SAND2012-5311C

Predictive uncertainty analysis of a highly parameterized groundwater model: Application of null-space Monte Carlo

Hongkyu Yoon, Sean A. McKenna, David B. Hart

Geoscience Research and Applications

Sandia National Laboratories

CMWR 2012

June, 2012

Sandia National Laboratories | U.S. Department of Energy, Office of Science | CFSES | Texas

This material is based upon work supported as part of the Center for Frontiers of Subsurface Energy Security, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001114.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Motivation

- Quantify predictive uncertainty in inverse parameter estimates
  - Basis for probabilistic predictive modeling
  - Computationally efficient and accurate
  - Robust for strongly heterogeneous media
- Culebra dolomite as a motivating example
- Calibration-constraint null-space Monte Carlo (NSMC) method to test and develop a practical means of addressing predictive uncertainty

Faster, Cheaper and At Least As Good!

Heterogeneous Field Parameterization

- “Pilot Points” (similar to a kernel-based approach)
- Choose locations in the model domain and update their properties to produce better fit to measured heads (“calibration points”)
- Spread influence of each point to neighboring model cells by using the spatial covariance function as a weighting scheme

Highly parameterized model

- For a traditional application of inverse modeling, the rule of thumb is that $n << m$
  - For numerical stability in the solutions, the number of parameters (n) had to be considerably less than the number of observations (m)
- This constraint still holds if each parameter is completely independent of all others
  - Rarely is this the case and almost certainly not the case for pilot points defining a spatially correlated T field
- Most applications of calibration by the Pilot Point method have considerably more estimated parameters than observed data
  - Need to reduce the effective number of parameters through regularization (e.g., supplementary information about parameters, relations between parameters) and subspace analysis (e.g., singular value decomposition)

Inverse Modeling

Linear Model

$$\mathbf{c} = \mathbf{X}\mathbf{b}$$

$$\Phi = (\mathbf{c} - \mathbf{X}\mathbf{b})^t\mathbf{Q}(\mathbf{c} - \mathbf{X}\mathbf{b})$$

$$\mathbf{b} = (\mathbf{X}^t\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Q}\mathbf{c}.$$

Non-Linear Model

$$\mathbf{c_0} = M(\mathbf{b_0})$$

$$\mathbf{c} = \mathbf{c}_0 + \mathbf{J}(\mathbf{b} - \mathbf{b}_0)$$

$$\Phi = (\mathbf{c}' - \mathbf{c}_0 - \mathbf{J}(\mathbf{b} - \mathbf{b}_0))^t\mathbf{Q}\,(\mathbf{c}' - \mathbf{c}_0 - \mathbf{J}(\mathbf{b} - \mathbf{b}_0))$$

$$\mathbf{u} = (\mathbf{J}^t\mathbf{Q}\mathbf{J})^{-1}\mathbf{J}^t\mathbf{Q}(\mathbf{c}' - \mathbf{c}_0)$$

SVD-Assist

$$\mathbf{c} = \mathbf{c}_0 + \mathbf{J}(\mathbf{b} - \mathbf{b}_0)$$

$$\mathbf{u} = (\mathbf{J}^t\mathbf{Q}\mathbf{J})^{-1}\mathbf{J}^t\mathbf{Q}(\mathbf{c}' - \mathbf{c}_0)$$

← *n* parameters →

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial O_1}{\partial P_1} & \frac{\partial O_1}{\partial P_2} & \frac{\partial O_1}{\partial P_3} & \cdots & \frac{\partial O_1}{\partial P_n} \\ \frac{\partial O_2}{\partial P_1} & \frac{\partial O_2}{\partial P_2} & \frac{\partial O_2}{\partial P_3} & \cdots & \frac{\partial O_2}{\partial P_n} \\ \frac{\partial O_3}{\partial P_1} & \frac{\partial O_3}{\partial P_2} & \frac{\partial O_3}{\partial P_3} & \cdots & \frac{\partial O_3}{\partial P_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial O_m}{\partial P_1} & \frac{\partial O_m}{\partial P_2} & \frac{\partial O_m}{\partial P_3} & \cdots & \frac{\partial O_m}{\partial P_n} \end{bmatrix}$$

↕ *m* observations

- Apply singular value decomposition (SVD) to Jacobian (the sensitivity matrix) to identify linear combinations of “sensitive” parameters – “super-parameters”
  - Truncated SVD by selecting parameters whose eigenvalues are smaller than a threshold value

- Decrease dimensionality of parameter estimation (factor of 3-10)
- Retain low frequency heterogeneity, but not small-scale heterogeneity

Null-Space Monte Carlo

$$c = Xb + \varepsilon$$

Full accounting for all parameters in physical system

$$X = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

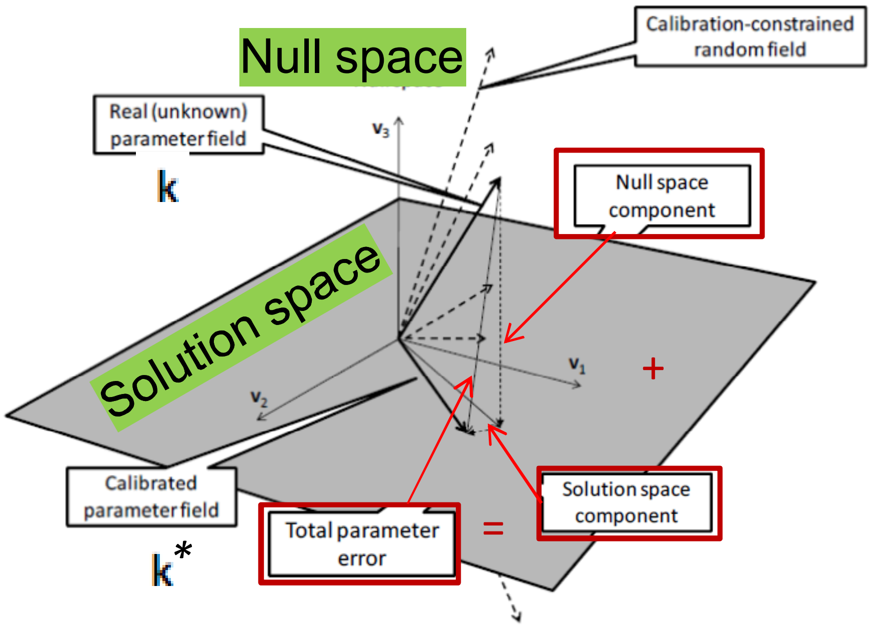
Separation of singular values into > threshold and < threshold

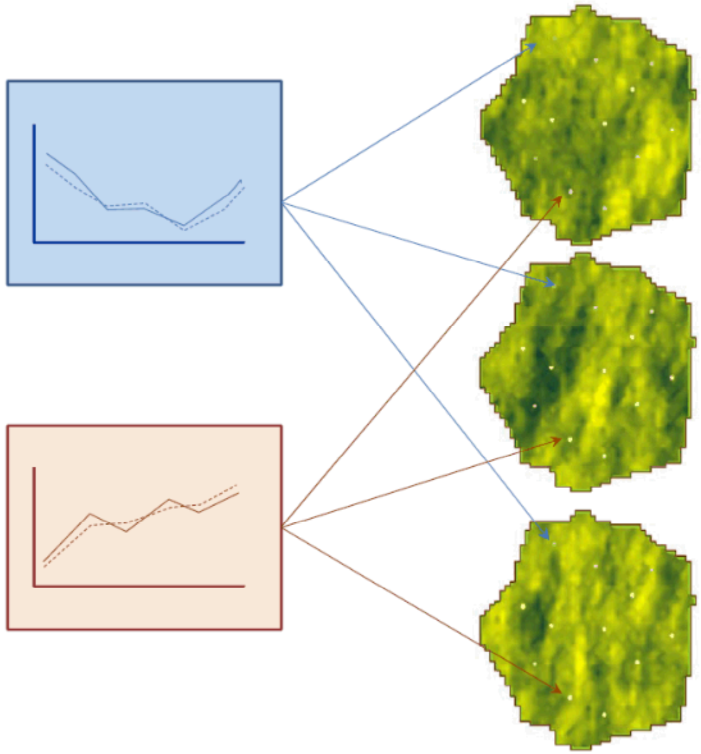
Replace X with parameters that significantly impact the calibration on given available information

$$c = U_1 S_1 V_1^T b + \varepsilon$$

U and V each contain orthogonal unit vectors covering the model range and domain (parameter) space

