Official Transcript of Proceedings

NUCLEAR REGULATORY COMMISSION

Title:	Advisory Committee on Nuclear Waste 150th Meeting
Docket Number:	(not applicable)
Location:	Rockville, Maryland
Date:	Thursday, May 27, 2004
Work Order No.:	NRC-1499

Pages 1-66

NEAL R. GROSS AND CO., INC.
Court Reporters and Transcribers
1323 Rhode Island Avenue, N.W.
Washington, D.C. 20005
(202) 234-4433

UNITED STATES OF AMERICA

NUCLEAR REGULATORY COMMISSION

+ + + + +

ADVISORY COMMITTEE ON NUCLEAR WASTE

150TH MEETING

+ + + +

THURSDAY,

MAY 27, 2004

+ + + +

The meeting commenced at 8:30 a.m. in Room T2B3, NRC Headquarters, Two White Flint North, Rockville, Maryland, B. John Garrick, Chairman, presiding.

PRESENT:

B. JOHN GARRICK	ACNW Chairman
MICHAEL T. RYAN	ACNW Vice Chairman
GEORGE M. HORNBERGER	ACNW Member
RUTH F. WEINER	ACNW Member

ACNW STAFF PRESENT:

HOWARD J. LARSON	Special Assistant, ACNW
ALLEN CROFF	ACNW Invited Expert
NEIL M. COLEMAN	ACNW Staff
RICHARD K. MAJOR	ACNW Staff

PRESENTERS:

THOMAS J. NICHOLSON	USNRC/NRR
SHLOMO P. NEUMAN	University of Arizona
PHILIP D. MEYER	Pacific Northwest National Laboratory
MING YE	Pacific Northwest National Laboratory

C-O-N-T-E-N-T-S

P-R-O-C-E-E-D-I-N-G-S

8:30 a.m.

CHAIRMAN GARRICK: On the record. The meeting will come to order. This is the third day of the 150th meeting of the Advisory Committee on Nuclear Waste. My name is John Garrick, Chairman of the ACNW. The other members of the committee are Mike Ryan, Vice Chair, George Hornberger and Ruth Weiner. Also present is our consultant Allen Croff.

Today the Committee will be briefed by the NRC staff and its consultants on a proposed strategy for the treatment of uncertainties in hydrologic models: conceptual model and parameter uncertainty. Secondly, we'll continue our discussion of proposed Committee letter reports. Neil Coleman is the designated federal official for today's session.

The meeting is being conducted in accordance with the provisions of the Federal Advisory Committee Act. The Committee hasn't received any written comments or requests for time to make oral statement from members of the public regarding today's session. But should anyone wish to address the Committee, please make your wishes known to one of the Committee staff.

If you do participate, it is requested

that the speakers use one of the microphones, identify themselves, and speak with sufficient clarity and volume so that it can be readily heard. Today our lead member of the Committee on the topic is George Hornberger. I'm going to ask him to carry forward.

MEMBER HORNBERGER: Thanks, John. Welcome, everybody. Today we finally get to talk about something exciting.

(Laughter.)

CHAIRMAN GARRICK: It took us until the third day to get there.

MEMBER HORNBERGER: We have three presentations today. The Office of Research has been supporting work on the important topic of how to deal with uncertainty in hydrological and hydrogeologic models including how one deals with differences or uncertainties in conceptual models.

So we have three presenters this morning; Tom Nicholson of the staff here, Phil Meyer from PNNL, and Shlomo Neuman from the University of Arizona. I think without further ado, we'll launch in. Tom, I understand you are going to be first.

MR. NICHOLSON: Good morning. Thank you very much for the introduction, George. I'd like to introduce Phil Meyer to my left who will be talking after me about the unified methodology that has been developed. Shlomo Neuman will get into some of the theoretical aspects of it and some of the testing of the methodology using the Apache Leap data.

At the table back there is Mark Thaggard. He is the Section Leader in Performance Assessment in the Decommissioning Area. He is our customer.

MEMBER HORNBERGER: He pays the bills.

MR. NICHOLSON: He pays the bills. When we do good and he acknowledges, then that makes our management feel very good. I just would like to briefly introduce the topic to you. I'll discuss the uncertainty issues, the research of Jack, the tasks, applications. Then I'll summarize very quickly. There's some information sources, the NUREGs that have been produced.

Uncertainties in the sources we think are a very integral part of performance assessments. We think in order to have full documentation that uncertainty has to be addressed. In the past, a lot of it was just looking at parameter uncertainty for us conception of what other people referred as structural uncertainty is an extremely important part of this.

There are a variety of sources of hydrogeologic uncertainty. The first one, which is probably the one that we focus on the most, is the incomplete knowledge of the system being analyzed. The incomplete knowledge is often how you interpret, how you do model extraction, understand how the system should be characterized and eventually modeled.

So the conceptualization is extremely important. We also may get uncertainties due to measurement errors and characterizing the system's features, events, and processes and, of course, the natural variability of the system, spatial properties, and the transient external stresses, for instance, infiltration.

Finally, we also would like to look at uncertainties that arise from the disparity between the sampling scale, the monitoring scale, and the simulation relative to the actual dimension of these features, events, and processes which may effect radionuclonic transport. Next please.

As I mentioned very briefly, it's this need to look at alternative representation system that is one of the key issues in the methodology. Shlomo Neuman, in a previous contract with this, has developed a very good report, NUREG/6805, which talks about a strategy for identifying and creating these alternative representations of hydrogeologic system.

Also, the methodologies produce a very rigorous and systematic approach to identifying and quantifying these very sources of uncertainties which are mentioned. So what we did was, many years ago we first briefed you people on the work that Shlomo was doing on conception model uncertainty and how to represent and develop model extractions of the system of interest.

Phil Meyer and his colleagues at PNNL developed a separate methodology on parameter uncertainty. We have asked them, and what they are reporting on today is this unified methodology in which they are bringing together the conception model uncertainty with the parameters. We've asked them to also look at the scenario uncertainty.

Phil and Shlomo will talk about the scenario uncertainty. We're focusing right now on hydrogeologic scenarios, for instance, irrigation strategies, ground water pumping, flooding, things of that nature. Next slide please.

Well, what are our research objectives? Our most important research objective is to develop the technical bases for the licensing staff so when they review performance assessments, they will have knowledge of and tools to assess uncertainty. We also want this detailed methodology that is evolving to be able to identify and compare alternative conceptual flow and transport models.

We want to apply this methodology to a variety of test cases. Phil will get into some discussion. It's already been tested from a feasibility standpoint on the Apache Leap database. But now, they want to apply it to some larger scale problems analogous to decommissioning.

Then finally, another extremely important objective is to educate the staff. Tomorrow Shlomo, Phil and Ming Ye, in the audience, will be educating the NRC staff on their methodology. We'll fully explore with them how to develop and create alternative conception models, how to look at parameter uncertainty and the theoretical underpinnings of it.

