

**QUANTIFYING UNCERTAINTY IN
GROUNDWATER MODELS**

**Stephen J. Cohen, Hydrogeologist
Uranium Recovery Licensing Branch
U.S. Nuclear Regulatory Commission**



U.S.NRC Protecting People and the Environment

WHY AM I TALKING ABOUT THIS?

TO DEMONSTRATE THAT UNCERTAINTY IN ANY TYPE OF MODEL COULD BE QUANTIFIED AND QUANTIFYING UNCERTAINTY CAN BE USEFUL TO BOTH THE REGULATORS AND THE REGULATED.

CONVEY AN UNDERSTANDING OF ONE METHOD OF QUANTIFYING MODEL UNCERTAINTY.

U.S.NRC Protecting People and the Environment

DISCUSSION TOPICS

- REASONS FOR QUANTIFYING MODEL UNCERTAINTY
- DESCRIPTION OF THE MAXIMUM LIKELIHOOD BAYESIAN MODEL AVERAGING (MLBMA) METHODOLOGY
- RESULTS OF MLBMA
- EXAMPLE OF MLBMA

U.S.NRC Protecting People and the Environment

WHY QUANTIFY MODEL UNCERTAINTY

- A THEORY HAS ONLY THE ALTERNATIVE OF BEING RIGHT OR WRONG. A MODEL HAS A THIRD POSSIBILITY; IT MAY BE RIGHT, BUT IRRELEVANT (Manfred Eigen, 1973).
- A MODEL BASED ON THE PRACTITIONER'S BEST GUESS IS INHERENTLY BIASED (Shlomo Neuman, 2006).
- QUANTIFYING UNCERTAINTY PROVIDES INFORMATION REGARDING THE DEGREE OF IRRELEVANCE, AND IT MINIMIZES BIAS.
- THIS DOES NOT REPRESENT A NEW REQUIREMENT!!!!



UNCERTAINTY ESTIMATION METHOD

- MAXIMUM LIKELIHOOD BAYESIAN MODEL AVERAGING (MLBMA).
- BASED ON PROCEDURES BY PNL, DRI-LAS VEGAS, UNIVERSITY OF ARIZONA (MEYER, P., YE, M., ROCKHOLD, M., CANTRELL, K., NEUMAN, S.)
- INCORPORATES MODEL, PARAMETER, AND SCENARIO UNCERTAINTY.
- CURRENT PAPERS AND GUIDANCE:
 - NUREG/CR – 6805
 - NUREG/CR – 6843
 - "ON EVALUATION OF RECHARGE MODEL UNCERTAINTY (MING YE, 2006)
 - "INFORMATION MATRIX" (JAY MYUNG AND DANIEL NAVARRO, 2004).



MLBMA METHOD

- BRIEF DISCUSSION OF BAYESIAN STATISTICS

BAYES' THEOREM

$$p(M_k | D) = \frac{p(D | M_k)p(M_k)}{\sum_{i=1}^K p(D | M_i)p(M_i)}$$

where:

$p(M_k | D)$ = posterior probability

$p(D | M_k)$ = likelihood of model M_k

$p(M_k)$ = prior probability

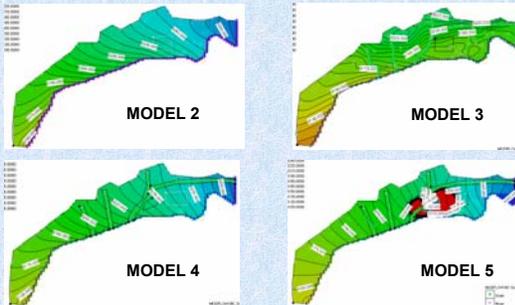


MODEL DEVELOPMENT

- MODELS DEVELOPED USING GROUNDWATER MODELING SYSTEM (GMS).
- MODEL 2: AVERAGE VALUES FOR HYDRAULIC CONDUCTIVITY (HC), RECHARGE, AND EVAPOTRANSPIRATION (ET).
- MODEL 3: AVERAGE VALUES FOR HC AND ET, ZONE VALUES FOR RECHARGE.
- MODEL 4: AVERAGE VALUE FOR HC, ZONE VALUES FOR RECHARGE AND ET.
- MODEL 5: SAME AS MODEL 4 WITH A GENERAL HEAD BOUNDARY, RECHARGE, AND ET.



MODEL DEVELOPMENT (cont'd.)



MODEL DEVELOPMENT (cont'd.)

- BOUND MODEL SELECTION FROM THE SIMPLEST TO THE MOST REASONABLY COMPLEX.
- MODEL ABSTRACTION – PROCESS OF MAKING COMPLICATED MODELS SIMPLE.
- SIMPLE MODELS MORE EASILY CALIBRATED AND TEND TO BE MORE ACCURATE.
- COMPLEXITY DOES NOT NECESSARILY TRANSLATE TO MORE ACCURACY, BUT WILL CERTAINLY INCREASE RUN TIMES.
- MLBMA PENALIZES COMPLEX MODELS.



