TOUGH Short Course

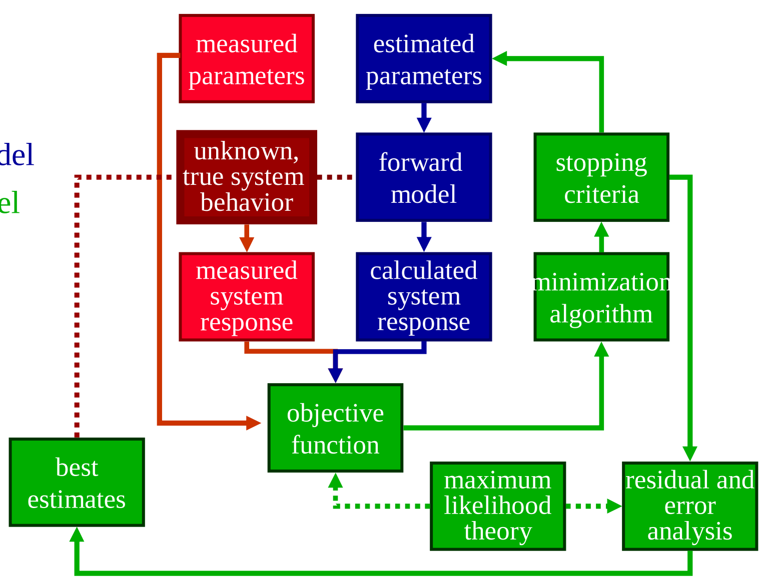
Lawrence Berkeley National Laboratory
Earth Sciences Division
Berkeley, California

iTOUGH2 Theory

- Definitions
- Stochastic Model
- Objective Function
- Minimization Algorithm
- Residual and Error Analysis

Inverse Modeling Procedure

True system
Data
Forward model
Inverse model



Calibration Point, Weighted Residual

- A *calibration point* is a point in space and time at which the observed system response z_i* and the calculated system response z_i will be compared during model calibration.
- The *residual* is the difference between the observed and calculated system response at calibration point *i*:

$$r_i = z_i * - z_i \qquad i = 1 \ldots m$$

- The *weighted residual* is the residual weighted by the inverse of the assumed measurement error:

$$y_i = \frac{z_i * - z_i}{\sigma_i} \qquad i = 1 \ldots m$$

3

Jacobian Matrix

- The Jacobian J is an $m \times n$ matrix holding the *sensitivity coefficients*
- The sensitivity coefficients are the partial derivatives of the calculated system response z_i at all calibration points *i*, i=1…m, with respect to each of the parameters p_j, j=1…n

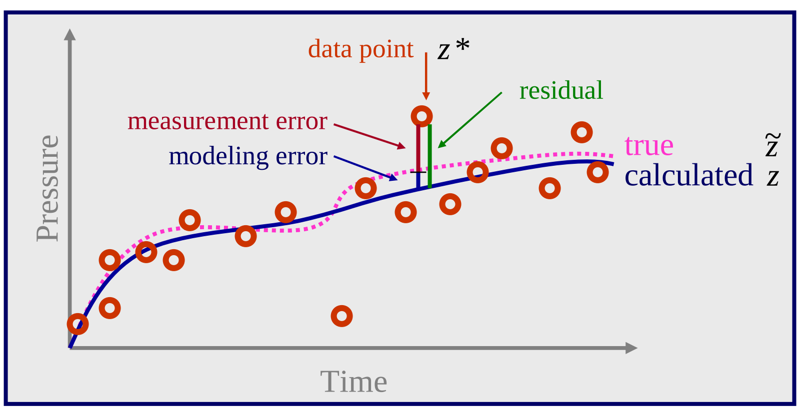
$$J_{ij} = \frac{\partial z_i}{\partial p_j}$$

4

The Stochastic Model

- Types of errors
- Functional and stochastic model
- Observation covariance matrix C_{zz}
- *a priori* error variance σ_0^2

Measurement Error, Modeling Error, and Residual



- Measurement error $e_d = z^* - \tilde{z}$
- Modeling error $e_m = z - \tilde{z}$
- Residual $r = e_d - e_m = (z^* - \tilde{z}) - (z - \tilde{z}) = z^* - z$

Systematic and Random Components

data point	=	true system state	+	measurement error
II		systematic identified conceptual functional		random unidentified distributional stochastic
calibration point	=	simulated system state	+	residual

7

Functional and Stochastic Models

- Functional Model
  - Attempts to capture the systematic, identifiable aspects
  - Attempts to mimic the true system state
  - Is represented by the governing equations of the forward model
- Stochastic Model
  - Describes random, unidentifiable aspects
  - Includes:
    - A distributional assumption about the final residuals
    - Estimate of expected size of residuals (*not* measurement errors!)
  - Is represented by the observation covariance matrix C_{zz}
- If the Functional Model were perfect, C_{zz} would be a stochastic description of the measurement errors.

8

Observation Covariance Matrix $\mathbf{C}_{zz}$

- $m \times m$ diagonal matrix
- Scales data of different quality
- Scales data of different type
- Weights fitting error
- Summarizes the Stochastic Model

$$\mathbf{C}_{zz} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_i^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_n^2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma_j^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \sigma_m^2 \end{bmatrix} = \sigma_0^2 \cdot \mathbf{V}_{zz}$$

9

Stochastic Model Development

Before inversion (*a priori*):

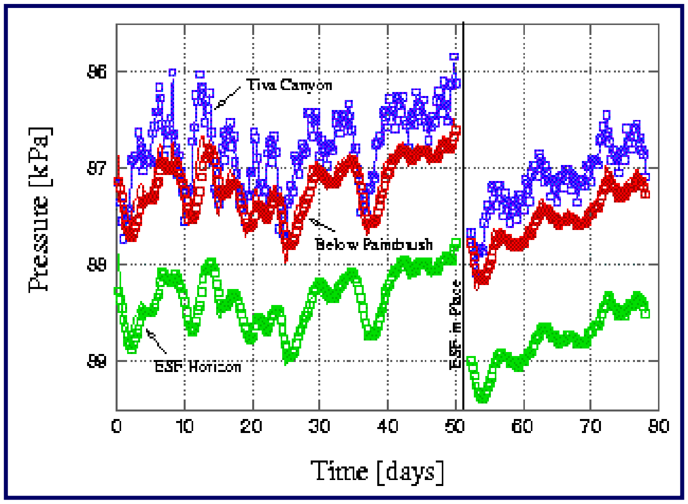
- Estimate *expected distribution of residuals* (type of distribution and standard deviation).
  - Consider which portion of the observed signal shall be explained by the functional model.
  - Consider *measurement errors* and also modeling errors.
  - Consider only *random* components.
  - Talk to experimentalist/data collector/data analyst!
- Set up observation covariance matrix C_{zz} (or $\boldsymbol{\sigma_0}^2\ V_{zz}$)

After inversion (*a posteriori*):

- Test randomness of residuals (look for systematic structure/bias in residuals)
- Perform Fisher Model Test (tests *a prior* error variance $\boldsymbol{\sigma_0}^2$ against *a posteriori* error variance $\boldsymbol{s_0}^2$).