This view graph is just simply to let you know that one of the things that we're most concerned about is structured media. There's a variety of ways of representing the database and phenomena, especially in the unsaturated zone, and that's what this is focusing on. We can look at the flow and later transport as it moves through course and fractured media. The question is, is it the matrix or is it the fractures that are controlling?

This is a view graph I would like to look at because it a reality check. Too often, we simplify models to the point where we don't look at the tremendous complexities involved in near surface and deeper process such as infiltration, development of perch water systems, the role of certain units either to become perching units or they may actually have fractures in them, such as the clastic dike, that allows water to migrate vertically. Then of course, there are other things such as wells themselves to be avenues for down home contamination.

So the research tasks, what are they? We have six of them. The first one has been accomplished. They have developed, and you should have copies of NUREG/CR-6843 which couples the conceptual model with the parameter uncertainty methodology. They are now incorporating scenario uncertainty into the methodology.

They are developing a test plane which Phil will discuss and test it on the 300 area database at the Hanford site, document the test case. As I said before, it isn't just one technology transfer. There's multiple ones in which they will come into the NRC headquarters and actually educate the staff on all the details of their methodology.

What are the applications? Well, the application is to apply their rigorous and systematic methodology to real test cases to formulate a set of plausible alternative conception models supported by field data. That's very important, supported by available field data, then to calibrate each one of these models to address parameter uncertainty and to estimate the model probability, and then finally to compute a weighted average of the model predictions with each model's results weighted by that model's probability.

In summary then, the research is to understand the various sources of uncertainty, to develop this systematic and rigorous methodology focusing on hydrogeologic flow and transport, formulate and compare alternative conception flow and transport models, and then to test the robustness and completeness of the methodology, and provide a technical basis for the staff. The last view graph is the three documents I have mentioned. So I would like to turn it over now to Phil. Phil, if you would walk through your view graphs with the gentlemen.

MR. MEYER: Sure. I'm going to go through this pretty quickly. There's a little bit of overlap between my slides and Tom's. I'm just going to hit the key points that I wanted to raise on those.

First off, I wanted to acknowledge not only Ming, who has been instrumental in this work, but also some other folks that have been involved; Mark Rockhold and Kirk Cantrell both at the lab who work as geochemists. Next slide please.

So from the perspective of the NRC staff dose assessments, the key issue for me is, where does the uncertainty come in? The approach that the NRC uses is a risk-informed, performance-based decision. The risk is assessed by evaluating uncertainty dose predictions. So that's where the uncertainty issues actually come in.

You typically have predictions that are made over a long period of time. There's complex processes involved. Therefore, the predictions of dose based upon that type of analysis are going to be uncertain. For our work, we are concentrating on the pathways only involving hydrologic transport.

Tom already went through the sources of hydrogeologic uncertainty that we're looking at. The key point here that I want to raise is that the uncertainty has the result that at a typical site there will be plausible alternative representations of the system and uncertainty about future behavior of the system. These alternative representations cannot always be resolved to a single representation that is the only one justified by the data.

So in terms of the project, our goal is to try to have an analysis of uncertainty for these problems that is somewhat comprehensive in the sense that it incorporates the parametric uncertainty, uncertainty about the conceptual models or the structural aspects of the representation, and also the scenario uncertainty where scenarios are conditioned. I'm going to talk next about each one of these just very briefly to raise a few key points.

So this is a picture taken from the near surface Hanford site by John Selker. It just illustrates the type of issues that can result in parameter uncertainty when looking at hydrogeology. Tom had a conceptual model slide from the Hanford site of tank waste leaks and potential transport and the various mechanisms that might be involved there.

I don't have that slide here, but that slide basically had things very homogenous. There were a few layers. There was Hanford gravels, Hanford sands which in that slide covered a fairly large area. Then there was a Caleche layer down below the tanks.

Well, this picture covers only a couple meters. But you can see the kind of variability that's at the Hanford site. This would be the variability that's within the Hanford sand unit that's in that picture that Tom showed. So you have physical and hydraulic properties here that are varying on the scale of just a few centimeters and the actual magnitudes are carrying over several orders of magnitude.

In addition, when you try to represent this, there's a limited number of samples that you can obtain from the site. Therefore, you can't actually discern this kind of variability from your sampling necessarily. And then there's scale differences between the scale of the measurements that you are taking and the actual representation within a model of the parameters of the site.

So our approach to the application of data and parameter estimation follows this little diagram. On the lower left, there's what we refer to as prior parameter values or prior parameter distributions which are based on generic or local information sources. That progresses and if you have site specific information, you can use that information to update, in a Bayesian sense, you could update those parameter values or distributions thereby reducing your parameter uncertainty.

In the upper right, if you have observations of the system behavior that you can apply to the calibration of the parameters, then you go ahead and do that using an inverse model and thereby reduce your parameter uncertainty even further. So there's a couple of points here. One is that the methodology that we want to apply needs to be able to incorporate systematically at any level parameter uncertainty. I guess that's my key point.

The other thing I wanted to point out here is that ultimately where you would like to be is up in the upper right where you are calibrating your parameters. That requires monitoring data. I know the NRC is sponsoring research on long-term monitoring.

The data that comes from such long-term monitoring would naturally fall into our methodology at the calibration point where as you collect more data, you can continue to refine not only your models but your parameter values. I'll discuss also in our methodology the probability of a model would get refined or updated in the same manner.

So that was parameter uncertainty. In

terms of conceptual model uncertainty, our perspective is as follows. Taking the site data and other information available, you can often formulate a number of conceptual models about the site and then also implement those potentially in different ways.

So the final conceptual mathematical model that you end up with, you may not be able to arrive at a unique representation of the system. That's represented here where at the bottom, there's three conceptual mathematical models that can be used to represent the site. Each one of them may be valid. That is, each one of them may be able to represent the data at the site to some degree. You may not be able to eliminate them all based upon the available data.

So in terms of evaluating conceptual model uncertainty, this is just a very brief summary of that. Shlomo is going to talk about this in more detail in terms of both the background and also application. But the basic idea is to postulate a set of plausible alternative conceptual models that are supported by the available data, then assign a prior probability to each alternative model where that prior probability represents your degree of belief and the suitability of that model for the site, and then estimate posterior model probability using observed behavior through process of calibrating each model and using the information from that calibration, and then compute the predictions with each model and combine the results using model probabilities as weights.

So this perspective doesn't try to lead you to a single model. In fact, the example that Shlomo is going to discuss, we demonstrate that if you use just a single model as opposed to a number of models, each of which is valid, that you will not have the best solution. That is, using multiple models and combining them in this way can lead you to better prediction, more reliable predictions.

There is a figure. This is entirely what Shlomo is going to be talking about. The Maximum Likelihood Bayesian Model Averaging is the name of this process. It's described in NUREG/CR-6843 and also in a Water Resources Research paper that just came out.

There is a flow chart that we put together in the NUREG that is in your notes. I'm not going to discuss that flow chart too much. But it summarizes the process of combined estimation of conceptual model and parameter uncertainty. Yes, it looks like this. (Indicating.) In the Water Resources Research paper and also in the NUREG, there is an application of this method that basically goes through the entire process.