NSMC Concept





Null space Monte Carlo generation of Calibration-constraint random fields

NSMC method can be used to compute many different parameter fields which all calibrate the model

From Young et al., 2009, Draft Report, Texas Water Development Board

Sampling Null Space

$$b - b^* = -(I - R)b + G\varepsilon$$

Unknown error between true and estimated parameters

Use information in *C(b)* and *C(ε)* to define *C(b-b*)*

$$C(b - b^*) = -(I - R)C(b)(I - R)^T + GC(\varepsilon)G^T$$

Parameter error from calibration null space (V_2)

Errors of estimates in solution space (V_1) driven by measurement noise

$$(b^* - b^*_{ST})' = V_2 V_2^T (b^* - b^*_{ST})$$

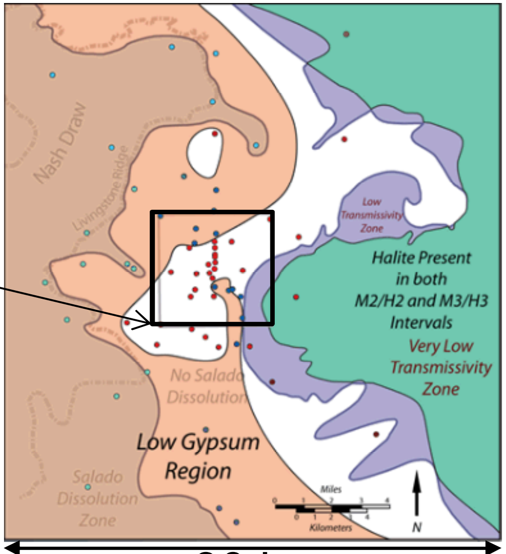
Project stochastic parameter differences that have zero impact on calibration into null space

Moore, C. and Doherty, J., 2005. The role of the calibration process in reducing model predictive error, Water Resources Research, Volume 41, No 5.

Tonkin M., and Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques, Water Resources Research, 45

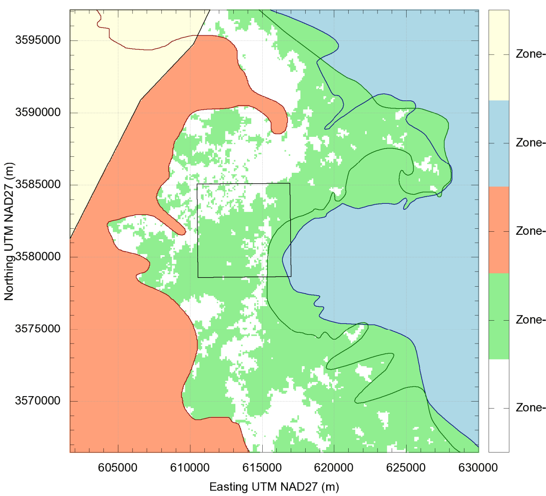
Culebra Dolomite

- The Culebra Dolomite near the WIPP site (NM,USA) => predictive performance measure is advective transport time to a prescribed boundary
- Observation data include two decades of steady-state head measurements and pumping test results



~ 30 km

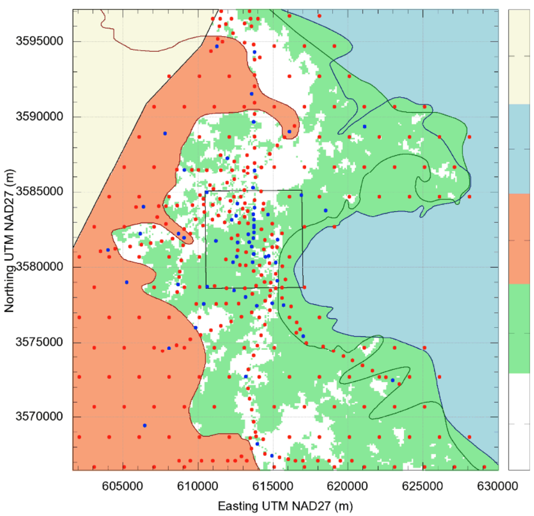
- Sequential indicator simulation (SISIM) generated the stochastic zones over the domain (Hart et al., 2009)
  - Both hard and soft data are used together
- This produces a large number of equally likely indicator fields, where high and low T may exist in the central zone



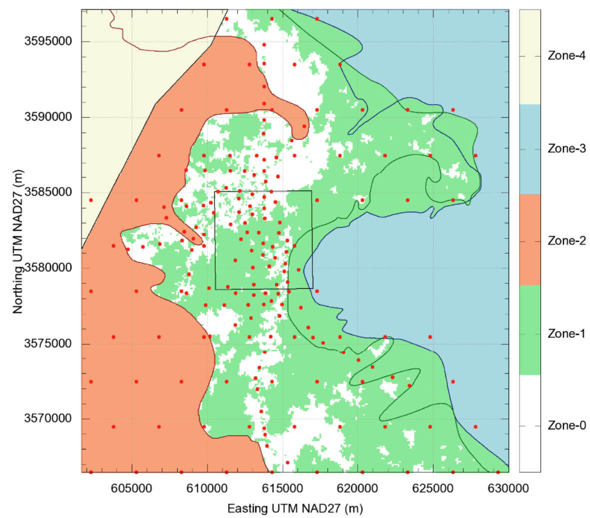
Three Property Fields

- Simultaneous estimation of three spatially correlated property fields

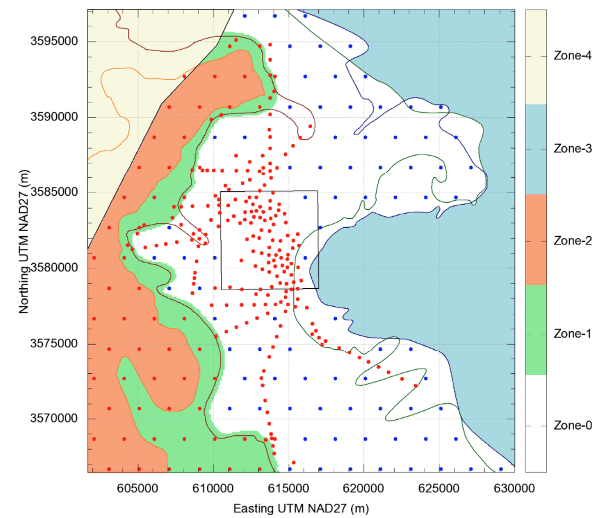
Transmissivity



Anisotropy



Storativity



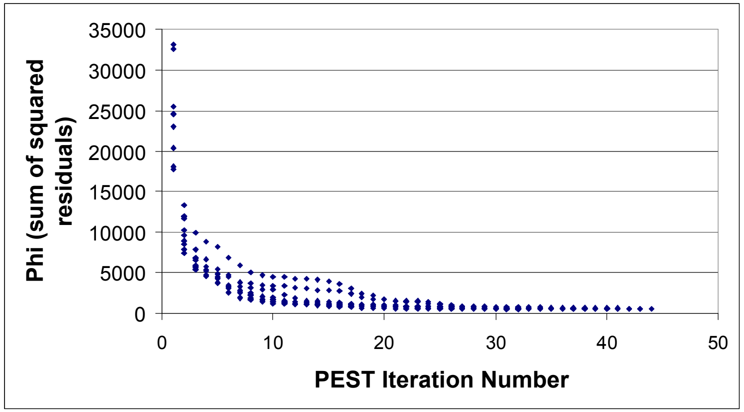
Locations in blue are fixed values and red are locations where parameters are adjusted. For transmissivity, 2 fields (Zones 0 and 1) are estimated, then combined

