Stochastic modelling and uncertainty

Jesus Carrera
Institute for Earth Sciences (IJA)
Higher council for Scientific Research (CSIC)
Barcelona, Spain
jcarrera@ija.csic.es
Contents

- GW Modelling and uncertainty evaluation:
  - Is it needed?,
  - Why?,
  - What for?
- Basic Concepts
  - Modelling procedure
  - Sources of uncertainty
  - Evaluation of uncertainty
- Discussion
  - Examples
- Conclusion:
  - Conceptual model: surprise
Traditional use of models

1. Understanding the past
2. Evaluating the present
3. Assessing the future state of aquifers

Semi-quantitative answers are sufficient, but

**Good qualitative answers require being very quantitative**
Modeling = Accounting

Lateral exchange, $f_{ij}$

Pumping, $Q_i$

Recharge, $r_i$

Storage var. $\Delta S_i$
Modelling: future needs

- A model is the (water or solute mass) accounting system for water bodies
- A well managed company needs a reliable accounting system. What about aquifers?
- If not, technical hidrogeology will continue to be a minor economic activity, despite of the importance of true hydrogeology.

But models need to be realistic, i.e., quantitatively accurate and reliable
Can models be accurate?

• Unknown parameters, extent and B.C.’s
• **Spatial Variability ➔ STOCHASTICS**
• Unknown actions. Pumping history is (one of) the best guarded secrets of any country!
• But, (long?) records of heads, and concentrations, and environmental isotopes, and well logs, and geophysics, and geology, and,.....
• Need to ensure consistency
1.- REALITY

How can this be achieved?

The trick:

We do not model reality but our view of it

... and then modify this view so as to fit actual observations.

4.- CALIBRATION

2.- CONCEPTUALIZATION

3.- DISCRETIZATION

\[ q(x) = q(P) \]

\[ Ah = b \]
Modelling Procedure

Conceptualization:
- Process Identification
- Model Structure Identification

CLASSICAL HIDROGEOLOGY

Science/knowledge

Site specific data

Discretization
Calibration
Error Analysis

Model selection
New experiment

OPERATION !!!

Prediction
Uncertainty Analysis

Multiple computer “realizations” are simulated using a range of input values for uncertain parameters.

Ensemble of realizations yields probability distribution for “performance metric.”

Stochastic Inputs
Multiple Computer Simulations (Flow & Transport Model)
Distribution of Results (Multiple Simulations)

Lanzhou, June, 2007
Uncertainty evaluation procedure

Transmissivity

Spatially variable

Other parameters (recharge, porosity, ...)

Deterministic, but uncertain,

Model Simulation

Water deficit
Capture from river
Travel time

Uncertainty in output is caused by uncertainty in input
Uncertainty evaluation procedure

Monte Carlo Method

Transmissivity

Other parameters (recharge, porosity, ...)

Deterministic, but uncertain,

Model Simulation

Model results

Water deficit, Capture from river, Travel time

Uncertainty is evaluated by repeated simulations with varying inputs according to their uncertainty
Issues regarding uncertainty evaluation

- Sources of uncertainty
- How to simulate inputs?
- How to quantify inputs uncertainty?
- How to condition on measurements?
- How to account for correlation?
- Which outputs should one look at?
- How many simulations?
- What about model and scenario uncertainty?
- CPU time
Sources of uncertainty

- Parameter uncertainty and variability
  - Feasible to quantify both its value and its effects on predictions

- Conceptual model uncertainty
  - Process: Feasible?
  - Structure: Feasible?

- Scenario uncertainty
  - ???
Probabilistic Performance Assessment Process

Formalization so as to ensure that all uncertainties are accounted for

Lanzhou, June, 2007
Assigning probabilities to conceptual models

- Given that one wishes to predict \( L \)
- That one has conjectured \( N_m \) models \( M_i \)
- That one has evaluated the pdf of the prediction for every model:

\[
 f(L / M_i) = \int f(L / p_i, M_i) \, \pi(p_i / M_i) \, dp_i
\]

- The total pdf is given by

\[
 f(L) = \sum_{i=1}^{N_m} f(L / M_i) \, \pi(M_i)
\]
How to evaluate $P(M_i)$?

1) Prescribe it: (e.g., equally probable)
   Even after data?

2) Estimate $P(M_i/data)$ (Kashyap, 1982; Carrera and Neuman, 1986; Medina and Carrera, 2004) from expected likelihood:

   $$P_i = P(M_i/data) = \int f(data/p, M_i) \delta(p_i/M_i) dp_i$$

   Strictly speaking:
   $$P(M_i/data) = \frac{\int f(data/p, M_i) \delta(p_i/M_i) dp_i}{\sum_{k=1}^{Nm} P(M_k/data)}$$

Which is feasible by linearization ($S=-2\ln(P_i)$)
Example: uncertainty in structure

\[ P_i = \alpha \exp(-S_i / 2) \]

All models lead to good fits, yet:

1000 orders of magnitude differences in posterior probability

It is wrong!
Example 2: Conceptual model = Surprise
Lanzhou, June, 2007

River Agrio

Aznalcollar dam
Characteristic pH after clean-up
Prior to barrier excavation

Based on:
Geophysics
41 EVS
Elect tomogr
Srf Mapping
Boreholes (27)
Trenches (10)
Hydraulic tests (3)
Prior to barrier excavation
After barrier
Revised Conceptual (Geological) model
The barrier (PRB) misses the aquifer!!!
Conclusions

- Uncertainty evaluation important
- Parameter uncertainty can be evaluated (just repeat simulations and examine outputs)
- But, Conceptual model = Surprise (Bredehoeft dixit). Always largest source of uncertainty
  - If I knew the painting before hand, I would not paint it (Picasso dixit)
- Bayesian methods not yet mature for evaluating posterior model probability
- Do not believe in full evaluation of uncertainty
Stochastics and uncertainty

- Two different concepts
- Uncertainty = ignorance
  - Random variables, probabilities, pdf’s
  - Statistics
- In many natural phenomena, ignorance is coupled to **spatial variability**
  - Random functions, variograms
  - Geostatistics, stochastics