I just want to briefly talk about scenario uncertainty because that is part of the issue too. Scenario uncertainty is a bit different than conceptual model uncertainty in the following sense. Similarly, we can postulate a set of alternative scenarios, a set of alternative future representations for the site in terms of things like Tom mentioned; irrigation, hydrologic events like flooding, stuff like that.

You can postulate these set of alternatives. In the same way that you can assign a prior probability to models, you can assign a prior probability to each scenario. That is, your degree of belief in the likelihood of that scenario occurring. Then there is a similar process to this. I'm not going to go into any detail.

But if you are comfortable with applying probabilities to scenarios, then you can incorporate that in a manner very similar to this flow chart just as an outer loop with this flow chart on the inside of that loop. If you are not comfortable with assigning prior probability scenarios, then you're stuck with something less than a formal assessment of scenario uncertainty because it's fundamentally different.

With the model probability, you can evaluate the probability of a model in the posterior sense from the system observations that you have. But you can't necessarily do that with scenario uncertainty. So if you are not comfortable with applying probabilities to scenarios, then you are limited to something like a sensitivity analysis for scenario uncertainty.

So in terms of evaluating method, the application that Shlomo is going to talk about is geostatistical modeling of air permeability in fractured rock. So in that case, the alternative models are geostatistic models of air permiability. That example is a complete application of the Maximum Likelihood Bayesian Model Averaging method.

It demonstrated the superiority of the model average result over the use of individual models. As I mentioned, that has just been published in Water Resources Research also. The other application that we're currently working on is uranium transport in the subsurface at the Hanford Site 300 area. I'm going to just briefly go through a few of the details about that application.

In the 300 area, there is a lot of process associated with the activities in the Hanford site that went on there. In disposing of some of their waste, they used liquid discharges to ponds and trenches. That waste had uranium in it which is now in the groundwater.

The site is outlined here in red on the surface there. That's a representation of the surface topography. The dark blue is the Columbia River. So the site is just a few hundred meters from the Columbia which makes it of some concern. This is the Columbia River here. (Indicating.) This is basically the fence line. The operations went on in here. There's a disposal pond here. Then these are some disposal trenches. This distance here is two or three hundred meters.

This is a representation of the major geologic units as the Hanford site geologist represents them shown here. The next slide, there is a cut away view that illustrates the layering, the three dimensional nature, discontinuities in layers. These are some of the data points represented by these yellow lines representing wells at the site.

We are currently developing what we're calling a nominal model for this site which is a three dimensional unsaturated/saturated zone model in which we will try to incorporate as much detail as possible, as much detail as we're willing to consider at the site. Then our plan is to have some relatively simpler models that we will actually apply the uncertainty methodology to.

This is a plan view of the nominal model, the most complex model, the representation of the grid discritization that we're using. This is a three dimensional model. This shows the data points that we're using. The three sources of contamination are located there. Next please.

One of the issues at this site because it is so close to the river is that there is an influence of the river on the groundwater. The river goes up and down in response to the seasonal cycles and also in response to the way the dams on the river are operated. This is just a time line from 1944, the beginning of the operation of the site, up to the present time of our reconstruction of the river stage.

You can see that it varies over ten meters. It has in the past. There was a discontinuity in terms of the statistical representation of the river stage when the last dam, Mica, went in up on the river in Canada. So this is just to illustrate that this is not only a three dimensional problem but it's also the transient issues. Transients of the transport is potentially an issue, and we will be representing that in our modeling.

So the uncertainty assessment is being applied here to a set of alternative models that are simplified from our nominal model. We are representing those models using the GMS, ground water modeling system, framework, MODFLOW, and MT3D to the greatest extent possible. The reason for doing that is, there are some NRC staff that have experience in GMS, and NRC is sponsoring work with the GMS folks.

The alternative representations that we will be using include homogeneous versus heterogeneous hydraulic parameterization and the steady state versus transient boundary conditions. Also, the chemistry at the site is somewhat complex. There's a lot of research going on now at the Hanford site related to that issue.

We will be representing a portion of the current chemical knowledge about the site in terms of the uniformity or non-uniformity of the adsorption model that's applied. Adsorption of the uranium is very sensitive to the total carbonate and solution concentration which varies with the river water and the ground water.

So just from a philosophical point of view, I wanted to finish up with a couple of thoughts. The value of uncertainty estimates is limited. So we have a process here, a methodology that we describe as comprehensive in some sense. But at the same time, it's important to recognize that the uncertainty estimates that are going to come out of any uncertainty analysis are lower bounds.

This is a quote from someone that we all know. "As we know, there are no knowns. There are things we know we know. We also know there are known unknowns. That is to say, we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know."

And I added a fourth category, those things that are unknown knowns, things we think we know but in fact we don't know. As I mentioned, the consequence of this is that any uncertainty estimates have to be looked at as lower bounds. But that doesn't mean that because of that you should do nothing. It's better to approach the problem from the point of view of trying to look at the uncertainty the best you can than to throw up your hands and say, "Well, it's so uncertain I can't do anything."

I will just end here with a quote from a personal philosophical inspiration of mine. "I know a lot of things, but I don't know a lot of other things. You have to stand for something or you are going to fall for anything." Thanks.

MR. NEUMAN: Good morning. As Tom and Phil have mentioned, I'm going to give you a brief summary of a paper that has just appeared on the Water Resources Research Journal website. The paper is right here. Essentially, it deals with this issue of conceptual parameter uncertainty assessment using a methodology that we have developed in the context of a previous NRC project which we are now trying to extent to the area of scenario uncertainty and applications in the context of the current PNNL projects of which I am involved.

The motivation for looking at conceptual model uncertainty stems from the recognition that environmental systems, in particular hydrogeologic subsurface systems, are open and complex. As such, if you were given a set of characterization monitoring data, there can be multiple interpretations of these data essentially leading to a system of possible conceptualizations and mathematical models.

It is common in hydrology to rely on a single conceptual model. We think that this may lead to what is known in statistics as Type 1 Model Error which arises from the rejection by omission of valid alternatives. I have been participating in many critiques and litigations associated with hydrogeologic systems. Almost always the focus is on the conceptual model underlying whatever mathematical model is used to support any given hydrogeologic calculations.

Type 2 Model Errors arise when one adopts by not rejecting an invalid model. This is especially critical if there is just one single model, as is always the case. This can be devastating from the standpoint of a person's reputation if he presents a conceptual model in the context of a scientific conference. In the context of litigation, it may cost millions of dollars. In the case of environmental issues, of course, it can lead to environmental damage.

Models are based on a single conceptual framework, therefore, underestimate uncertainty by undersampling the valid model space. This is the Type 1 Error. And they may introduce statistical bias by relying on an invalid model which is the Type 2 Error. And these uncertainty and bias may be significant.

So in order to address these issues, we

have, in the context of our previous NRC project with the University of Arizona, developed a comprehensive strategy of hydrogeologic modeling with special emphasis on uncertainty assessment. The strategy is summarized in NUREG/CR-6805 published by myself and Peter Virenga in 2003.