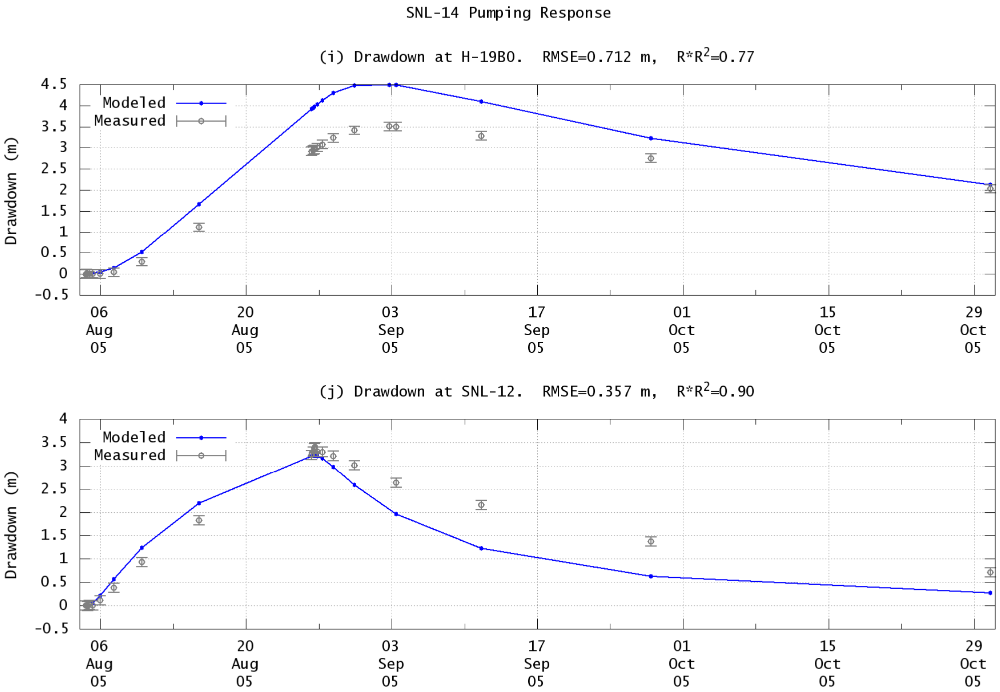
1380 observations, ≈ 1200 parameters

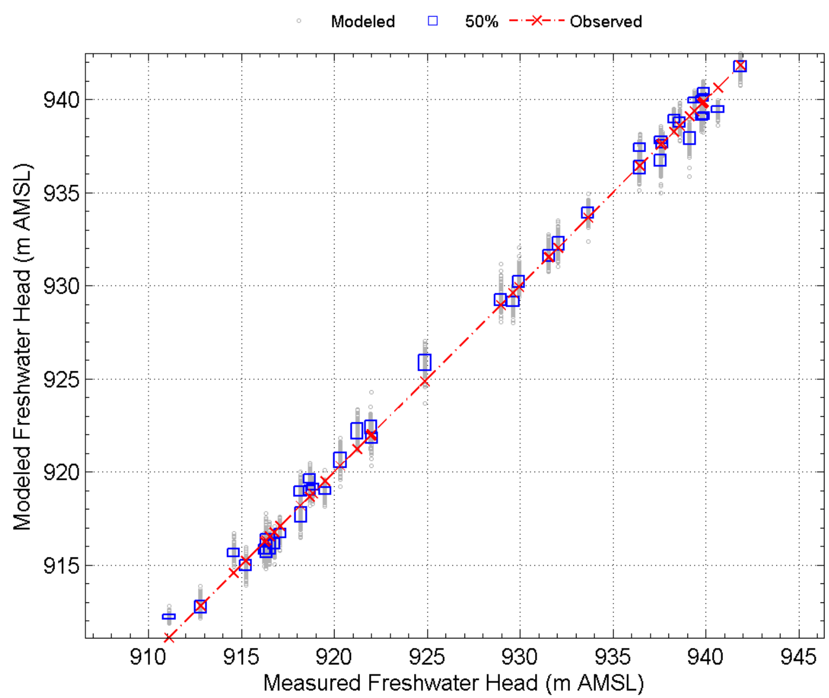
MSP Calibration Process

200 starting fields are calibrated to 1380 steady-state head and transient drawdown observations