10

Exercise: Determine $\mathbf{C}_{zz}$



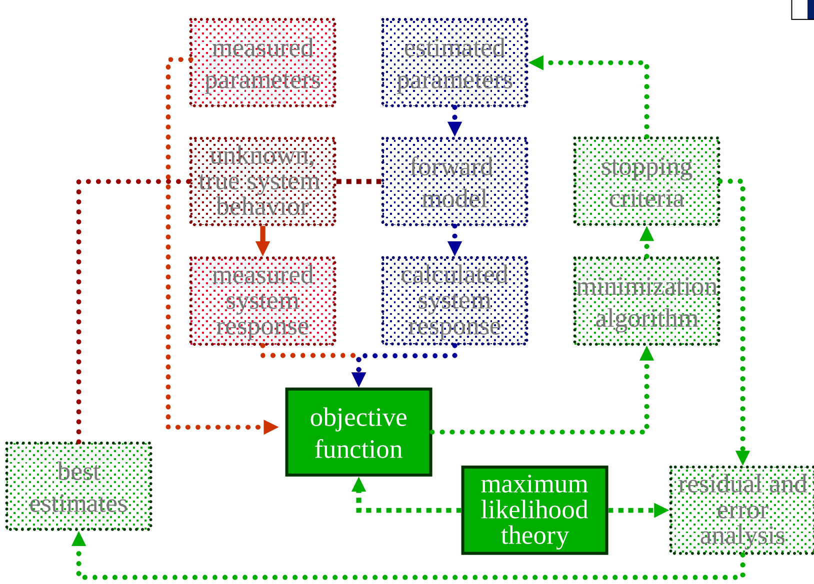
Stochastic Model: Questions

- List types of errors?
- Provide examples of systematic errors?
- Provide examples of random errors?
- How are random errors described?
- What is the purpose of the Stochastic Model?
- Why do we base the Stochastic Model on measurement errors?
- What is the value of σ_0^2?
- What is the distribution of a sum of many random errors?

Objective Functions

- Objective Function
- Maximum Likelihood
- Least Squares
- Robust Estimators

13



14

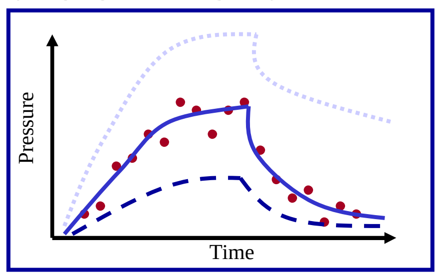
Objective Function

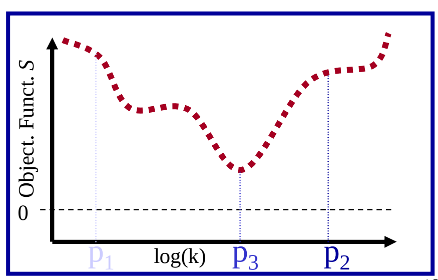
- Aggregate *measure of misfit* between measured data and model prediction.
- Scalar, ***S***; function of unknown parameters p.
- To be minimized (*minimizing **S** = improving fit*)
- Can be based on maximum likelihood considerations.
- Examples of objective functions:
  - Least absolute value (L_1)
  - *Least squares* (L_2)
  - MinMax (L_∞)
  - Others (Cauchy, Andres, Huber, …)

15

Objective Function: Aggregate Measure of Misfit

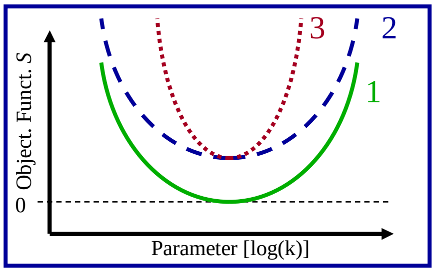
- Pressure data from well test.
- Predicted pressures depend on parameter log(*k*).
- Objective function *S* is aggregate measure of misfit.
- *S* is function of log(*k*).
- log(*k*) that minimizes *S* yields best fit.
- p_3 is considered best estimate of log(*k*).
- p_3 is parameter that most likely “produced” the observed pressure data.

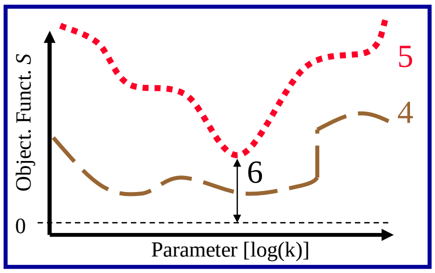




16

Objective Function for n = 1





1 ______________________
____
____

2 ______________________
____
____

3 ______________________
____
____

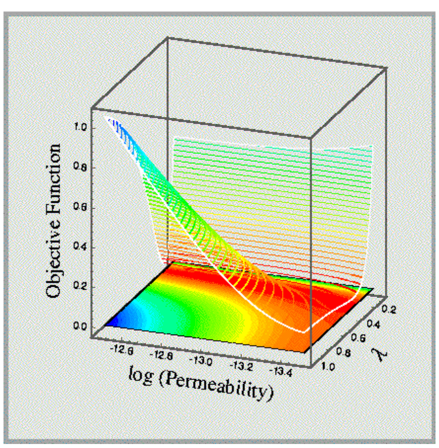
4 ______________________
____
____

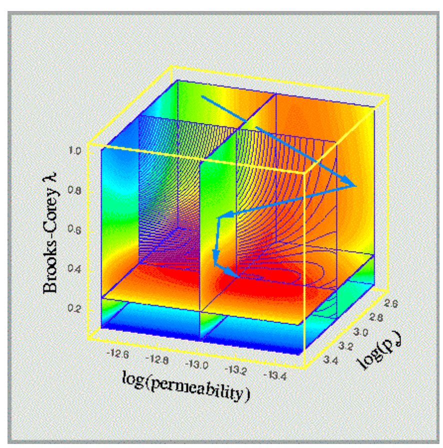
5 ______________________
____
____

6 ______________________
____
____

17

Objective Function for n=2 and n=3





18

Well- and Ill-Posed Inverse Problems

- Topology of objective function defines inverse problem…
- Data
- Sensitivity

… not the minimization algorithm!

ill-posed

well-posed

19

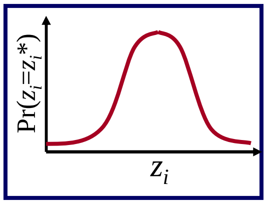
Properties of Objective Function

- ***S*** **=** ***S*****(**z(p)**)** is a *nonlinear* function in *n*-dimensional parameter space.
- Assumed *convex*, but may exhibit local minima, saddle points, long and narrow valleys, etc.
- Assumed *continuous* and *differentiable*, but may be discontinuous and numerically unstable.
- Close to *quadratic* near minimum.