The basic idea there is to account for uncertainties due to three major sources. The most important one that we were focusing on, because it was novel and there were no known ways for addressing it, was the conceptual model uncertainty which of course is manifested in the mathematical model which summarizes the underlying concept. We will refer to this as structural model uncertainty.

Model parameter uncertainty has been handled in the past. We have well developed techniques to handle it. But of course, the question is how to combine this with the conceptual model uncertainty aspect. It is relatively easy. The literature is full of techniques that allow one to account for uncertainty enforcing terms. In the case of hydrogeology, that would be source terms, boundary conditions, initial conditions and so on.

It is very possible that certain scenarios could be embedded within this level of uncertainty but not all of it perhaps. One key element of this more comprehensive strategy is this Maximum Likelihood Bayesian Model Averaging concept. So what is Bayesian Model Averaging? It is a technique developed by statisticians, especially by the Statistical School of the University of Washington in Seattle. But others have been developing it.

It started perhaps ten years ago or so appearing in our literature and has been summarized in a very nice tutorial by Hoeting in 1999. There have been some additions to that since then where the idea is that one considers a set M, call it, of possible conceptual models translated into mathematical models.

So we have a set, M1 through MK, of mathematical models, each one based on a different conceptual framework. Suppose we want to predict a quantity Delta, which in the context of hydrogeology could be hydrolic head, velocity, flux of the contaminant, whatever it is that we want to predict. Of course, there can be multiple Deltas. But we'll be focusing on one of these.

So what we would like to know is the probability that this Delta is correct given the data. Or what is the probability or distributions of our predictions? In other words, what is the uncertainty of the predictions?

The idea here is that we would write this posterior distribution of the Delta posterior because it is based on observation of data D as a weighted sum over all the models that we have adopted for our analysis rather than relying on a single model where P Delta/MKD is the posterior distribution of Delta given by a single model and P MKD is a weight which represents the posterior probability of this model being a correct model.

All of these probabilities are implicitly conditioned not only on the data but on our choice of models. So everything is going to be relative to our choice of models. We do not believe that it is possible to assess predictive uncertainty in an absolute sense but only in a conditional sense given a certain set of models, given a certain set of data.

One can then easily come up with expressions for the prediction or posterior mean of Delta given as the ensemble average or the statistical average of the quantity we are trying to predict, Delta, given the data, which again is a weighted average of the predictions or ensemble averages given by individual models weighted by the posterior probability of each model.

And the variance can be expressed in a similar manner and can in fact be decomposed into two components; a variance associated with the predictions of a single model, again, weighted by the posterior probability of each model and a variance which arises from differences between the models, the between model variance, and again weighted in the same way. This has been shown by Draper and others in the statistical literature.

What is Maximum Likelihood BMA, to which we refer as MLBMA? BMA requires prior information about the parameters of the model. It also would entail for implementation a very large number of Monte Carlo rounds of each model. The idea behind BMA is to enhance the computational efficiency of BMA and also to eliminate this need to rely so heavily on prior information.

So the idea then is to approximate some of these probabilities using Maximum Likelihood estimates of the parameters. Theta hat would be a likelihood estimate of the parameter space. Theta K, K being the designation of a particular model. We have models running from M1 to MK.

In particular, what I have proposed in 2002 as part of this previous NRC project is to use a so-called model discrimination criterion developed by Kashyap, to which we refer as KIC, to estimate the posterior probabilities of the model MK/D. There are well-established techniques in hydrogeology and of course not only in hydrogeology, but my focus is on hydrogeology, to obtain these Maximum Likelihood estimates and calculate the Kashyap model discrimination criterion.

One can do it with or - and I want to stress that - without prior information about parameters. Very often in hydrology, we do not have reliable prior estimates of the parameters. We rely on monitored observation of the system to calibrate a model through inversion against those data and this way, estimated parameters.

The approach is valid for both deterministic and stochastic moment models of the subsurface or for that matter any other system. One can then use Monte Carlo or stochastic moment models to estimate the predictive uncertainty of Delta so to obtain an ensemble mean E of Delta given MK, the model, the estimates, Theta hat, and a given set of data, and the same with respect to the variance.

Both BMA and MLBMA include a system M of models. The question of course is, how should one choose these models? Of course, we will want to work with models which are physically most plausible. They appear to be qualitatively consistent a priori with the available knowledge and the data so that they form what is sometimes referred to as Occam's window.

Otherwise, there would be an infinite set of models that one could consider. So we have to limit ourselves to something that is practical. To the extent that these models are clearly distinct from each other, then it would make sense perhaps to assign prior probability to each model as being simply 1/K where K is the number of the models. Otherwise, there may be some questions about how to assign these prior probabilities.

This is an open question. How should these prior probabilities be determined? What impact will they have on the final result? What we believe is that the more one conditions the models on data, the less important it is what the prior probabilities will be because the posteriors will essentially overwhelm the priors. But nevertheless, it's an open issue that we need to address.

So the overall strategy then is to postulate alternative conceptual mathematical models, which in itself is a whole issue, assign prior probability to each model, another major issue, assign prior probabilities parameters of each model - and in MLBMA this is optional, in BMA this is the essence - obtain posterior parameter estimates for each model and an estimation covariance - this is critical - by statistically-based model calibration or inversion, calculate posterior probability for each model using the formula that we have just looked at, predict quantities of interests using each model, assess prediction and certainty, the distribution and the variance in the least, for each model using Monte Carlo or a stochastic moment method which does not require Monte Carlo and is therefore computationally potentially more efficient, weigh predictions and uncertainties by corresponding posterior model probabilities - this is the BMA concept - and sum these over all the models so that there is a weighted average prediction of both the quantity of interest and the uncertainty associated with it.

I'm very quickly going to go through our first application of this which was done primarily for demonstration and analysis purposes. It may not be directly relevant to NRC interests. But nevertheless, from a purely scientific standpoint, we think that it has provided us with a pretty good case study.

Some of you may remember the Apache Leap Research Site in Arizona which is unsaturated tuff. We have conducted a number of single hall and cross hall pressure interference tests at the site. You can see the boreholes there. Here, I'm going to talk about one meter scale packer tests which have provided us with over 180 measurements of air permeability in this fracture domain.

The question we are going to ask ourselves is, what is the best geostatistical model of spatial correlation to apply to these data? If you look at those data and plot a sample correlation representation in the form of a variogram between those data - these numbers by the way indicate how many pairs were available for each point on this correllogram, variogram type plot, lag distance is the distance between data - there is a variety of models that one can fit to this spatial correlation model.

We are going to, in particular, look at a fractal power model, models that treat the medium as homogeneous statistically. Those are the exponential in this model, and models which superimpose on this homogeneity a trend or a drift. Those are the first order and the second order polynomial drift models, altogether a number of models.

It is well known that if one tries to estimate jointly by Maximum Likelihood, both the variogram and the drift parameters in some of these models, which have most of them, one can obtain biased estimates. So we have come up with a two step procedure to avoid this bias. I will not go into those details because they are technical.

