# SENSITIVITY STUDIES WITH A PROBABILISTIC RADIONUCLIDE TRANSPORT MODEL FOR GEOLOGICAL DISPOSAL IN MEUSE/HAUTE-MARNE, FRANCE

Richard Codell (consultant), Sitakanta Mohanty, Stuart Stothoff
*Southwest Research Institute®, 6220 Culebra Road, San Antonio, Texas, USA 78228; e-mail: smohanty@swri.org*

Gregory Mathieu, Marc Bourgeois, Delphine Pellegrini
*Institut de Radioprotection et de Sûreté Nucléaire, PRP-DGE/SEDRAN/BERIS, BP 17, 92262 Fontenay-aux-Roses Cedex, France; Email: delphine.pellegrini@irsn.fr*

*This paper examines selected sensitivity analysis approaches, compares these methods using results from MC-MELODIE, a probabilistic hydrologic flow and contaminant transport code, and highlights experience gained from using these methods in the context of deep geologic disposal of intermediate long-lived and high-level wastes in France. The paper compares three sampling methods — Factorial Designs (FD), Box-Behnken Designs (BBD), and Latin Hypercube Designs (LHD). A simplified but realistic surrogate for the actual performance model was developed from a regression based on 100-realizations of MC-MELODIE ($R^2$=0.89). Results, based only on the coefficient of determination, show that LHD appears to perform best, followed by BBD sampling and full-factorial two-level sampling. For experimental designs limited to few samples (such as million year MC-MELODIE runs), the three-level sampling designs allow planning of further experiments to take advantage of the additional information on the curvature of the response surface. This property of three-level designs, while desirable, is not necessarily important for characterizing the sensitivity of parameters for a natural system. Given the results of this computational experiment, and our experience with using sampling methods for repository performance assessments, we see no particular advantage for using sampling techniques other than LHD.*

## I. INTRODUCTION

The French Institute for Radiological Protection and Nuclear Safety (IRSN) has carried out field-scale experimental and modeling studies to support its technical review of the national radioactive waste management agency's (Andra) proposed waste disposal facility project for intermediate-level long-lived wastes (IL-LLW) and high level wastes (HLW) in a deep geological formation at the Meuse/Haute-Marne (MHM) site. The disposal facility is proposed for the argillaceous Callovo-Oxfordian (COx) formation[1]. Low-permeability clayey materials planned for the entrance of the disposal cells, in the drifts and in the shafts or ramps at the level of the top of the COx formation, are intended to minimize water flow through engineered components and to limit/delay the release and transport of radionuclides to the shafts exits. IRSN's review of Andra's Safety Case for the disposal of radioactive waste in the COx formation is being complemented with a stochastic approach using MC-MELODIE, a probabilistic version of IRSN's deterministic MELODIE hydrologic flow and contaminant transport code representing the details of the hydrologic system by ~11 million node points and a time scale of 1 million years. IRSN is investigating various methods to further analyze the impact of uncertainty on the performance level of the disposal facility.

The study of uncertainty impacts and the associated sensitivity studies for the French repository program have been carried out by Deman *et al*.[2]. This paper highlights analyses, which involve computational experiments with a detailed computer model to assess the influence of uncertainties in hydro-dispersive parameters on modeling mass transport within the MHM aquifer system. They carried out tens of flow and mass transport simulations using the finite element simulator GroundWater[3], and also developed approximation functions (e.g., polynomial and spline functions) to serve as a substitute for the detailed numerical model for performing risk analysis to support decision making.

They concluded that classical Monte Carlo (MC) schemes would not be appropriate because of the long run times and high computational expense. They instead based their analysis on the Response Surface Method (RSM) to conduct sensitivity analysis on selected variables and rank them on the basis of their relative contribution to the total variance of the model outputs. RSM is used to (i) establish an approximate relationship between the sampled parameters and the performance model responses, (ii) determine the significance of the sampled parameters, and (iii) find the optimum values of the sampled parameters that result in a maximum or minimum response[4]. While any experimental design has the same general goals, RSM often is concerned with higher-order relationships, such as a quadratic, between the system responses and the parameters[5].

The objective of this paper is to examine sensitivity analysis approaches used in Deman *et al*[2] with other methods, using results from IRSN's detailed probabilistic flow and transport model MC-MELODIE, and highlight experience gained from using these methods. The paper first introduces existing sensitivity analysis methods, then compares, based on IRSN's MC-MELODIE model, three sampling methods — Full-Factorial Two-Level Designs (FD)[6], Box-Behnken Designs (BBD)[7], and Latin Hypercube Designs (LHD)[8]. The paper primarily focuses on the authors' experience with these methodologies.

## II. SENSITIVITY ANALYSIS METHODS

### II.A. Factorial Designs

The term "Factorial Design" is usually defined as a plan for experiments for which each parameter of the system is defined at two or more discrete levels[6]. A "full-factorial design" is one in which experiments are performed on all possible combinations of factors. Such a design requires $K^n$ experiments, where $K$ is the number of discrete levels (usually 2 or 3) and $n$ is the number of parameters defining the system. FDs usually sample over a range of each parameter at fixed intervals (e.g., the 5th and 95th percentile for a 2-interval design or adding the 50th percentile for a 3-interval design). Figure 1 shows the samples for a 3-parameter, two-level FD, with 8 points at the vertices of a cube. By sampling all parameters in a system in this manner, it often is possible to determine unambiguously the effects of the variations in a parameter and all combinations of parameters. Two effects result from an experiment: (1) the "main effect" is the sensitivity of the result to a parameter itself; and (2) "interactions" are the sensitivity of a result to combinations of parameters. The number of experiments often is limited by computational expense and time, and FDs are frequently cost prohibitive. For example, a system having only 10 components (i.e., parameters) with 2 levels would require 1,024 experiments, or 59,049 experiments with 3 levels.