20

Maximum Likelihood: General

- Probability density function



$$\Phi(\mathbf{z};\mathbf{p}) = \Pr(\mathbf{z} = \mathbf{z}^* \,|\, \mathbf{p}) = \prod \Phi(z_i;\mathbf{p})$$

- Likelihood function

$$\Phi(\mathbf{z};\mathbf{p}) \Leftrightarrow L(\mathbf{p};\mathbf{z}^*)$$

- Maximum likelihood

$$\text{maximize } L(\mathbf{p};\mathbf{z}^*)$$

$$\text{minimize } S = -2\ln\left[L(\mathbf{p};\mathbf{z}^*)\right]$$

21

Maximum Likelihood: Normal Distribution → Least Squares

- Gaussian probability density function

$$\Phi(z_i;\mathbf{p}) = (2\pi\sigma_i)^{-1/2} \exp\left[-\frac{1}{2}\frac{(z_i^* - \hat{z}_i)^2}{\sigma_i^2}\right]$$

$$\Phi(\mathbf{z};\mathbf{p}) = \prod_{i=1}^{m} \Phi(z_i;\mathbf{p}) = (2\pi)^{-m/2} \left|\mathbf{C}_{zz}\right|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z}^* - \hat{\mathbf{z}})^T \mathbf{C}_{zz}^{-1} (\mathbf{z}^* - \hat{\mathbf{z}})\right]$$

- Likelihood function

$$L(\mathbf{p};\mathbf{z}^*) = (2\pi)^{-m/2} \left|\mathbf{C}_{zz}\right|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z}^* - \hat{\mathbf{z}})^T \mathbf{C}_{zz}^{-1} (\mathbf{z}^* - \hat{\mathbf{z}})\right]$$

- Maximum likelihood

$$\text{minimize } S = (\mathbf{z}^* - \mathbf{z})^T \mathbf{C}_{zz}^{-1} (\mathbf{z}^* - \mathbf{z})$$

22

Weighted Least Squares

- *Weighted least squares*: Sum of squared residuals, weighted by inverse of observation variance
- If residuals are *normally* distributed, minimizing the least-squares objective function

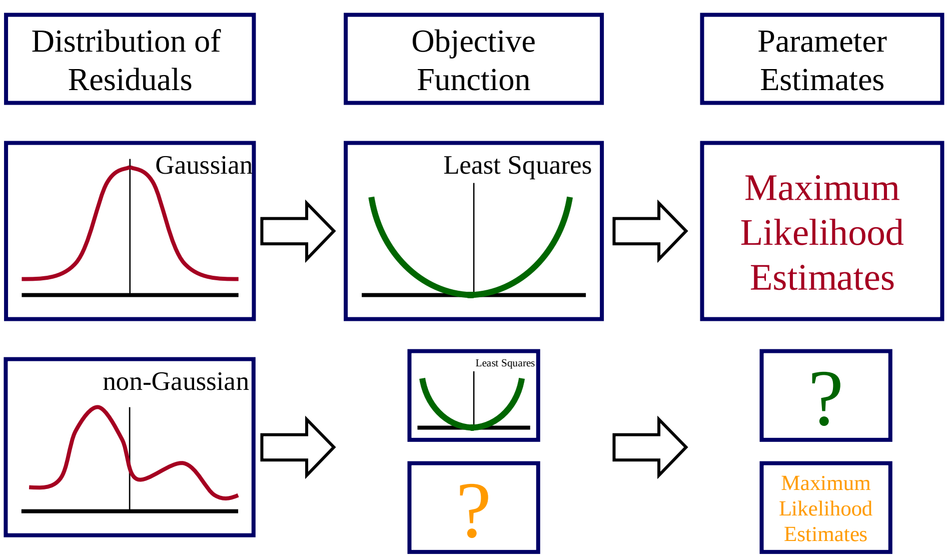
$$\text{minimize} \quad S = \sum_{i=1}^{m} \frac{(z_i * - z_i)^2}{\sigma_i^2} = (\mathbf{z} * - \mathbf{z})^T \mathbf{C}_{zz}^{-1} (\mathbf{z} * - \mathbf{z})$$

yields *maximum-likelihood estimates*.

- Least squares estimation (by definition) minimizes error variance.

23

Non-Gaussian Residuals

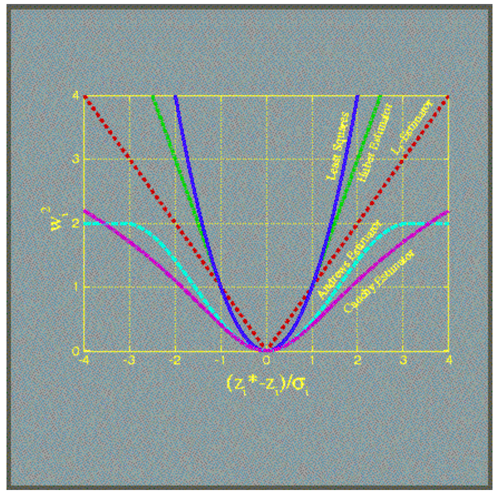


24

Robust Estimators

Robust estimators reduce the weight and thus impact of large residuals (from outliers in the data or systematic modeling errors) on the estimated parameter set.

Water Resour. Res., *34*(11), 2939-2947, November 1998



25

"Theoria Cominationis Observationum Erroribus Minimis Obnoxiae" Carl Friedrich Gauss (~1820)

The integral ∫xϕx.dx, i.e., the mean value of *x*, indicates the presence or absence of constant error, as well as its magnitude. Similarly, the integral ∫xxϕx.dx taken from x=-∞ to x=+ ∞ (the mean square of *x*) **seems most appropriate** to generally define and quantify the uncertainty of the observations. Thus, given two systems of observations which differ in their likelihoods, we will say that the one for which the integral ∫xxϕx.dx is smaller is the more precise.

Now if someone should object that this convention has been chosen arbitrarily with no compelling necessity, I will gladly agree. In fact, the problem has an intrinsic vagueness about it that can only be resolved by a more or less **arbitrary principle**. It is not out of place to compare the estimation of quantity by means of an observation subject to larger or smaller errors with a game of chance. Since any error to be feared in an observation is connected with a loss, the game in one in which nobody wins and everybody looses. We estimate the outcome of such a game from the probable loss: namely, from the sum of the product of the individual losses with their respective probabilities.

It is by no means self-evident how much loss should be assigned to a given observation error. On the contrary, the matter depends in some part on our own judgment. Clearly we cannot set the loss equal to the error itself; for if positive errors were taken as losses, negative errors would have to represent gains. The size of the loss is better represented by a function that is naturally positive. Since the number of such functions is infinite, it would seem that we should **choose the simplest function** having this property. **That function is arguably the square**, since the principle proposed above results from its adoption.

Laplace has also considered the problem in a similar manner, but he adopted the **absolute value of the error** as his measure of loss. Now if I am not mistaken, **this convention is no less arbitrary than mine**. Should an error of double size be considered as tolerable as a single error twice repeated or worse? Is it better to assign only twice as much influence to a double error or more? T**he answers are not self-evident, and the problem cannot be resolved by mathematical proofs, but only by an arbitrary decision**. Moreover, it cannot be denied that Laplace's convention violates continuity and hence resists analytic treatment, while the results that my convention leads to are distinguished by their **wonderful simplicity and generality**.