But just to give you a very quick idea, it is possible using a method called Universal Kriging coupled with a Maximum Likelihood parameter estimation scheme, to which we refer as the Adjoint State ML Cross Validation scheme, it is possible to estimate variogram parameters without estimating the drift parameters. Once we do this, to the extent that a model includes drift, we then estimate it by so-called generalized lead squares.

There is a table here showing our calculated posterior model probabilities for each one of these. Let's start from the top. You can see the various models designated Pow0. This is the power model. Exp0, this is the exponential correlation model with added drift. Sph0 is a spherical model without a drift. One indicates a linear drift and two indicates a quadratic drift. This is all in three dimensions.

The second row is important because it indicates the number of parameters associated with each one of these models. The third one is the negative log likelihood, a measure of model fit to the data that I was showing you. If one went strictly by the model fit, one would probably select the exponential two model which has the lowest value of NLL.

And yet, the model discrimination criterion KIC would select other models. We have, by the way, looked at various other model discrimination criteria. For those of you who are familiar with this concept, there are others called IKE and VIC and so on. We have tried them all. They do not give a consistent ranking of these models.

What is typically done in situations such as this is, people do the parameter estimation, look at these model discrimination criteria, and use them to select one model and discard all the others. It's very clear from this example that doing so is really without justification. First of all, these criteria are very close to each other. Second, their ranking is not entirely consistent.

So this is where we come in and say, "How about selecting several of these models and analyzing them jointly?" We do it twice. The first time we assign a probability p(Mk) to every one of these models equal to 1/7 because there are seven models. The second time, based on the calculated posterior model probabilities which essentially give zeros for three of these models, for Exp2, Sph0, and Sph2, the second time, we ignore those three saying they have very low probability.

One could also ignore the one with the very low probability of 0.51. But we keep this in the picture and redo this by assigning a 1/4 to each one of these four non-zero probability models and essentially get very similar results in this particular case. It's not clear that that's what's going to happen always.

We will run very quickly through some figures which show you two dimensional sections through a three dimensional volume over which we estimate log permeability and plot it for the various models on the top. At the bottom, we plot the corresponding estimation variance. If you look at these pictures, you will see that the models give very similar estimates of the parameters.

So if all you wanted was an estimate, you could use almost any one of these models and the differences wouldn't be large. Where the differences are really large is in the variance of the estimation, meaning at the bottom. So it's the bottom where you will see differences.

Let's go to those other two. So this is Exp1 and Sph1. We now have four models only that we have retained. We have eliminated three of those based on the posterior probabilities that you have seen. And now, BMA. So the posterior mean here is the weighted sum - we used a method called Kriging to do the estimation - of the Kriging estimates.

The posterior variance according to the formula I have shown you before is the weighted sum of the within-model and between-model variances and the weights, of course, as the posterior model probabilities. Again, the estimate is very, very similar but the variance now is different.

So the summary of this. The posterior probability is the weighted sum of the model probability. You can see that the heavy solid black line is a compromise between the various models. The variances are shown at the bottom. Again, it's a compromise between the variances of the various others. In this case, we are looking at the variance average over all the points or pixels within this three dimensional block of pixels.

More important is cross validation. In order to see how well each model individually can predict data and how well or poorly MLBMA will do, we look at six boreholes. We have data from six boreholes. So we ignore measurements of log permeability data from one borehole at a time, use the remaining data to estimate these, and then compare with the known values that have been measured in each one of these boreholes.

So we estimate the value of the parameters and model probabilities based on the remaining data, assess and compare the predictive capabilities of the models, and BMA. This just indicates the sensitivity to the data. We have compared this - we don't see the comparison here - sensitivity to other model discrimination criteria which would also be used in our context such as IKE and VIC and so on and come to a conclusion that is not new that the KIC discrimination criterion appears to be the most sensitive to data.

This is one major reason why we advocate using KIC because otherwise someone could use something else as well. More importantly, a measure of predictive capability is the so-called log score, the negative natural logarithm of the posterior probability of predicting DT or data that were ignored using a model Mk and the set of data DB which have been used or are being used for the purpose of the prediction.

So for a single model, it's just a minus log of the posterior probability of DD for a given model, and then for BMA we have summed them up weighted by the posterior probabilities of the model. You can see that BMA provides the least predictive log score meaning the highest probability that its predictions are correct in comparison to the individual model. The smaller, the less information is lost.

Another measure of predictive capability is the so-called predictive coverage where we generate by Monte Carlo simulation the whole range of possible results. We look at the 90 percent interval of the generated values and want to know to what extent the actual data lie in this in-depth prediction interval. Here you want, of course, the largest amount of data to lie in the predictive interval. Again, BMA covers a larger range than the other model. It's very close to the power model, but it's certainly very different from the other two models. The larger, the better the model's predictive capabilities.

So to summarize, we found that MLBMA provides a theoretical as well as a working framework for prediction under uncertainty which accounts jointly for model structure uncertainty, the conceptual framework, the nature of the mathematical equations that are going to the model, the parameters that go into this equation, and though we haven't really looked at it from a theoretical standpoint, we know that we can account for forcing terms, which again I want to suggest already embed at least a certain class of scenario ranges.

By changing the forcing terms, you essentially change the regime, the scenarios under which things happen and all of this in a manner which is consistent with everything that we know about the system and the available data. In this particular example, we have shown that MLBMA is superior to individual geostatistical models of data at DLRS. Thank you.

MEMBER HORNBERGER: Thank you very much. Is that it, Tom?

MR. NICHOLSON: That's it.

MEMBER HORNBERGER: Great. Very interesting. So I'm sure there are questions. I think we should give our Bayesian first crack here.

(Laughter.)

CHAIRMAN GARRICK: You've given us a little more than I can digest in 30 minutes. I find consistency in a lot of areas between what you are trying to do particularly with respect to modeling uncertainty which is the key one that we need to deal with in many respects and the way we have done it for a couple of decades in some of our large risk assessments.

But I do have a few issues of clarification. My biggest problem is trying to connect what you are doing with the way I have been practicing this business for a long time. Maybe I should start with that in a simplistic way. The way we build risk models and try to account for information uncertainty - we sometimes prefer to call it information uncertainty over parameter uncertainty - and modeling uncertainty is we kind of look at a risk assessment as a structured set of scenarios.

That right there brings us to a different interpretation of what is meant by a scenario. In the work that we've had a lot of experience with, what a scenario is is basically a pathway from some sort of an issue condition or initiating event to some consequence. Each pathway may result in a different consequence.

So what we do is we structure our scenarios usually in some sort of an event tree format such that we can clearly account for the intervening events between the initiating event or the initial condition and the endstate or the consequence. As a result of that and all the combinations and permutations you get, you get a lot of endstates depending on what intervenes with the scenario as it progresses.