Computation time is <u>several months</u>

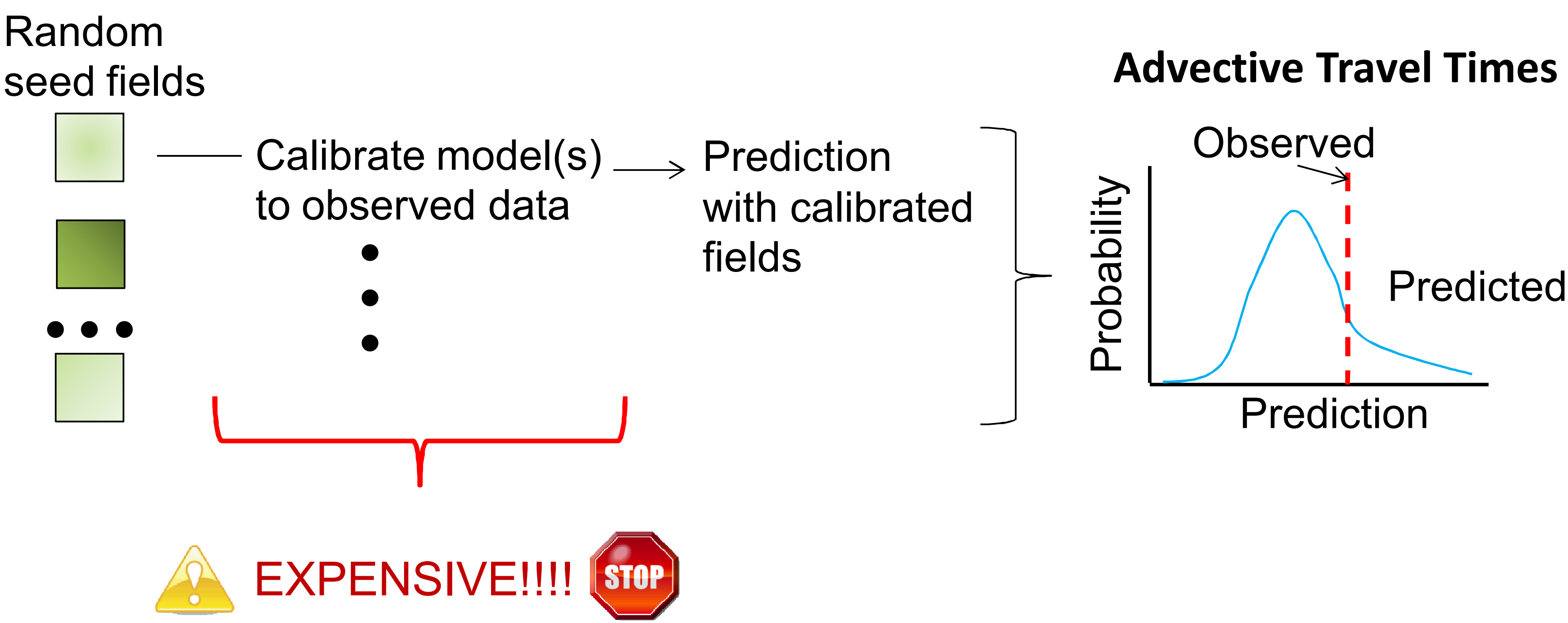






Approach 1: MSP

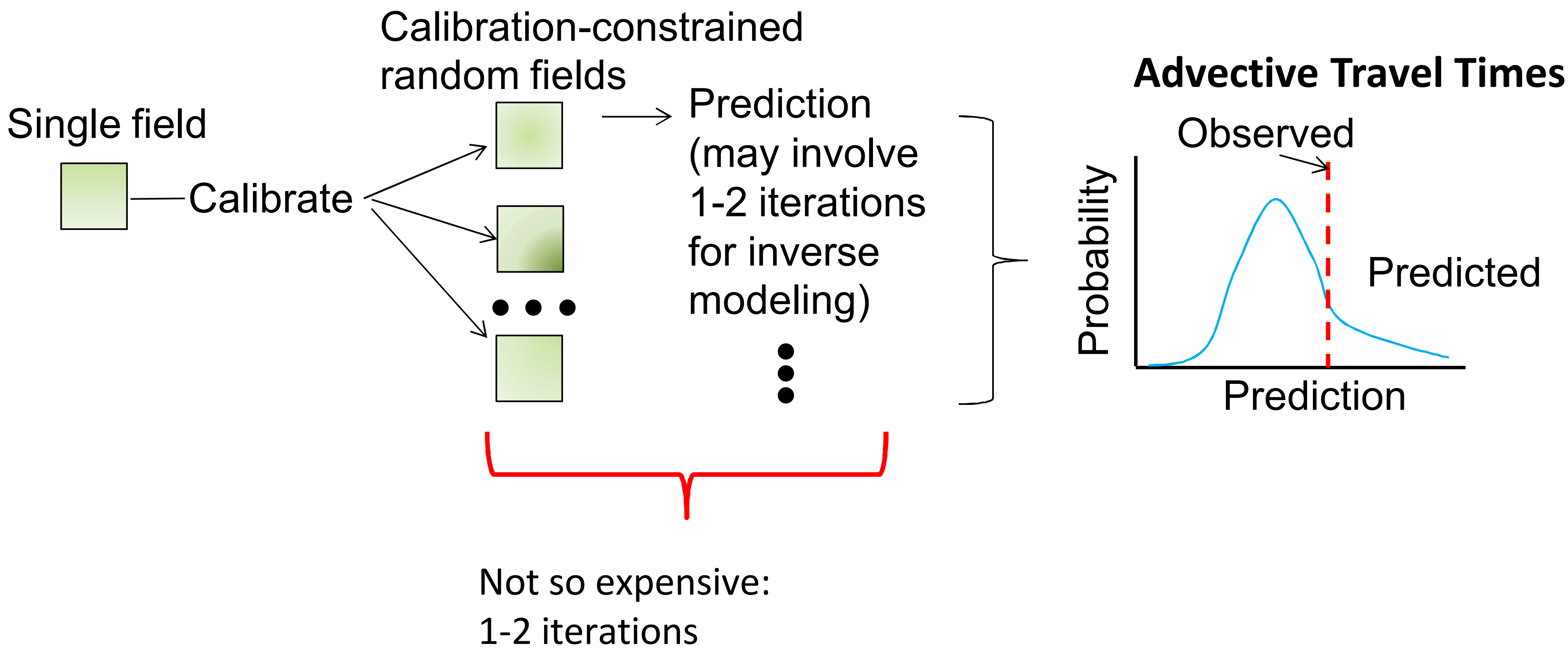
- Conceptual model is stochastic – poorly known location and extent of highly fractured zones. Leads to multiple, equally probable starting points (seed fields for inversion). Calibrate each seed field.
- Stochastic Inverse Modeling (~ Ramarao et al (1995, WRR))



Up to 50 iterations

Approach 2: NSMC

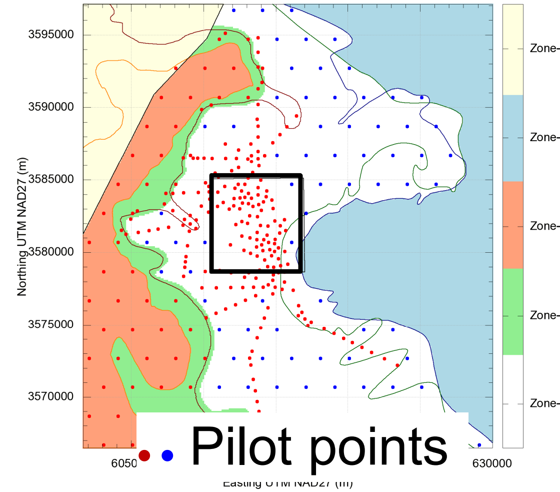
- Shift the approach: Calibrate one seed field and then modify that field to create probabilistic distribution of results
- Null-Space Monte Carlo Method (~ Tonkin & Doherty (2009,WRR))



Ground Water Parameter Estimation

- Calibrated parameters (>1200 parameters in total)

  transmissivity (T), horizontal hydraulic anisotropy, storativity (S), and a section of recharge
- 200 multiple random seed fields are calibrated to observed data, and then 100 best fields are selected for travel-time analysis

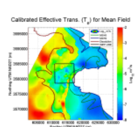
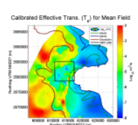
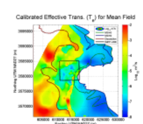
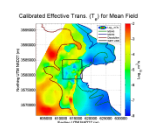
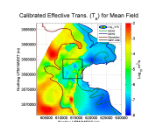


NSMC Analysis (5 groups -21 fields)

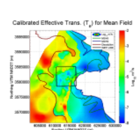
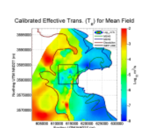
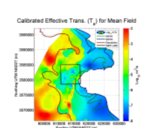
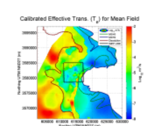
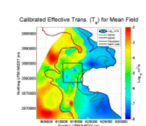
Percentile in 200 multiple fields

