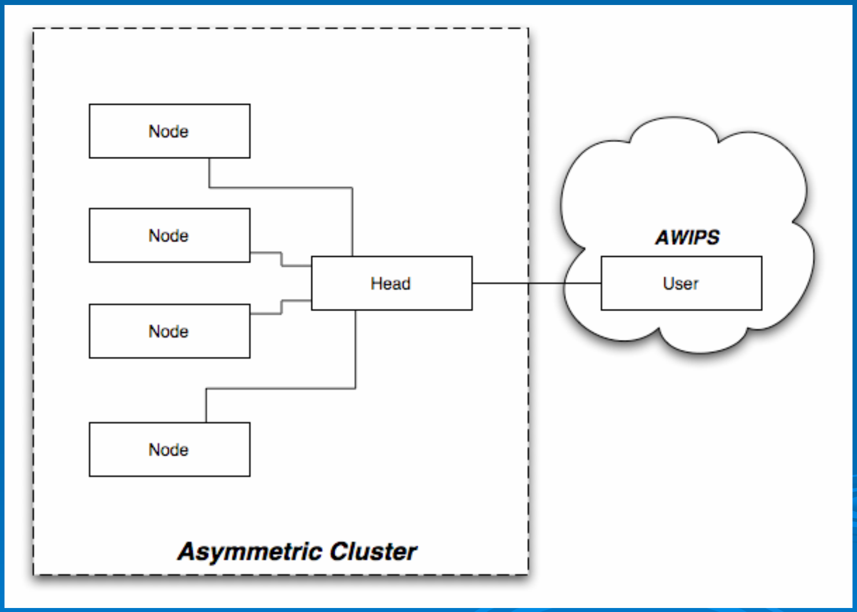
(t_{mn}, p_{mn}) : Temperature, t & precipitation, p for Basin, m for period n

MM5 Modeling Details

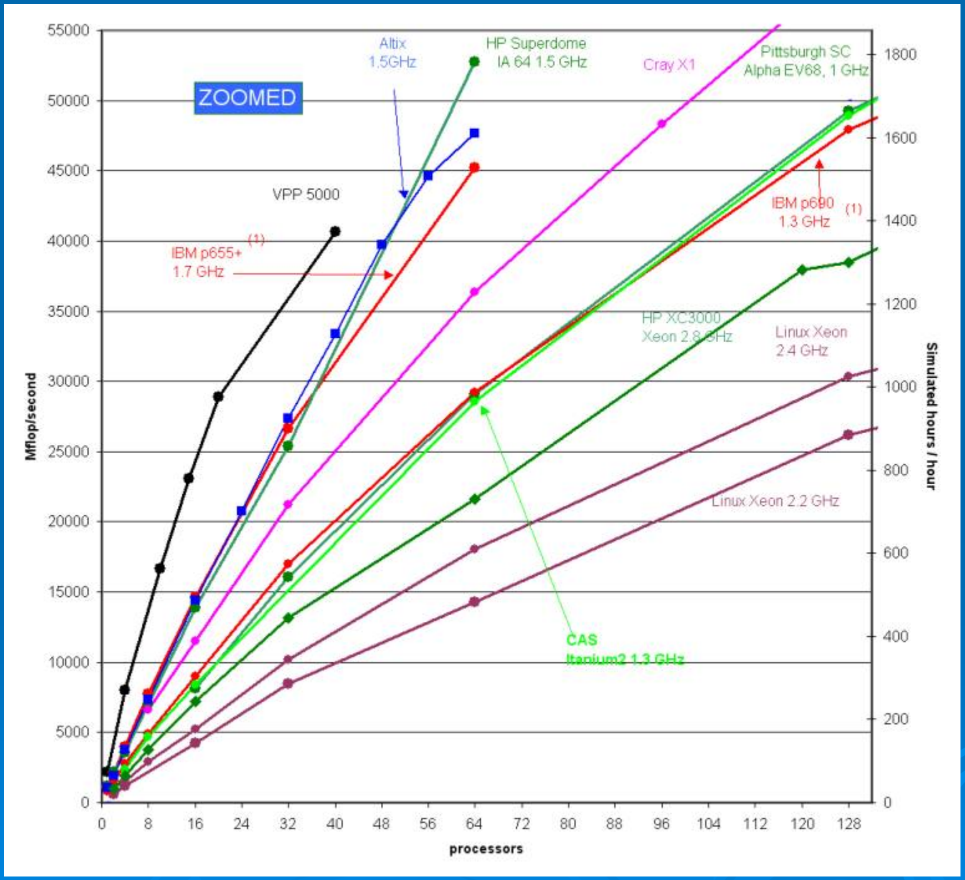
- Runs locally on a dedicated non-AWIPS Linux server
  - ~2 GHz Dual CPU Linux system
  - Portland Group F90 compiler
  - Output grids viewed in AWIPS and GRASS GIS
- 00 Z GFS model initializations
- 2 model domains
  - Outer: 150x90 grid, 27 km spacing — 60 hr run
  - Inner: 114x120 grid, 9 km spacing — 36 hr run
  - Lambert Conic Conformal projection
- Simulation runs take ~5 hours (1:30 to 6:30 local time)
- Runs on cron, automatically
- 2 OHRFC staff members sent to UCAR training in 2003

OHRFC Linux Cluster

- 1 head node
  - Dual CPU
  - 2.4 GHz
  - 2 ethernet cards
- 4 nodes
  - Dual CPU
  - 2.8 GHz
- Gigabit ethernet
- Under construction



MM5 Cluster Performance



References

• Cressie, N., 1993. Statistics for spatial data. John Wiley & Sons, New York.
• Gel, Y., A.E. Raftery, and T. Gneiting, 2004. “Calibrated Probabilistic Mesoscale Weather Field Forecasting: The Geostatistical Output Perturbation Method”, J. Am. Stat. Assoc., 99(467).
• Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press, Oxford and New York.
• Isaaks, E.H. and Srivastava, R.M., 1990. An introduction to applied geostatistics. Oxford University Press, New York.
• Kitanidis, P.K., 1997. Introduction to geostatistics : applications to hydrogeology. Cambridge University Press, Cambridge, England.
• Venables, W.N. & Ripley, B.D., 2002. Modern applied statistics with S, 4th edition. Springer-Verlag, New York.
• Deutsch, C. V., & Journel, A. G., 1998. GSLIB: Geostatistical software library and user’s guide. 2nd Ed., Oxford University Press, Oxford.

D G Rossiter: http://www.itc.nl/personal/rossiter/

R: http://www.r-project.org/
R spatial projects: http://sal.agecon.uiuc.edu/csiss/Rgeo/
GRASS: http://grass.itc.it/
gstat: http://www.gstat.org/
gslib: http://www.gslib.com/

Discussion

Questions?