So one very convenient structure has been to look upon a risk assessment as a set of scenarios. For each of these scenarios, we determine a probability of the scenario. That probability is based on, of course, all of the evidence. For the most part, the work has not been as accountable for modeling and conceptual uncertainty as it has been for information and parameter uncertainty.

Then when we get all the scenarios, we convolute those scenarios on the basis of reordering them in terms of increasing consequences and then cumulating them from the bottom into a family of complimentary cumulative distribution curves. Then we have a very nice display of not only the risk associated with each scenario but the total risk of the system that we're analyzing. That's the basic structure that we've done for two or three decades.

Now, one of the things here that's very different, of course, is what is meant by a scenario. Although, it may not be as much of a difference when I look into it more carefully than I'm able to do just on the basis of your presentation. But a couple of things that I have questions about are, when you talk about a scenario and calculating the uncertainty of a scenario, embedded in that analysis, of course, could conceivably be the so-called structural uncertainty and the modeling uncertainty so that that becomes a result that embraces both parameter uncertainty and our information uncertainty and modeling uncertainty.

So I don't look at that as a different kind of uncertainty as the methodology that you have been discussing about seems to kind of imply that this is a different uncertainty. That may be just because, as I said earlier, we're talking about a different definition of what we mean by scenario.

The other thing that you said that I'm having a little trouble wrestling with is - I guess it was said by Phil - that single conceptual model inevitably leads to an underestimate of uncertainty.

There I have a misunderstanding of what you mean because I have seen a number of conceptual models that led to an overstatement of uncertainty. It's just a lousy model, a conservative model.

One other thing I would suggest on things like this curve here of the posterior probability and the weighted sum of model probabilities where you show the results of the different models, I assume those results are mean values.

DR. NEUMAN: Yes.

CHAIRMAN GARRICK: It would be very informative to see the family of curves representing the uncertainty of each of those models.

DR. NEUMAN: We have both the mean and the values.

CHAIRMAN GARRICK: Right. But I mean if you were to plot between some reasonable bounds, say, of five percent and the 95 percent because that would communicate to you not only what the results are in terms of the central tendency parameters but how the model works with respect to the treatment of uncertainty.

DR. NEUMAN: Right. Well, actually the results that we have –- Let me start from the back and move backwards in addressing the issues that you have raised, each of which, I think, is very, very well taken. As far as presenting the results, we present those mean values which are the predicted values and then look at the variance of the estimation error or rather the prediction error. So I think we do have, as well as the distribution of the estimation errors, both the prediction and various measures of how good those predictions are.

The second point that you raised remind me please. What was that? Because now I'm confused. You raised three points. The single model. The idea of the single versus multiple models is that what we normally do, at least in hydrogeology, is adopt a single conceptual model on which we build a hydrogeological mathematical model for a site, for example, Yucca Mountain or whatever and then study uncertainty on the basis that the model is correct and the uncertainty results from our inability to evaluate exacting what the parameter values are. So essentially it's the parameter uncertainty that is normally being evaluated.

If you assume that the model is correct and only look at the uncertainty associated with the parameters, then you have undersampled the space of potential models because there may be other models that also are associated with their own parameter uncertainty and if you were to add those too fro multi-curves, you would most probably have a wider range of uncertainty to superimpose. That's the idea.

CHAIRMAN GARRICK: Yes, I see what you're saying.

DR. NEUMAN: As far as scenario uncertainty is concerned, the type of scenario that I mentioned, I specifically suggested it is a very limited definition of scenario, scenarios that result from forcing terms in a particular conceptual framework and parameterization framework. We fully recognize that uncertainty in the model itself or changes in fact in the system may represent a scenario and changes in the parameters may represent a scenario. But the focus of this particular proposal is to go way beyond that in the definition of scenarios. Here I would rather defer to Phil in filling in this information about what we will be meaning by scenarios.

DR. MEYER: Let me just first comment on your comment about the question about the single conceptual model and how you said you've seen cases where the uncertainty was grossly overestimated with the single model because it was a poor model.

CHAIRMAN GARRICK: Yes.

DR. MEYER: So our perspective on that issue is if you only have a single model and your model happens to be poor, then you're going to make a poor decision one way or the other. The advantage from the outset of trying to look at multiple models acknowledging that it's valuable to try to formulate alternative models from the get-go will lead you into the process doing so and then the quantitative methods that we describe can be used to assess the posterior probability of those models.

In the case if you go out and have three or four models, it depends upon your data, of course, but in a situation where one of those models is very poor, that is, it's poor with respect to representing observations that you have at the system, then that model, like was the case with some of the models that were considered in the Apache Leap example, ends up with a very small posterior model probability. So you eliminate those models from the analysis from any further consideration. That may be the case of what happened.

CHAIRMAN GARRICK: Right.

DR. MEYER: But Shlomo's point is accurate that if you're considering additional models from a common sense point of view, you can only be increasing the total uncertainty to consider between the models.

DR. NEUMAN: Just to add one point to that and that is if you just have a single model which is wrong and estimate --

CHAIRMAN GARRICK: But just picking up on the quotes that you showed, you don't always know that it's wrong.

DR. NEUMAN: Right. You don't know what is wrong, so the result of that is statistical bias.

CHAIRMAN GARRICK: Yes.

DR. NEUMAN: You may or may not know that your model is bias. Typically, you will not because you start from the premise that your model is correct.

CHAIRMAN GARRICK: Right.

DR. NEUMAN: And then you superimpose on this a probability distribution of an uncertainty evaluation or assessment which is purely based on uncertainty of parameters and may be the input functions, the forcing terms. So now you have a distribution about an incorrect mean.

CHAIRMAN GARRICK: Yes.

DR. NEUMAN: You take another model. The mean is going to be different and the distribution is going to be different.

CHAIRMAN GARRICK: Sure. Right.

DR. NEUMAN: So that was the idea.

CHAIRMAN GARRICK: One of the things that I was interested in also was your frequent comments about the difficulty of establishing prior probabilities. In practice, we haven't found that to be such a big issue as a Bayesian.

I never really quite understood why people who are somewhat anti-Bayesian say "It's okay except where do you get your priors?" Well, you get your priors from what you know. And then you proceed from there to try to infer from additional information through Bayesian methods what the impact is on that prior. As you said, in many instances, what was the prior distribution didn't matter much anyhow because the posterior information dominated the outcome. So in practice, it really hasn't been the issue that we often hear in people who are not extensive users of Bayesian methods or more specifically, people who are somewhat anti-Bayesian.

DR. MEYER: The reason we emphasize that point is because in our experience of presenting this stuff in the past including to an audience filled with experts that have had a lot of experience in the general area of uncertainty assessment that we've gotten a lot of questions about that and concern raised along the lines that you suggest. So that's why we've put the point in.

CHAIRMAN GARRICK: Yeah. Well this is fascinating work and very important work because I think we have a long ways to go to get a real handle on the contribution to uncertainty from the conceptual model from the modeling standpoint. And I have many more questions than I want to take the time to deal with now, but let me just encourage you to continue. You might search for a simplification of the methods in some areas.

DR. MEYER: Can I just make a comment about the description you gave of probabilistic risk versus risk assessment.