Best/fastest 25th 50th 75th Worst/slowest

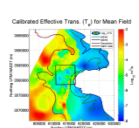
(Un)calibrated Φ-based five fields

(Un)calibrated travel time-based five fields

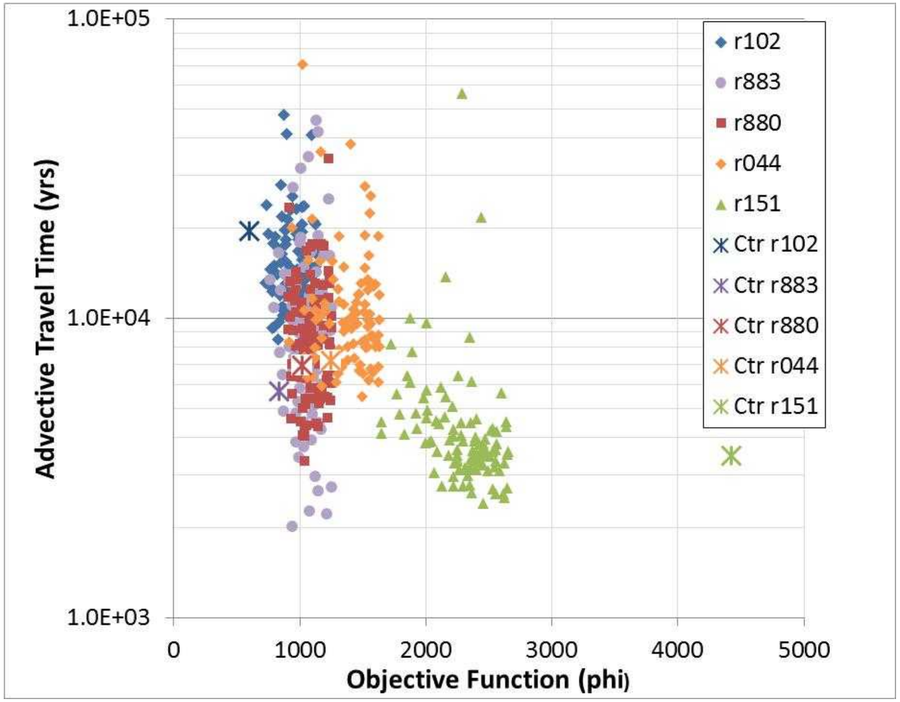
Mean field of 200 random seed fields



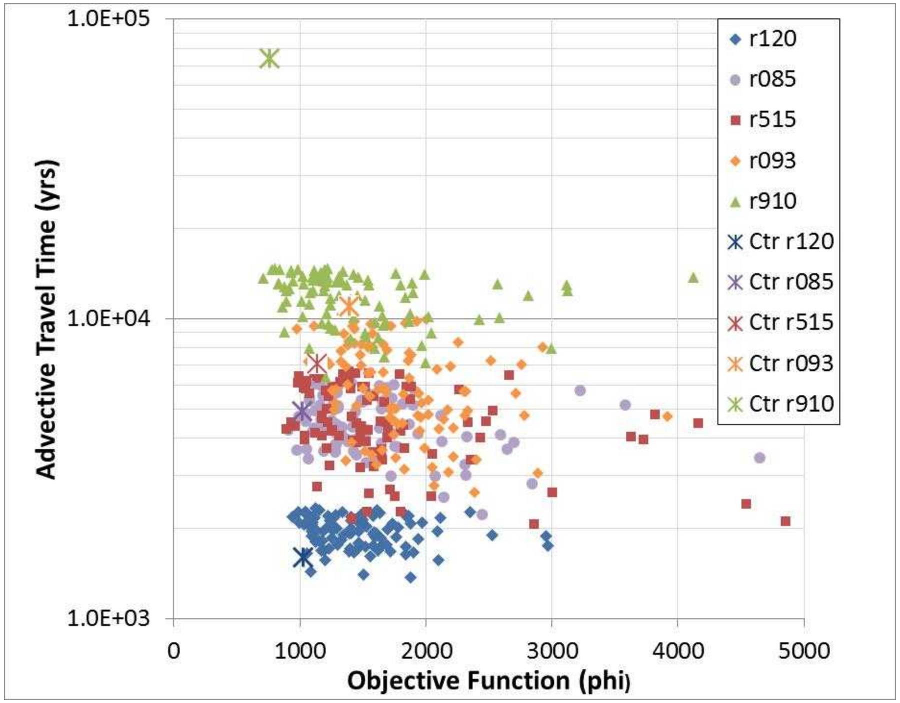
NSMC generation of 200 random fields from each calibrated field

Travel Time and Obj. Function

Objective Function Selections



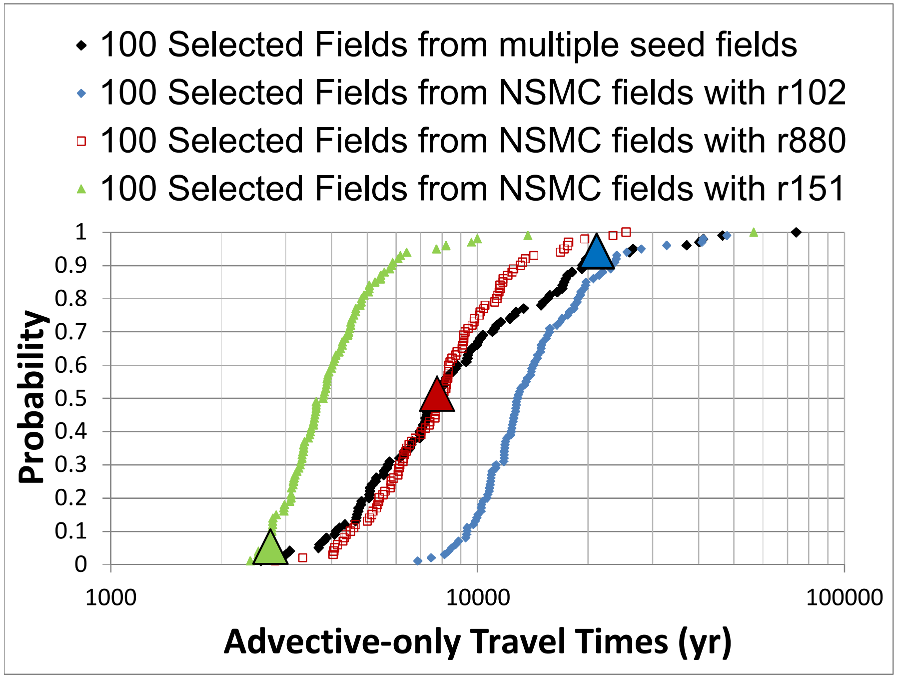
Travel Time Selections



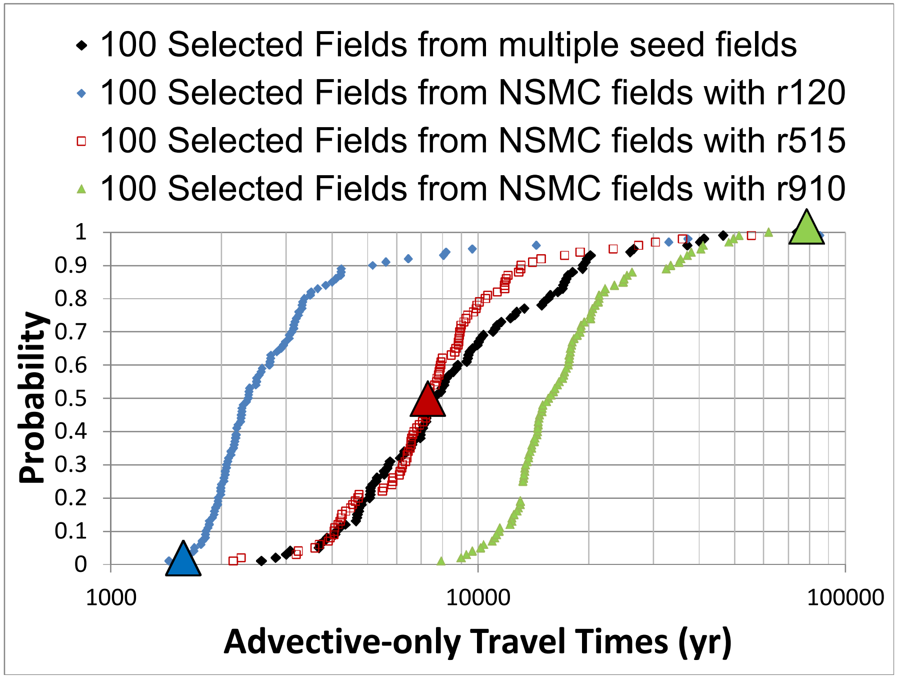
5 calibrated fields selected with each criterion
100 NSMC realizations shown for each calibrated field