("Theory of the Combination of Observations Least Subject to Errors", translated from Latin by G.W. Stewart, SIAM, 1995; emphases added)

26

Objective Function: Questions

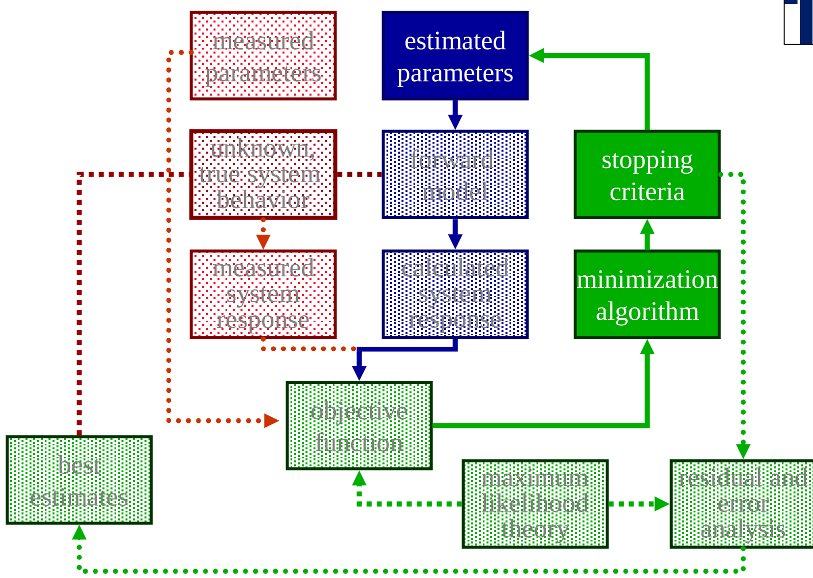
- Purpose of objective function?
- Properties of objective function?
- Well-posed inverse problem?
- Ill-posed inverse problem?
- Reasons for choosing least-squares?
- Potential problems with least-squares?
- Sketch contours of objective function (n=2) for nonlinear model, well-posed inverse problem, noisy data, correlated parameters.

27

Minimization Algorithm

- Overview and Classification
  - Principles
  - Advantages
  - Disadvantages
- Gauss-Newton
- Levenberg-Marquardt

28



29

Purpose of Minimization Algorithm

- *Find the minimum* of the objective function
- Automatically *update parameter vector* p such the the objective function ***S*** is reduced.
  - Recall: The objective function ***S*** is a function of the model output z, which is a function of the parameter vector p:

$$\text{minimize} \quad S = (\mathbf{z}^* - \mathbf{z}(\mathbf{p}))^T \, \mathbf{C}_{zz}^{-1} (\mathbf{z}^* = \mathbf{z}(\mathbf{p}))$$

30

Properties of Objective Function

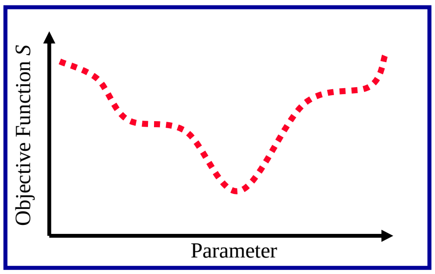
REAL

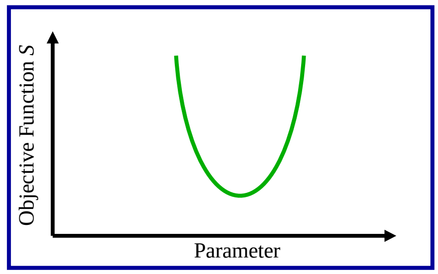
- Non-linear
- Complicated topology
- Many local minima
- Discontinuous
- Unstable

approximate ?

IDEAL

- Quadratic
- Symmetric and convex
- One global minimum
- Continuous
- Stable





31

Classification

- Direct Search Methods
  - Evaluate objective function many times
- Gradient-Based Methods
  - Move along gradient of objective function
- Second-Order Methods
  - Evaluate Hessian (or approximation to Hessian) of objective function

32

Direct Search Methods

- *Principle*
  - Map objective function by evaluating the objective function for systematically or randomly selected parameter combinations
- *Advantage*
  - No assumption about topology of objective function.
  - Obtain complete picture of parameter sensitivity and well- or ill-posedness of inverse problem
- *Disadvantage*
  - Inefficient
- *Examples*
  - Trial & Error
  - Grid Search
  - Simplex Algorithm
  - Simulated Annealing
  - Genetic Algorithm

33

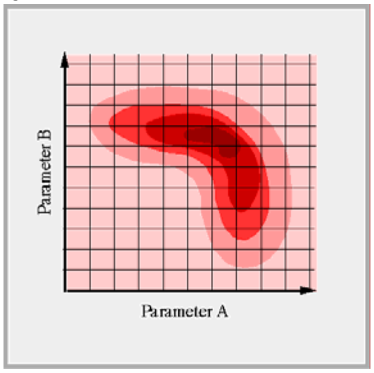
Trial & Error

- *Principle*
  - Update parameters based on expert's insight into system behavior and parameter sensitivities
- *Advantage*
  - Incorporation of "soft" information
  - Obtain feel for system behavior and sensitivities
- *Disadvantage*
  - Subjective
  - Tedious/inefficient
  - No formal error analysis

34

Grid Search

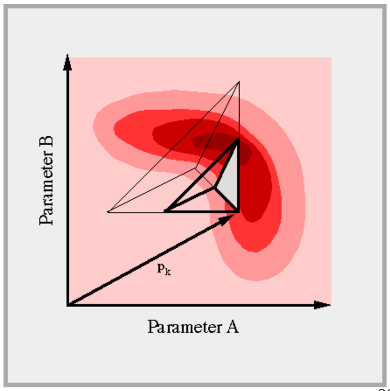
- *Principle*
  - Evaluate objective function "everywhere" in the parameter space
- *Advantage*
  - Obtain complete information:
    - Local minima
    - Sensitivities
    - Uncertainties
    - Nonuniqueness
- *Disadvantage*
  - Very expensive
  - Only practical for up to 3 parameters



35

Simplex Algorithm

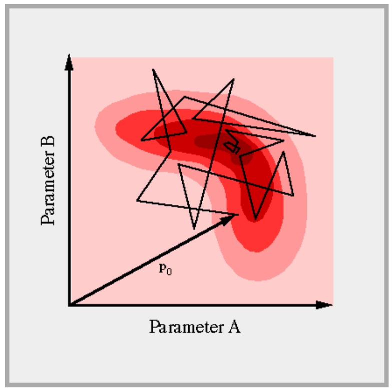
- *Principle*
  - Obtain downhill direction from *(n*-+1)-dimensional simplex. Move on by reflection, expansion, and contraction of simplex.
- *Advantage*
  - No derivatives
- *Disadvantage*
  - Inefficient
  - No formal error analysis