CHAIRMAN GARRICK: Right.

DR. MEYER: Very well developed particularly in a reactor safety area. I've thought about the differences and how to reconcile the terminology and the applications and I haven't really reached a determination.

CHAIRMAN GARRICK: Well, we're really having a problem with that in other nuclear materials, so I can appreciate that. Although I think that there's some real basic approaches and practices that are transferrable. Those are the ones, of course, we want to take advantage of as much as we can.

DR. MEYER: One of the issues that I see in that area, and someone from the NRC Staff can correct me, but in the reactor safety area, it seems like the regulations, the requirements at the end, the end state, has been probabilistically based for a long time. Whereas, in the nuclear waste area, the endpoint, the criterion, is not probabilistically based. It's a deterministic one.

CHAIRMAN GARRICK: Yeah, we're working on that. Mike?

VICE CHAIRMAN RYAN: I don't have any questions to that.

CHAIRMAN GARRICK: Ruth?

MEMBER WEINER: First of all, I want to agree with my colleague, Dr. Hornberger. This was an absolutely fascinating presentation and I want to thank you all very much. Now my questions are a lot more naive than Dr. Garrick's, so please excuse their naivete ahead of time. Dr. Neuman, what's the most valid counter-argument to your approach?

DR. NEUMAN: One of the counter-arguments that I have already received from colleagues is that it doesn't make sense to speak of numerous conceptual and mathematical models because in order to develop only one for a given site, it takes a tremendous amount of time, effort and money. So when it comes to major site based models of flow and transport and three dimensions over a given region and so on, the chances of people actually postulating more than one model and willing to actually work fully through the entire modeling process including uncertainty assessment with more than one model is not going to be practical.

My answer to that is it depends on how important this is to the project. If it is important, then the money should be found to be done. Typically what happens is that people get together in a room, argue out based on the available data and understanding of a given site, a given system, their various viewpoints and then one group that does the actual modeling will go and decide "Okay, based on what everybody else has said here is how we are going to conceptualize the site." But many viewpoints, then, remain unrepresented. So that's one counter-argument.

The other counter-argument is the anti-Bayesian argument that Dr. Garrick has pointed out and that is there are fundamental issues associated with what is the meaning of prior probabilities. How do you know that you have selected the correct set of models within that set M of so many models that you are working with?

My answer to the first one is very similar to what Dr. Garrick has said and that is I am hoping that I will be working in a situation where the data will eventually overwhelm my priors. So my priors represent in our view our understanding of the system. It's subjective. What is our current concept of how this system may operate? What the uncertainty associated with the various models is?

There is a valid question raised by statisticians about the possibility of including in a set of seven models three that are very similar to each other and in our example from the Apache Leap site for example, one could argue that the three exponential models essentially belong to one family and to spherical models or three spherical models belong to the same family. So maybe what one should do is dilute their prior probabilities.

We have played with that concept in our paper. I haven't shown you the results. The results are sensitive in our case to the priors, not to a great extent, but to a sufficient extent to raise some concern. The only way that I see that that can be addresses is sensitivity analysis of these kinds where you try different kinds of priors and then if you establish that you cannot distinguish between them, maybe use this model as a means of guiding future data collection so as to if you believe that it's worse, then reduce the uncertainty and the ambiguity associated with that.

Another possible weakness which I think is a strength on one hand, but weakness on the other is the maximum likelihood approximation because it's an approximation. If it's not done correctly, it may lead to statistical bias in the estimates. That's why we were concerned with that in our particular application.

We don't really have a full answer to the question "How good that approximation is as compared to the full BMA without the ML, maximum likelihood, approximation". So there are quite a number of open questions, I think. It's the first application of Disto hydrology (PH) and the first application of ML of this kind of ML application because they are related in statistics that I think need to be addressed.

MEMBER WEINER: Have you applied it to any systems other than the geohydrologic systems? I ask because intuitively having worked with single models, you always get a different model that's going to give a different answer. That's always true and intuitively you think "Well, if I use more than one, if I put more than one into this construct, I will get a better result." Have you applied this to any other system?

DR. NEUMAN: Not we, but statisticians have done so. Over the last decade, there have been a number of papers that this concept of BMA has become quite widespread among biogen statisticians. Typically they would apply this to much simpler systems than the ones that we deal with. We are hoping and, of course, we are developing this Hanford application which I think is going to be interesting. But, yes, there is in the literature a number of examples worked out by statisticians.

MEMBER WEINER: I have a lot more questions like Dr. Garrick, but I won't take up the time of the audience. Thank you very much by the way. I did have a question for Dr. Meyer. It's a really simple question. You mentioned that in your vertical drill-down at the Hanford site that properties varied by orders of magnitude. Is that true for adsorption onto the soils also? Does that also vary by orders of magnitude if you go down?

DR. MEYER: The laboratory data shows that the uranium adsorption at the Hanford site is very sensitive to ph and total carbonate in solution. That does vary over the range of possible values by orders of magnitude, but the belief is that the ph is buffered very quickly by the soil, so ph is really not that important at the site.

The total carbonate does vary because you have rainwater coming in and then it interacts with the solids. You have river water that is mixing at the river zone, but that variation in the KD value of the linear model in your adsorption model is like 1^{10} maybe.

MEMBER WEINER: I just wondered about that. I have one more that I cannot resist. Is the result from an invalid model always bad? Is there mathematical proof of that?

MR. NEUMAN: Huh. That's a very interesting question. Is it always bad? I guess not. It depends on how you use the model. It is possible very often to fit almost any model to a wide range of models to given data and clearly, not all of these models represent the system equally well. Because we deal with natural systems, not engineered systems or at least not fully engineered systems, we really don't know what the correct model is.

What we do know is that using the wrong model for long term prediction is an issue of extrapolation which is always dangerous. It's not always bad, but it is always dangerous. I have at least two examples. In the 1970s, the late Professor Raimi from Stanford and very well known petroleum engineer, developed a very, very simple model for pressure and temperature evolution in the Wairaqui geothermal field in New Zealand which they then used to predict these temperatures and variations in pressure within this system over several decades.

They discovered after about ten years that the predictions are completely off. The reason for that was that while they were developing the model and calibrating this simple model against existing data the Wairaqui system was dominated by hot water. But in ten years, the system flushed and developed into a two-phase water vapor system. At that point, of course, the governing equations were completely different.

The other example is some looks back by Conoco of the U.S. Geological Survey and others at how well did models of groundwater contaminant transport develop in the `70s and the `80s turn out to predict actual situations in aquifers at various sites. They found, of course, that with time the predictions started deviating as one should expect actually from what was actually found.

One reasons for that was that the forcing turn simply did not correspond to what they assumed when they developed the model. Another reason is at that time our modeling capacities were nothing as good as what they are today. But the third one is simply that the model themselves are not entirely reliable. I know that in the petroleum area where modeling is used continuously to plan the production modes and quantities of petroleum and gas from reservoirs they never use a model for more than just a few years without recalibrating it to the data as more data come in fully being cognizant of the fact that long term predictions are a problem.