Travel Time Distributions

Objective Function Selections

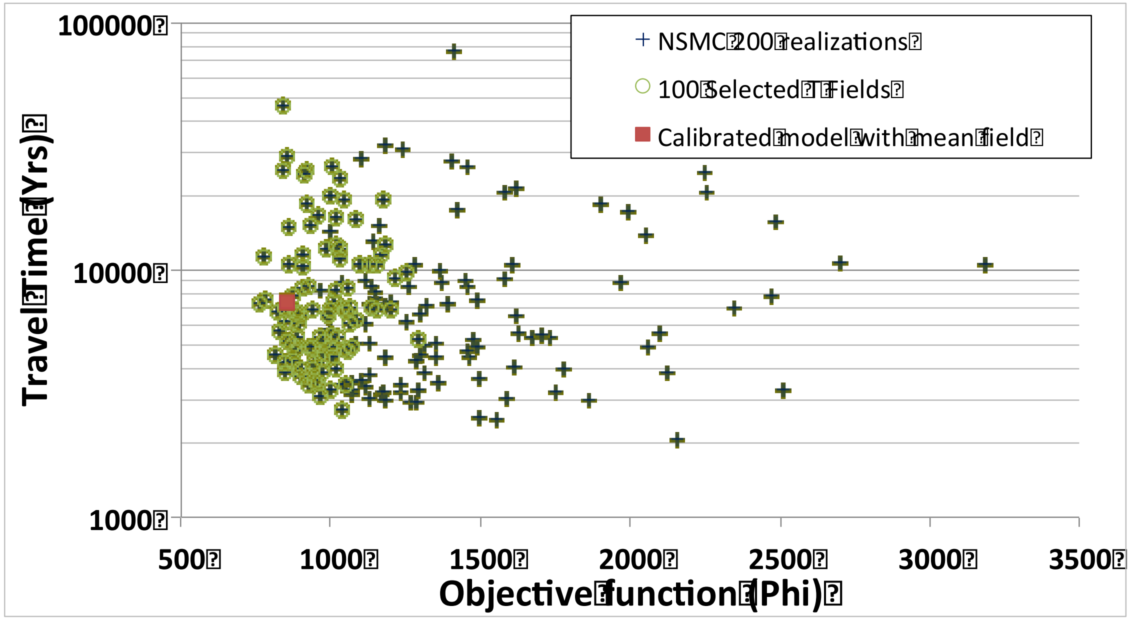


Travel Time Selections

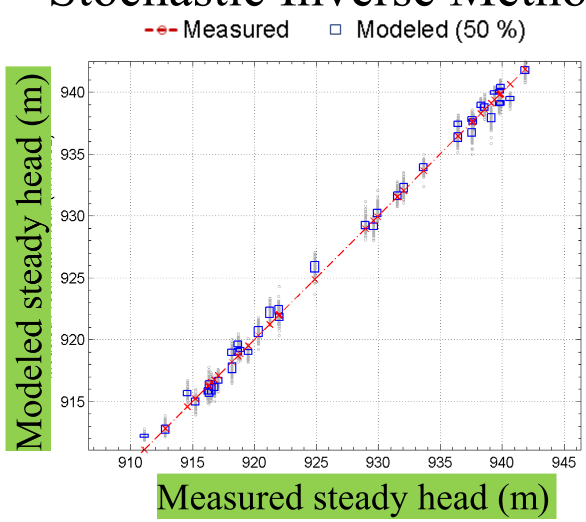


3 calibrated fields selected with each criterion (low, median, high)
Original MSP calibration results
100 NSMC realizations shown for each calibrated field

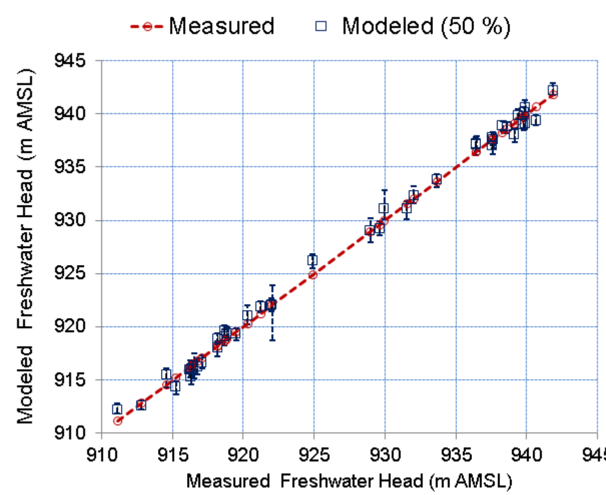
Mean field of 200 multiple seed fields can be used to get a calibrated model for NSMC method