36

Simulated Annealing

- *Principle*
  - Perform random steps in parameter step. Accept uphill steps with a certain, decreasing probability.
- *Advantage*
  - No derivatives
  - Can get out of local minima
- *Disadvantage*
  - Inefficient
  - No formal error analysis



37

Gradient-Based Methods

- *Principle*
  - Perform step along gradient direction.
- *Advantage*
  - Robust for sufficiently small step sizes
  - There are efficient algorithms for calculating gradients
- *Disadvantage*
  - Inefficient stepping close to minimum
- *Examples*
  - Steepest descent
  - Quasi-Newton methods
  - Conjugate gradient methods

38

Second-Order Methods

- *Principle*
  - Evaluate second derivative of objective function or approximation thereof.
- *Advantage*
  - Quadratic convergence rate
- *Disadvantage*
  - Requires second derivatives
  - Not always robust
- *Examples*
  - Newton
  - Gauss-Newton
  - Levenberg-Marquardt

39

Gauss-Newton Method

- Linearize model:

$$\mathbf{z} = \mathbf{z}_0 + \mathbf{J}\Delta\mathbf{p}$$

- Substitute into objective function:

$$S = (\mathbf{z}^* - \mathbf{z}_0 \mathbf{J}\Delta\mathbf{p})^T \mathbf{C}_{zz}^{-1} (\mathbf{z}^* - \mathbf{z}_0 - \mathbf{J}\Delta\mathbf{p})$$

- Set derivative of objective function to zero:

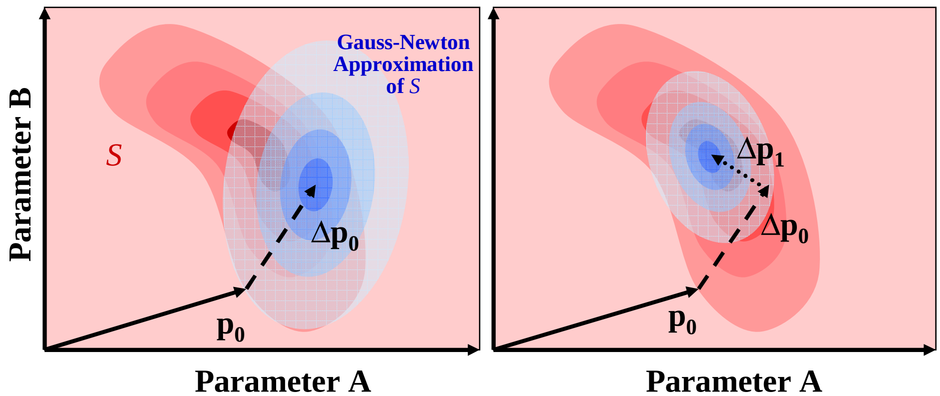
$$\frac{\partial S}{\partial \mathbf{p}} = 0$$

- Obtain solution Δp:

$$\Delta\mathbf{p} = \left(\mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{J}\right)^{-1} \mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{r}$$

40

Gauss-Newton Method



41

Gauss-Newton Method

- Gauss-Newton identifies the minimum in a *single iteration* if:
  - Model is *linear*
  - Quadratic objective function
- Quadratic convergence rate for weakly nonlinear models
- Too large steps for strongly nonlinear models

42

Levenberg-Marquardt Method

- Modification of Gauss-Newton method for strongly nonlinear models.
- Mixes gradient and Gauss-Newton method:
  - Performs robust steps along gradient far away from the minimum.
  - Performs efficient Gauss-Newton steps near the minimum.
- Automatically adjust relative weight between gradient and Gauss-Newton strategy.

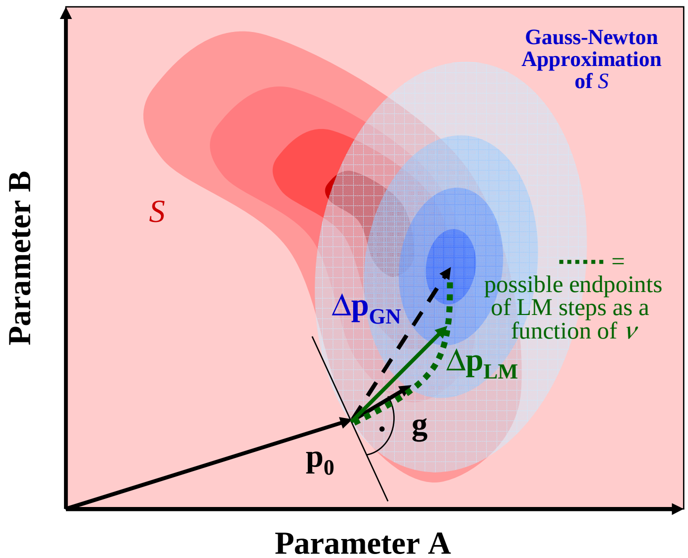
Levenberg-Marquardt Method

- Add diagonal matrix D to linear approximation of Hessian:

$$\Delta \mathbf{p} = \left(\mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{J} + \lambda \mathbf{D}\right)^{-1} \mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{r}$$

- If step is successful, move toward Gauss-Newton strategy (reduce λ by ν)
- If step unsuccessful, move toward gradient strategy (increase λ by ν).

Levenberg-Marquardt Method

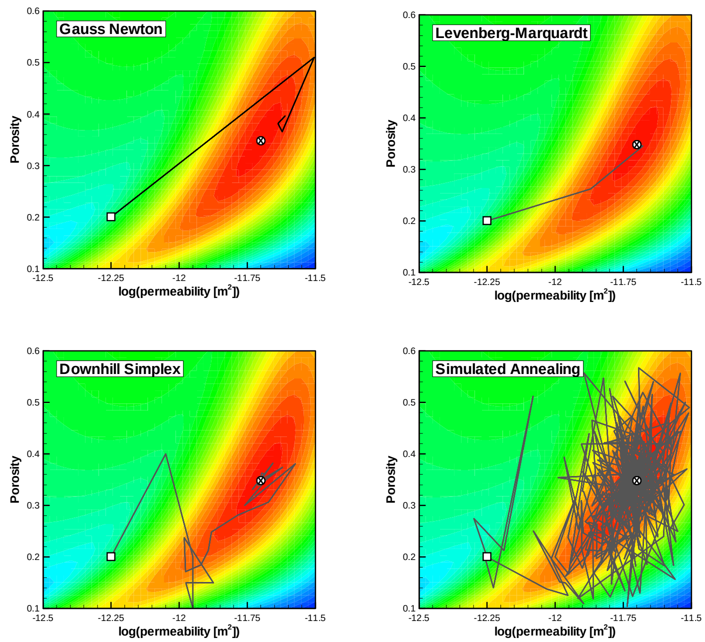


45

Levenberg-Marquardt Method

- The Levenberg-Marquardt modification of the Gauss-Newton algorithm allows for efficient minimization of objective functions from strongly nonlinear models.
- If stepping is successful, large Gauss-Newton steps are taken.
- Otherwise, small, robust steps along steepest-descent direction are taken.