MEMBER WEINER: Thank you.

MEMBER HORNBERGER: At the risk of exposing myself as being Phil's Class 4, *i.e.,* thinking I know something that I don't, it strikes me, if I understand correctly, your maximum likelihood approach that if you use the KIC or the information criterion to basically as a likelihood wait to evaluate your posterior probabilities for your model. Is that it?

DR. NEUMAN: No, not directly. It enters into a formula for the posterior model probability.

MEMBER HORNBERGER: Right.

DR. NEUMAN: The actual formula is in one of the appendices of the paper that you received.

MEMBER HORNBERGER: Okay. So loosely speaking, that's --

DR. NEUMAN: Right. Loosely speaking, right.

MEMBER HORNBERGER: I guess what I don't know and if this is a long answer, we'll just skip it. You intimidated that if one did Monte Carlo simulations that this posterior probability would automatically pop out. That's obscure to me.

DR. NEUMAN: What I was talking about actually is a comparison of -- Okay. I don't have that formula here, but it's in the paper. Do you have the paper?

MEMBER HORNBERGER: Not in front of me.

DR. NEUMAN: Okay. In BMA - I'm looking at the paper now at a formula which is not in the slides - equation 3 is an expression for the posterior

- of course, posterior - likelihood function of a given model and it is given as the integral of the likelihood function of a given model and its parameters. So you have an integral P of D/θ, the parameter, M, the models, multiplied by the probability of the parameters for a given model integrated over the parameter space.

So if you use this formula, you absolutely have to have a prior probability of the parameters for a given model and then you integrate that over the entire probability space meaning you have to generate, unless it's a very simple case which you could do it analytically, by Monte Carlo simulation a huge number of these. You fully rely on priors whereas in ML because of our experience in hydrology that we very often do not have good priors, but in fact, rely upon observations of actual system behavior to get at these, we kind of kill both birds at the same time, introduce this question of how good is the approximation, but nevertheless, skip this.

MEMBER HORNBERGER: Just one last question then. Shlomo, you started out by saying that you've been involved in critiques and litigation. So I have a question. If you were involved in a critique and somebody had - we'll use your Apache Leap example -

chosen a single model and let's say it was your EXP0 model.

DR. NEUMAN: Yes.

MEMBER HORNBERGER: So you go through all of yours and you say "Yes, your maximum likelihood Bayesian give you better examples than the cross-validation anyway." But if you're critiquing, would there be any reason that you would then say "Aha, you picked the wrong model" because you picked your EXP0 model?

SR. NEUMAN: Well, the one thing that would come out of this approach is a comparison of these models both in terms of the various discrimination criterion, particularly the KIC criterion. If there was a big difference between these, you would say "Aha, there's another model we just match better." But I think more telling would be the posterior probability and you saw that at least three of the models, in fact, four of the models, had zero or extremely low posterior probability.

Based on that, if you selected one of them, I would be able to tell you "Look, you have selected a model which has a very low probability of being the correct one given the existing data." You could still come back and argue "Well, maybe if I had other data that this model would turn out to be better" and you might be right and I wouldn't be able to argue against you.

MEMBER HORNBERGER: But if I had picked one of your models that had a posterior probability of 35 percent, would you say that I was wrong?

DR. NEUMAN: No, I would not say that you are wrong. I would say you had probably picked a model which is almost equally likely to my other models and you are fine. So, yes, absolutely.

MEMBER HORNBERGER: Questions from Staff? Mike?

MR. MAJOR: As the committee -- Well, let me back up or pitch this differently. Can the maximum likelihood approach be used to homogenize competing conceptual models? If a decision maker can't choose between competing conceptual models, can you use this approach to -- I think this goes to your slide 18 on BMA results.

DR. NEUMAN: Yes, I'm glad you asked this question because actually I've been asked the same question many times after making this presentation, apparently not making myself clear enough that the whole idea of this approach is to do precisely what you are suggesting rather than selecting the best among a selected model which is what people used to do.

MR. MAJOR: Right.

DR. NEUMAN: We look at these models, evaluate them in light of the data from an uncertainty aspect and then say "We can get rid of these, but we should keep the other four and produce a weighted average prediction with all four." So that's actually what we do which another way is to actually average out homogenize the predictions. Yes.

DR. MEYER: I'll just make a comment that your question is directly related to that one. Why just not pick one model? You have three there about equally weighted. In this particular application, the intermodel variability was relatively small. As someone pointed out, if you pick one of those three models, you're probably okay, but that might not be the case in some other situation.

You might have a very large intervariability and what that means is if you pick one model, you could get very different results than if you pick another model even though they have equal probability, the predictions, if they are significantly different. That means that an approach like this where you keep all those models is more valuable than in our application where the models really gave very close predictions. Does that make sense?

MEMBER HORNBERGER: Yeah, and I do understand that. My question wasn't quite that naive, but I also think that in the case you just outlined, you might get some arguments from people if you were in a litigation situation.

DR. MEYER: Yeah, and also the greater the variability between models the more incentive there is, in terms of their prediction, to go out and try to resolve those differences.

DR. NEUMAN: My point about the litigation was not that people look at alternative models, but that the easiest thing to do in a litigation situation is to attack the underlying concept, so you have this very elaborate model, but "Wait a second. Where do the assumptions come from?"

MEMBER HORNBERGER: Absolutely. Neil.

MR. COLEMAN: Most of the sites that NRC and EPA look at in licensing work, it's remediation work for when it's looking at different contaminants that have been introduced. Now those, in my experience, provide very useful traces for better understanding and differentiating conceptual models and the plausible ranges of parameters just as physicians use tecnicium-99M to understand processes in the human body. Does the documentation that you folks have developed give guidance on how best to use the early information that one has on patterns of contaminants at any site to early on narrow down the range of plausible conceptual models?

DR. NEUMAN: I don't think that we have done so specifically. If you look into the NUREG CR6/CR6805, the one that I reference here on the strategy, we have looked at and discussed and not precisely documented but discussed various ways in which hydrogeologic data of all kinds should enter into the process of constructing alternative conceptual models. One example in which some tracer data –- Well, actually, not tracer data. I'm trying to think if there were any.

The only example in which tracer data actually entered, strictly speaking, was the Frai au Gere (PH) example, which is an abandoned uranium mine in fractured granite in France where we have tracer data, but not in the context that you are mentioning where both hydraulic and tracer data entered into the comparison of two different models. I'll be discussing that tomorrow, but not specifically what you are saying.

MEMBER HORNBERGER: Okay. We're actually on a fairly tight time schedule this morning. We thank you very much and I'm going to turn the meeting back to the Chairman.

CHAIRMAN GARRICK: Thank you. My reading of the agenda says that this is pretty much the time when we can adjourn the meeting. I would like to ask the Committee to hang around for a little while because we might have some other things to talk about one on one outside of the agenda. So with that, I think we will adjourn. Off the record.

(Whereupon, the above-entitled matter was concluded at 10:07 a.m.)