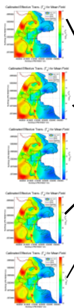
Stochastic Inverse Method



NSMC method with mean field

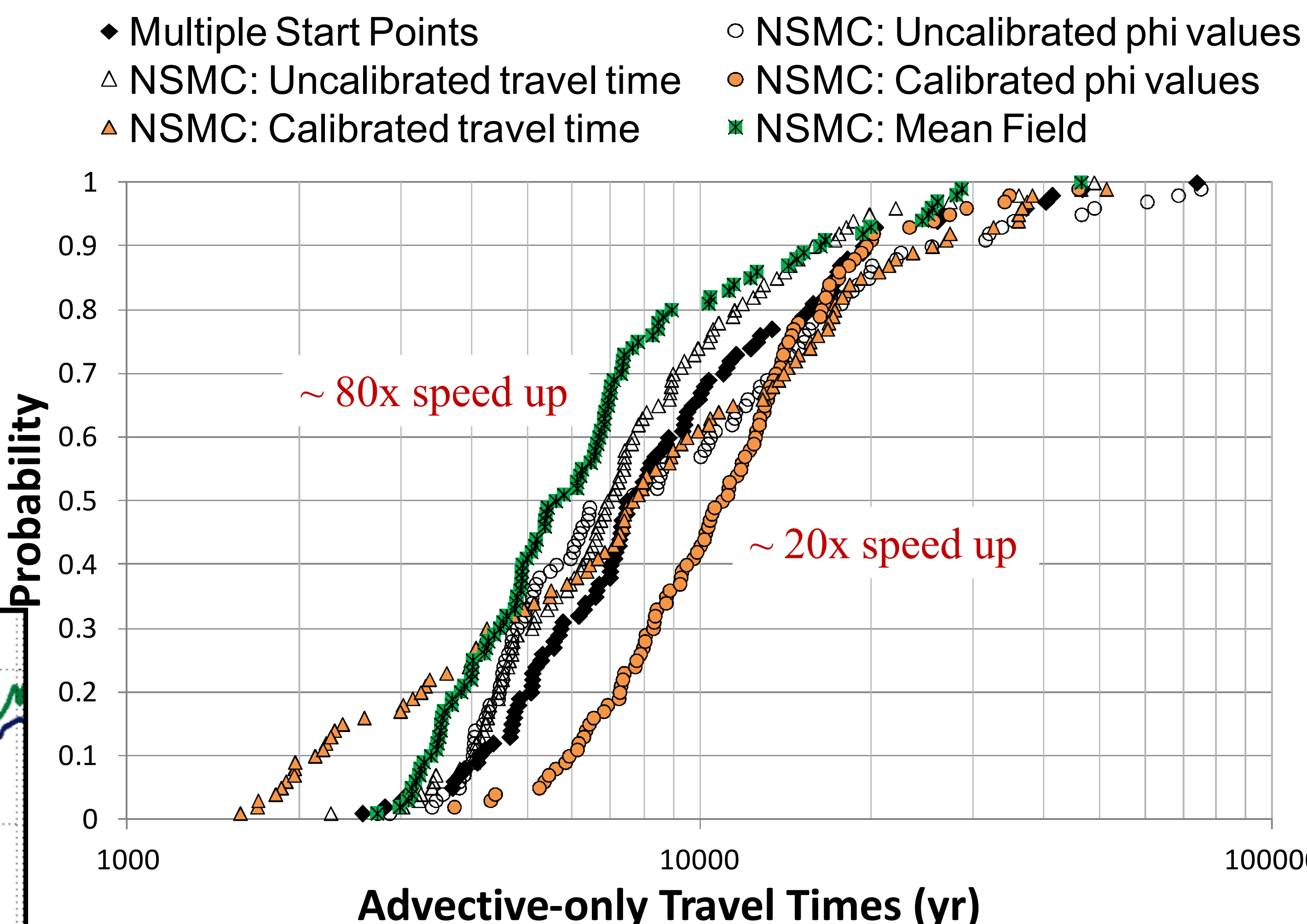


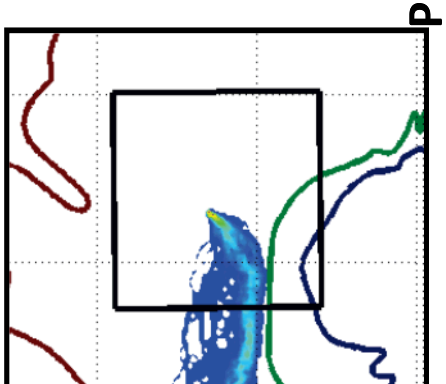
Comparison of travel times for different methods shows the effectiveness of the NSMC method



40 RFs from each calibrated model

100 Selected fields



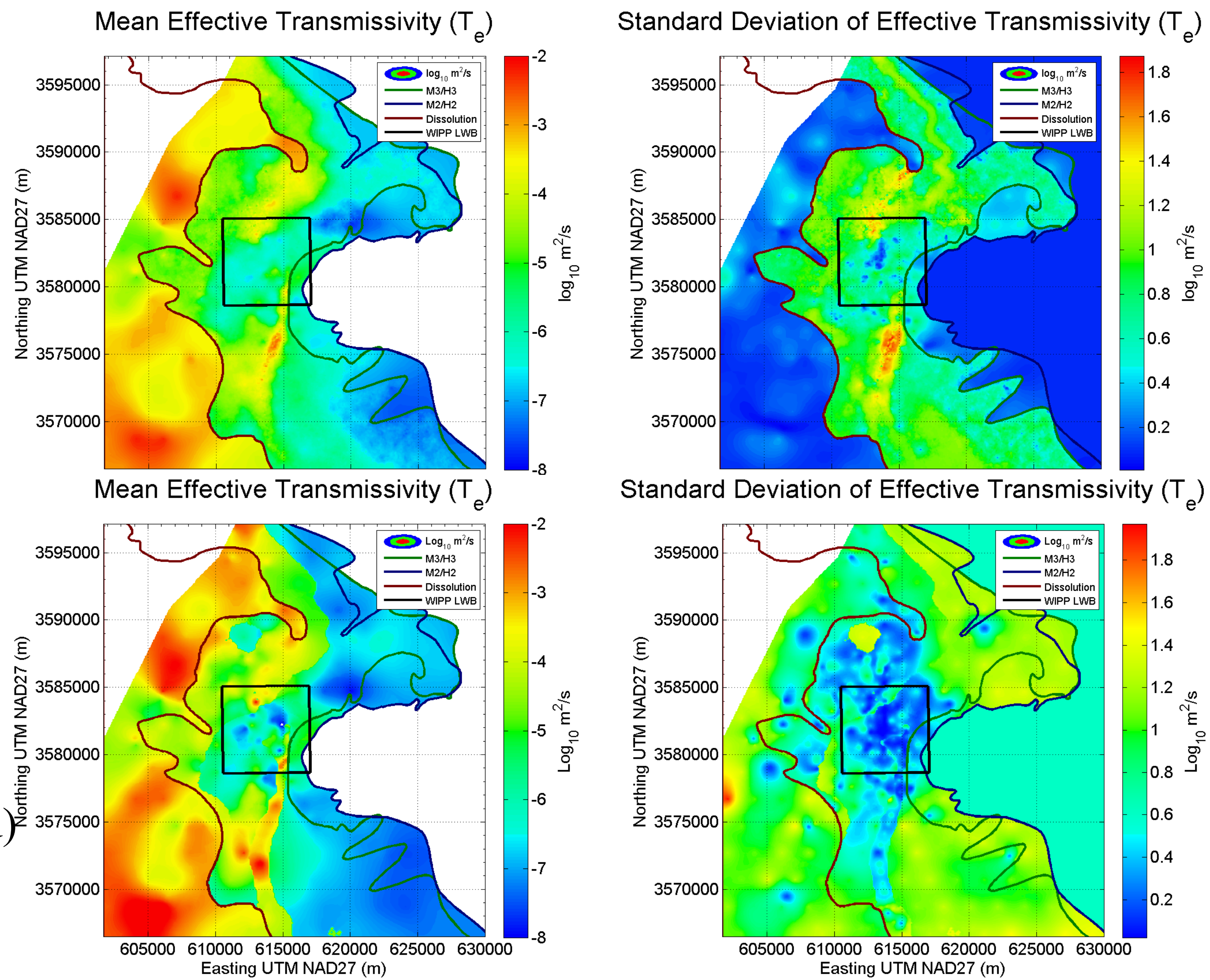


Effect of high T channel

T_e has a similar distribution for both fields, but S.D. of T_e distribution is quite different

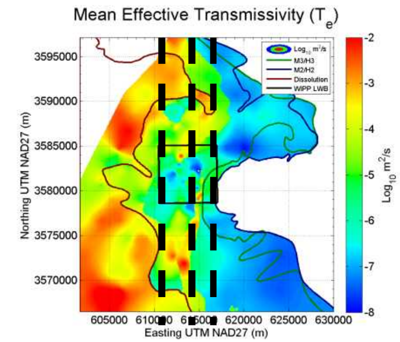
100 selected Fields MSP (Multiple Starting Point)

100 selected Fields (NSMC method with the mean field)