46



47

Minimization Algorithm: Questions

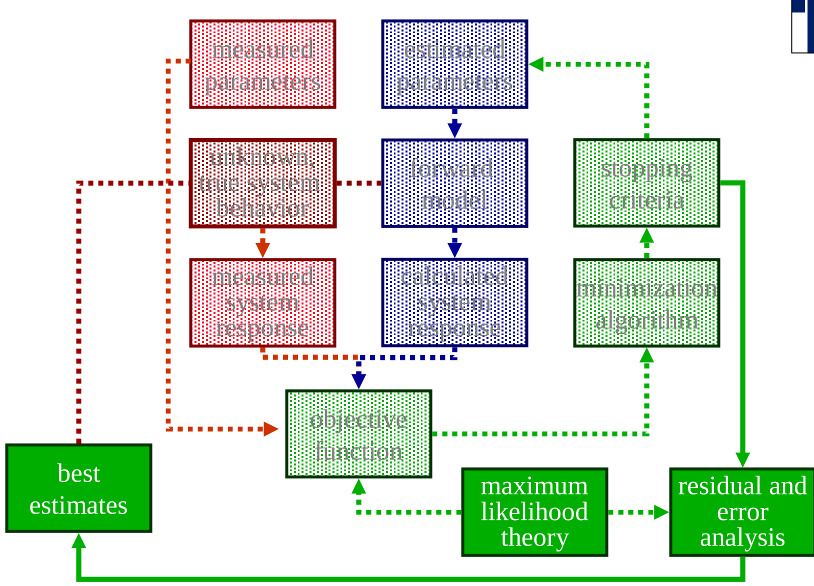
- Purpose of minimization algorithm?
- List methods and their advantages and limitations.
- Key criteria for selecting the minimization method?
- Sketch:
  - Objective function for two parameters, nonlinear model
  - Select starting point (initial guess of parameter vector)
  - Draw Gauss-Newton approximation
  - Perform two Gauss-Newton steps
- Principle of Levenberg-Marquardt method?

48

Error Analysis

- Estimated Error Variance s_0^2
- Fisher Model Test
- Covariance Matrix of Estimated Parameters C_{pp}
- Parameter Correlations
- Uncertainty Propagation Analysis

49



50

Why Error Analysis?

- Parameter estimates may be worthless if:
  - Model is unlikely match to the data
    goodness-of-fit, Fisher Model Test
  - Estimates are biased by systematic error or outliers in the data
    residual analysis
  - Estimation uncertainty is large
    C_{pp}, *correlation coefficients*
  - Solution is *non-unique* or *unstable*

51

Estimated Error Variance s_0^2

- The estimated error variance or *a posteriori* error variance is a measure of *goodness-of-fit*, representing the *mean weighted residual*:

$$s_0^2 = \frac{\mathbf{r}^T \mathbf{C}_{zz}^{-1} \mathbf{r}}{m-n}$$

- If *a priori* assumptions about residuals (expressed through matrix C_{zz}) were reasonable, the ratio is close to 1.

52

Fisher Model Test

- The ratio s_0^2 / σ_0^2 is a random variable following an *F*-distribution: $F_{m-n,\infty,1-\alpha}$

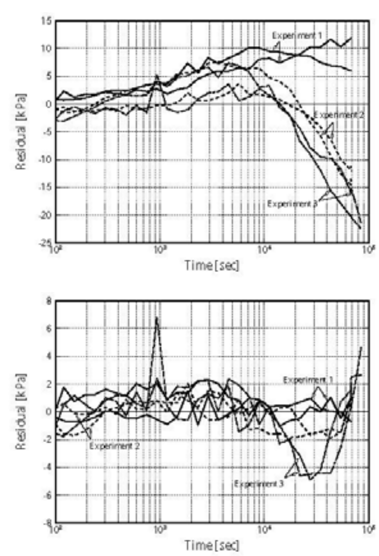
$F_{m-n,\infty,1-\alpha} < s_0^2 / \sigma_0^2$	**: *Error in functional or stochastic model***
$1 \le s_0^2 / \sigma_0^2 \le F_{m-n,\infty,1-\alpha}$	**: Model test passed**
$s_0^2 / \sigma_0^2 \le 1$	**: Error in stochastic model**

- Fisher Model Test only meaningful if reliable stochastic model available

53

Residual Analysis

- Residuals should be *random*.
- Trends and patterns in the residuals indicate systematic errors.
- Try to *remove systematic errors* by:
  - Refining the functional model.
  - Correcting data.
- Top figure indicates systematic errors caused by gas leakage.
- Bottom figure shows random residuals after gas leakage is explicitly modeled.



54

Covariance Matrix of Estimated Parameters $\mathbf{C}_{pp}$

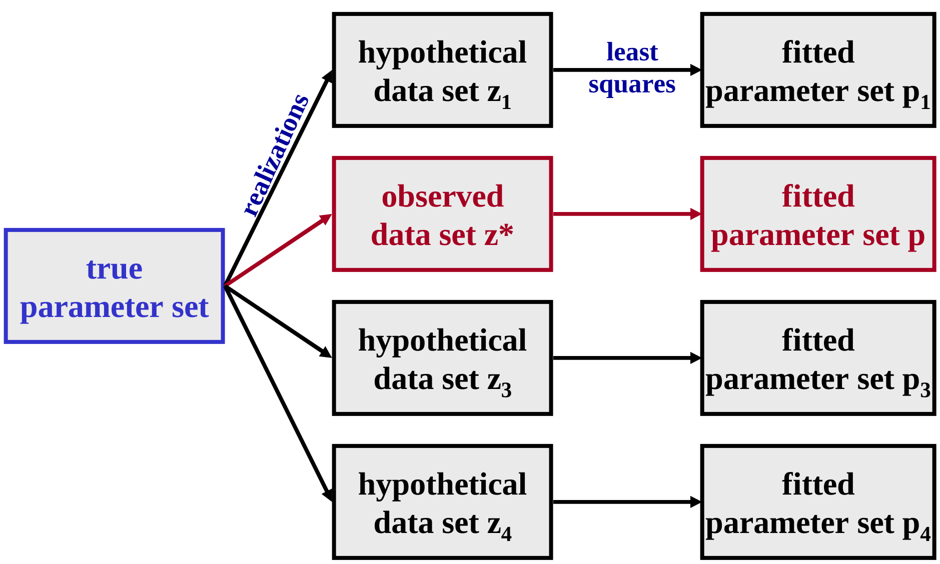
- The covariance matrix C_{pp} is an estimate of the uncertainty of the estimated parameters:

$$\mathbf{C}_{pp} = s_0^2\left(\mathbf{J}^T\mathbf{C}_{zz}^{-1}\mathbf{J}\right)^{-1}$$

- C_{pp} is an approximation of the actual parameter uncertainty; it is based on a normality and linearity assumption.