JOINT CALIBRATION (cont'd.)

- CALIBRATE MODELS TO FIND OPTIMAL PARAMETERS.
- USE PEST WITH LOG TRANSFORMATION FOR HYDRAULIC CONDUCTIVITY.
- AFTERWARDS, IN "PARAMETERS" DIALOG BOX, INSERT OPTIMAL PARAMETERS AND UNCHECK "LOG TRANSFORM".
- SET NOPTMAX TO -1. THIS WILL GIVE YOU THE MATRICES YOU NEED FOR RANKING COMPUTATIONS.



JOINT CALIBRATION (cont'd.)

NOPTMAX

LOG TRANSFORM



PRIOR PROBABILITY

- ESTIMATE BY MODELER OF THE CONFIDENCE HE/SHE HAS IN EACH MODEL (SUBJECTIVE).
- COMMON METHOD IS EXPERT ELICITATION
- PRIOR PROBABILITY BECOMES LESS IMPORTANT WITH MORE DATA

$$p(M_i | D) = \frac{p(D | M_i)p(M_i)}{\sum_{i=1}^n p(D | M_i)p(M_i)}$$

where:

$p(M_i | D)$ = posterior probability
 $p(D | M_i)$ = likelihood of model M_i

$p(M_i)$ = prior probability



PRIOR PROBABILITY (cont'd.)

PRIOR PROBABILITIES FOR THIS EXAMPLE

- MODEL 2 – 0.35
- MODEL 3 – 0.3
- MODEL 4 – 0.2
- MODEL 5 – 0.15



LIKELIHOOD

- USE DATA FROM JOINT CALIBRATION TO CALCULATE LIKELIHOOD
- NEED TO OBTAIN THE MAXIMUM LIKELIHOOD PARAMETER ESTIMATES TO CALCULATE NEGATIVE LOG LIKELIHOOD (NLL) AND FISHER INFORMATION MATRIX.
- DATA FOR NLL AND FISHER INFORMATION MATRIX IS FOUND IN THE .REC FILE THAT IS WRITTEN AFTER PEST RUNS.
- NLL = OBJECTIVE FUNCTION = WSSR



MODEL RANKING WITH KIC

- RANKING ACCOMPLISHED BY CALCULATING KASHYAP'S INFORMATION CRITERION (KIC)
- INFORMATION CRITERION PROVIDES AN INSIGHT INTO THE AMOUNT OF INFORMATION ONE CAN OBTAIN FROM A SET OF DATA.
- FISHER INFORMATION DEFINED AS THE COVARIANCE OF THE FIRST PARTIAL DERIVATIVES OF THE LOG-LIKELIHOOD.
- OTHER CRITERION AVAILABLE – AIC, BIC. KIC CONSIDERED BEST FOR GROUNDWATER APPLICATIONS.



PARAMETER UNCERTAINTY (cont'd.)

- FOR Δ (i.e. HEAD AT A WELL, CONCENTRATION AT A WELL) OF CONCERN, TABULATE RESULTS AND COMPUTE PROBABILITY DISTRIBUTION FUNCTION (PDF).
- COMPUTE MODEL EXPECTATION (MEAN OF PDF) AND VARIANCE.
- COMPUTE MODEL-AVERAGED PDF.
- THESE ARE YOUR FINAL ANSWERS.



PARAMETER UNCERTAINTY (cont'd.)

- Δ = HEAD AT A PARTICULAR WELL
- PDF COMPUTED USING HISTOGRAM ANALYSIS
- EXPECTATION AND VARIANCE COMPUTED FOR EACH MODEL
- MODEL EXPECTATION AND VARIANCE (PREDICTIVE UNCERTAINTY COMPUTED FOR OVERALL MODEL
- MODEL-AVERAGED PDF CALCULATED



PARAMETER UNCERTAINTY (cont'd.)

- PDF, EXPECTATION, AND VARIANCE COMPUTED FOR EACH MODEL

Bin	Frequency	Probability	Expectation	Variance
5200	4	0.08	416	3695.52
5400	39	0.78	4212	112.32
5600	7	0.14	784	4948.16
More	0			
Expectation			5412	
Variance				8656

EXPECTATION FOR MODEL 2

$$E[\Delta | D] = \sum_{i=1}^K E[\Delta | D, M_i] p(M_i | D)$$

TOTAL MODEL EXPECTATION

Posterior Mean	E	p(M _i D)	E ² p(M _i D)
Model 2	5412	0.999999995	5411.999974
Model 3	5400	4.72269E-09	2.55025E-05
Model 4	5636	4.21149E-12	2.3736E-08
Model 5	5832	2.28318E-17	1.33155E-13
Sum - Total E			5412.0