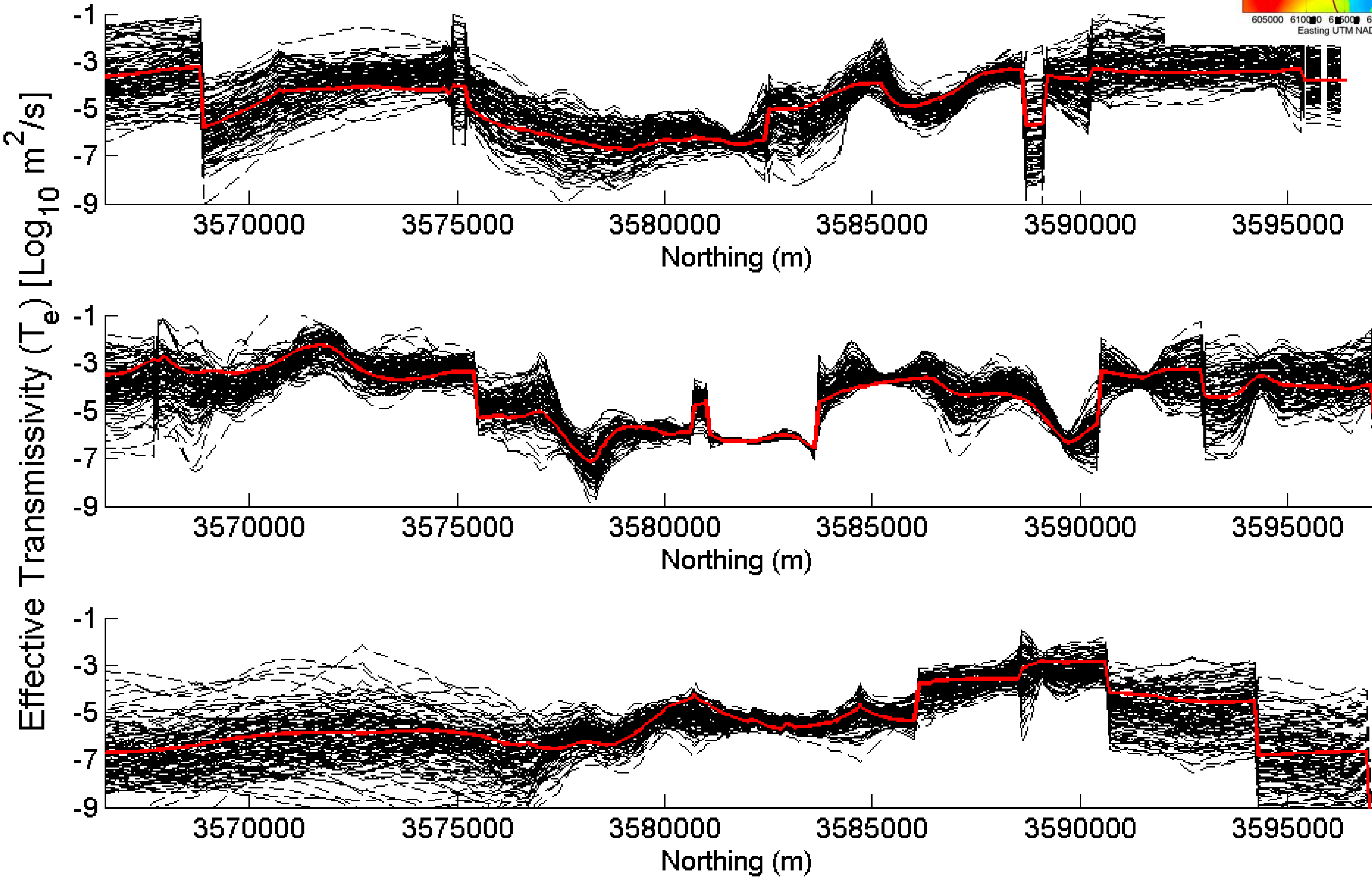
Sandia National Laboratories

T_e distribution along three transects capture calibrated T_e trends

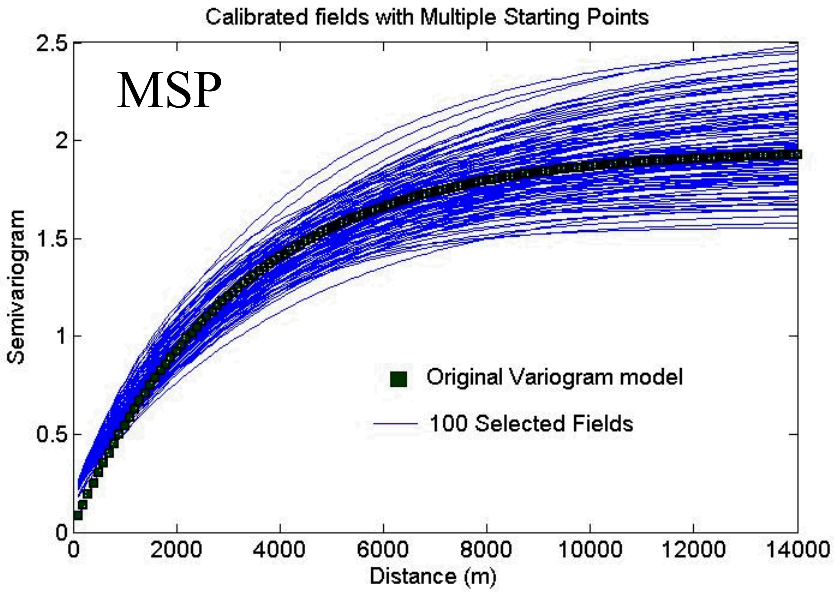


Calibrated model with mean field

NSMC 100 selected fields



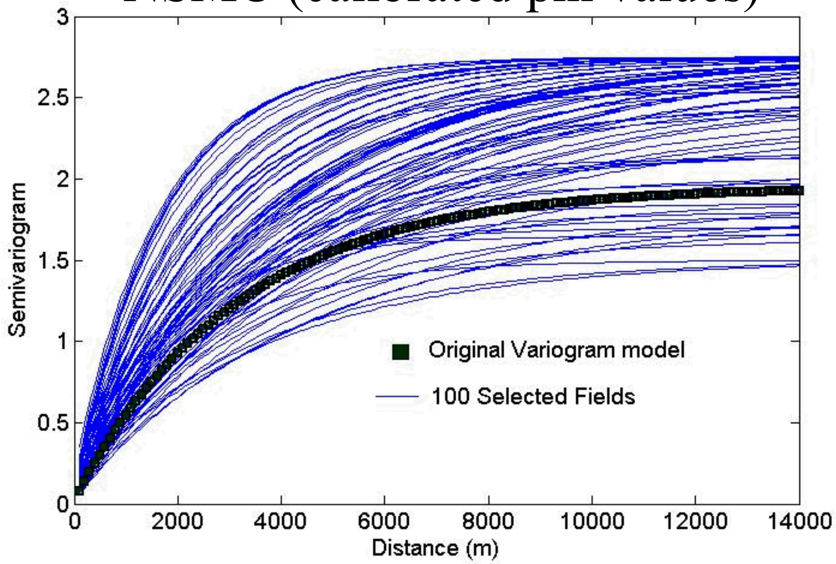
Variograms



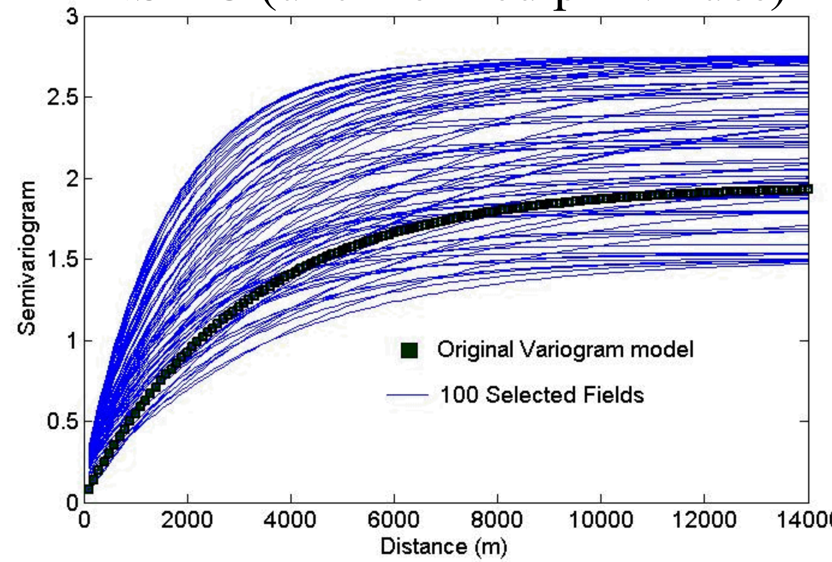
Variograms of 100 selected MSP Fields follow the original variogram model well with a small variation

For all NSMC sampling method, variograms of selected fields have more variations

NSMC (calibrated phi values)



NSMC (uncalibrated phi values)



Conclusions

- Can NSMC approach approximate ensemble predictions obtained with MSP runs?
  - Yes, but the calibration constraint will bias estimates and predictions to values proximal to the initial calibration
- Given a set of previously run models, what is an effective means of expanding the predictive ensemble?
  - Select final ensemble from larger set of NSMC realizations using calibration quality and non-uniform sampling from NSMC realizations
- Without existing calibrations, can a mean-field representation and/or initial forward runs serve as an initial starting point?
  - Yes, NSMC realizations provide good match to observations and reasonable approximation of MSP predictive distribution

Questions