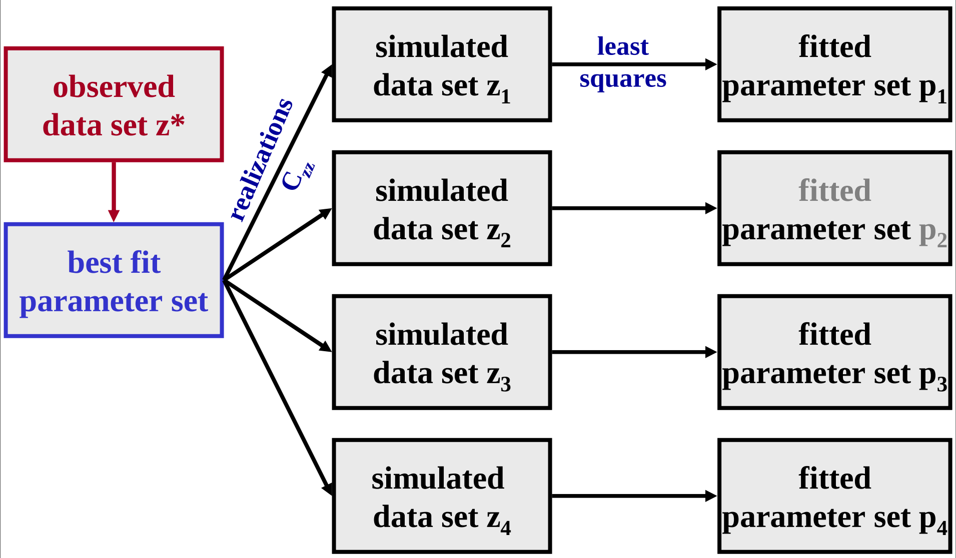
55

Statistical Universe of Data Sets



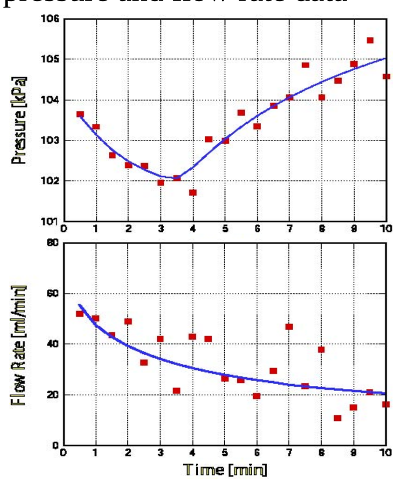
56

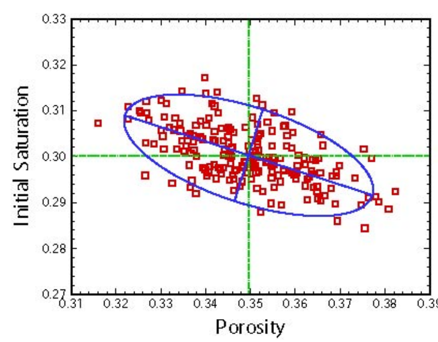
Simulating Statistical Universe of Data Sets



Simulating Estimation Uncertainty

200 realizations of hypothetical pressure and flow rate data ⇨ 200 best estimates of porosity and initial gas saturation

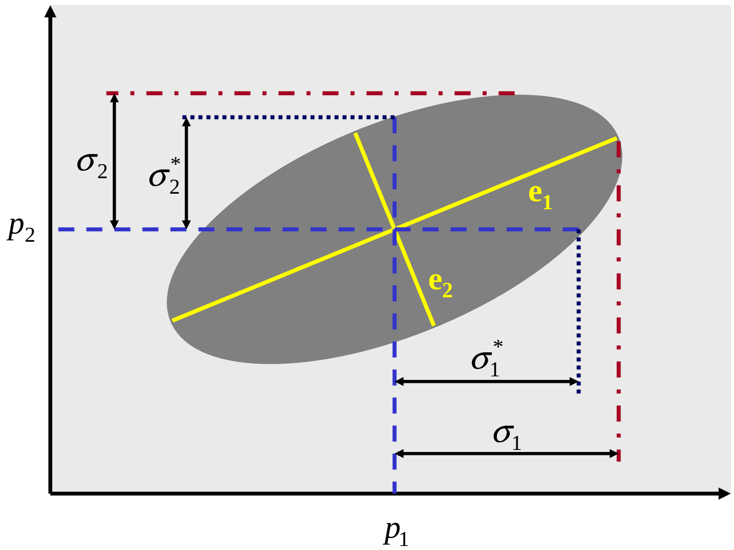




C_{pp} gives reasonable estimate of uncertainty region

58

C_{pp} Visualized



59

Covariance Matrix – Correlation Matrix

- Diagonal elements of C_{pp} contains variances σ_{ii}^2 of estimated parameters.
- Off-diagonal elements are covariances c_{ij} between pairs of parameters. They can be "normalized" to yield *correlation coefficients*:

$$-1 < r_{ij} = \frac{c_{ij}}{\sqrt{\sigma_{ii}^2 \cdot \sigma_{jj}^2}} < 1$$

- C_{pp} is directly proportional to goodness-of-fit (s_0^2).
- C_{pp} is inversely proportional to sensitivity matrix (J).

60

Estimation Uncertainty

- Decreases with improvement of fit
  - *Use good data and good model*
- Decreases with increasing sensitivity
  - *Use sensitive data*
- Decreases with decreasing correlations
  - *Use data that allow for independent determination of each parameter*
  - *Avoid overparameterization*
- Design tests accordingly!

61

Overparameterization

- A match can *always* be improved by adding more parameters to p.
- Adding new parameters increases correlations and thus *increases estimation uncertainty*.
- Check C_{pp} for large variances, correlation coefficients close to –1 or 1, and large condition numbers.
- Add parameters only if the fit can be *significantly improved* without introducing strong parameter correlations.
- *Avoid overparameterization*!

62

Correlation Coefficients

- A correlation coefficient of *zero* indicates that the two parameters can be estimated *independently*.
- A correlation coefficient of *–1 or 1* indicates *non-uniqueness*.
- A negative correlation coefficient indicates that a statistically similar match can be obtained by increasing one parameter and decreasing the other.
- If correlations exist, the uncertainty in one parameter affects the uncertainty in the other parameters.
- Design experiment as to minimize correlations.

63

Confidence Region

- The probability that the true parameter set lies within the ellipsoidal confidence region region represented by C_{pp} is $(1-\alpha)$.
- The true confidence region is bounded by the contour line of the objective function at level:

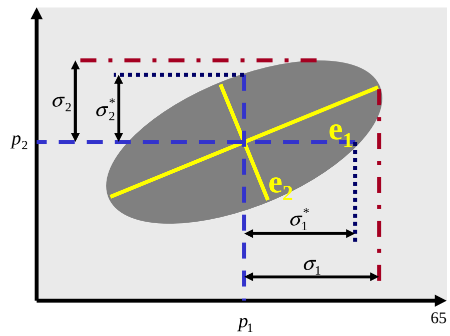
$$S_{\min} + s_0^2 \cdot n \cdot F_{m-n,\infty,1-\alpha}$$

- The confidence region in`creases with decreasing α.

64

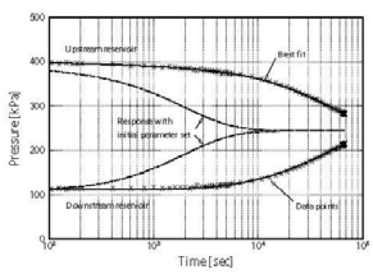
Overall Correlation

- The conditional standard deviation σ^* is the estimation uncertainty assuming that all other parameters are perfectly known.
- C_{pp} holds the marginal standard deviations σ.
- The ratio σ/σ^* is a measure of overall correlation.
- The ratio σ/σ^* should be close to 1

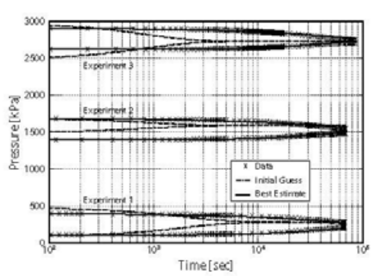


65

Reducing Correlations: Example



	log(k)	log(b)	ϕ
log(k)	1.67	-0.99	-0.87
log(b)	-0.99	2.16	0.87
ϕ	-0.87	0.87	0.003



	log(k)	log(b)	ϕ
log(k)	1E-4	-0.52	-0.12
log(b)	-0.52	4e-4	-0.02
Φ	-0.12	-0.02	0.01

66

Confidence Region

- The contours of the objective function visualize:
  - Confidence region
  - Correlation structure
  - Appropriateness of linearity assumption
  - Appropriateness of normality assumption
  - Well-posedness of inverse problem

67

Uncertainty Propagation Analysis

- Calculate prediction uncertainty as a result of parameter uncertainty.
- *Linear analysis* (First-Order Second-Moment)
  - Fast (*n*+1 forward runs)
  - Easy to report (mean and covariance matrix)
  - Based on linearity and normality assumption
- *Monte Carlo simulations*
  - Expensive (many forward runs)
  - Difficult to repost
  - Full distribution
  - No distributional assumptions

68

Linear Error Propagation Analysis (First-Order-Second-Moment)

- *Assumptions*
  - Change in model prediction Δz can be approximated by a linear function of the parameter changes Δp.
  - Δp is (log-)normally distributed.

$$\mathbf{C}_{\hat{z}\hat{z}} = \mathbf{J}\mathbf{C}_{pp}\mathbf{J}^T$$

- Error band is *symmetric*, representing (log-)normally distributed prediction errors.
- May assign certain probability to *unphysical* system behavior.

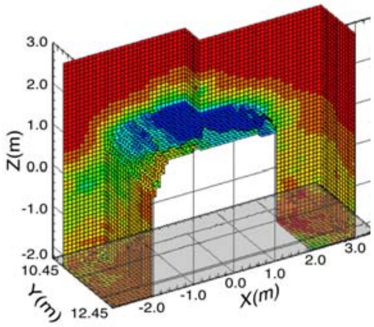
69

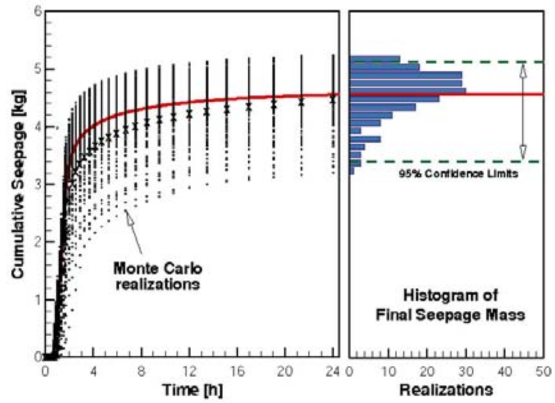
Monte Carlo Simulations

- Run many simulations with randomly selected parameter combinations drawn from the given probability density function.
- Provides *full distribution* of prediction uncertainty (histogram), which can be analyzed statistically.
- *Nonlinearities* are automatically taken into account.
- Results are always *physically reasonable*.

70

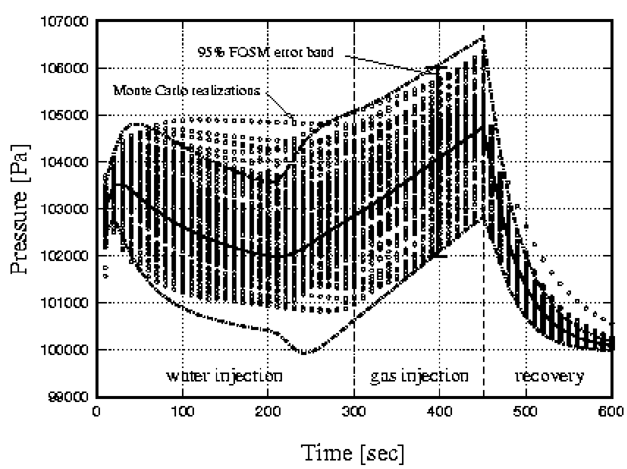
Monte Carlo Simulations





71

Comparison FOSM-MC



72

Error Analysis: Questions

- The *a posteriori* error variance s_0^2 turns out to be significantly greater than the *a priori* error variance σ_0^2.
  - What does “significantly” mean?
  - What does that result indicate?
- Discuss $\mathbf{C}_{pp} = s_0^2 \left(\mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{J} \right)^{-1}$
- Under which conditions is C_{pp} a good approximation of the actual confidence region?
- How can you reduce estimation uncertainty.

73

Error Analysis: Questions

- $C_{zz} = \sigma_0^2 V_{zz}$ is the *a priori* observation covariance matrix.

  If all elements of C_{zz} were multiplied by a factor of 4, how would this affect:
  - The value of the objective function *S*?
  - The estimated parameter set p?
  - The estimated error variance s_0^2?
  - The uncertainty of the estimated parameters C_{pp}?
  - The outcome of the Fisher Model Test?

74

Error Analysis: Questions

- Discuss "overparameterization".
- Describe the main differences between FOSM and Monte Carlo simulations.
- Assume you have to estimate the (hopefully small) probability that the TCE concentration at a drinking water well does not exceed a certain level.
  - Which uncertainty propagation analysis method would you choose?
  - Justify your choice.
  - Describe the procedure.

75

Final Remarks

- Identify and focus on parameters that most strongly affect the model predictions.
- *Design* an experiment that allows you to determine these parameters with sufficiently small estimation uncertainty.
- Use inverse modeling to estimate *model-related, process-specific, scale-dependent* parameters:

Model	+	Parameter	=	Prediction
correct	+	correct	=	correct
correct	+	wrong	=	wrong
wrong	+	correct	=	wrong
wrong	+	wrong	=	? (most likely wrong)
wrong	+	fitted	=	best we can get

76