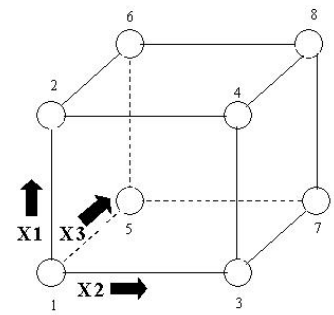


Fig. 1. Full-factorial for three parameters and two levels

#### *II.A.1. Fractional Factorial Designs (FFD)*

These designs require only a fraction of the number of samples that would be required in a full-factorial design. However, the reduction in the number of experiments can come at the expense of ambiguous sensitivity results. The main problem with FFD is that the apparent response for a particular parameter or a combination of parameters (the "main effect" and "interactions", respectively) can be caused by the interaction of two or more other parameters. This phenomenon is called "aliasing" or "confounding."[6] FFDs are characterized by their "resolution" (i.e., the more sparse the sampling, the lower the ability to distinguish the main effect or interactions from the confounding effects of other parameters.) For example, a sparse "resolution III" design estimates the main effects, but main effects are confounded by two-factor or higher interactions (i.e., we cannot be sure whether we are measuring the actual sensitivity of a parameter, or we are measuring the interaction of two other parameters.) A resolution V design, which has much higher sampling density, has less ambiguity and therefore estimates main effects without confounding from less than four-factor interactions. Two-factor interactions are not confounded by other two-factor interactions, but three-factor interactions may be confounded by other three-factor interactions[9].

### II.B. Three-Level Designs

Where responses to parameter changes are not linear, RSM often relies on sampling schemes that can efficiently extract higher-order information with relatively few experiments. The $3^k$ factorial design is based on three samples for each of the parameters. However, FDs for three or more levels are often prohibitive because of the large number of samples required. Ordinary FFDs for three levels often are not efficient ways to fit quadratic effects from experiments because they usually lack sufficient degrees of freedom for fitting all terms[5]. Two FFD methods have been developed explicitly to improve the efficiency of sampling for three levels: The Central Composite Design (CCD) and the BBD[7,9]. Only the latter will be discussed in detail here. The BBD is a fractional factorial design of Resolution V that can evaluate all linear and quadratic direct effects and all linear two-factor interactions. Figure 2 shows a BBD for three parameters and three levels, with 11 sampled points; whereas a corresponding FD would require 27 sampled points. A significant difference between the BBD and the three-level FD is that the former does not consider extreme combinations of parameters (i.e., there are no samples taken in the vertices or corners of the cube[7]). The $3^k$ factorial design considers the points on the vertices of the sample space. The CCD requires sampling at more than three levels for each parameter and also considers the points on the vertices of the sample space. When the number of experiments is limited because of cost or time considerations, it often is more efficient to restrict samples to a more central location and exclude unlikely extremes of parameter combinations.

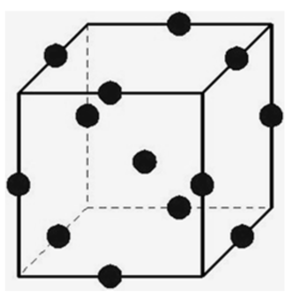

Fig. 2. Box-Behnken design for three parameters and three levels[7]

### II.C. Latin Hypercube Designs (LHD)

In the LHD method[8], each parameter is divided into N equally probable levels, where N is the number of runs or experiments to be made. LHD forces an equal sampling of each parameter and an even distribution of samples within the sample space. LHD has the advantage of simplicity, good coverage of the parameter space, and because each experiment is equally probable, the results can be used directly in computation of uncertainty.

### II.D. Performance Function Model

MC-MELODIE is a probabilistic framework model developed by IRSN with assistance from the Center for Nuclear Waste Regulatory Analyses (CNWRA®) of Southwest Research Institute® (SwRI®) to augment the deterministic 2-D and 3-D IRSN flow and transport computer code MELODIE. Mohanty *et al*[10] provides the details on MC-MELODIE. The combined characteristics of the radionuclides, waste package failures, host rock, and engineered barrier system are expected to control activity releases from the host rock to the surrounding formations. Total activity release from a failed waste package likely will depend on the interplay among diffusion rates, solubility limits, and decay rates. IRSN uses MELODIE to evaluate radionuclide transport at two scales: (i) within the COx formation, (i.e., COx model), and (ii) at the regional scale (MHM model).

The COx-scale model embodies the details of the repository at the level of excavated shafts. IRSN developed a simplified representation of the facility to focus on the IL-LLW zone, which is closest to the shafts and ramp. The overall domain includes a 135 m thick COx clay formation block that is 3 km by 2.6 km in the horizontal, with 86 cells receiving radioactive waste in packages backfilled and sealed into shafts. Each drift is assumed to be surrounded by an excavation damaged zone (EDZ) grading into a micro fractured EDZ. The cells are connected with backfilled access drifts to a set of two backfilled ramps and two backfilled shafts connecting to the surface.

Boundary conditions include (i) a no-flow boundary on the four lateral sides and (ii) zero activity concentration at the bottom and top faces. The top and bottom boundary conditions maximize radionuclide releases to the overlying and underlying aquifers. At the COx scale, the focus is on two transfer pathways: (i) diffusive transport pathway through the host rock and (ii) advective transport pathway through the drifts to the exits of the ramps and shafts. The source term is assumed to consist of one soluble and nonsorbing radionuclide ($^{129}I$) instantaneously released from all cells. The parameters defining this system include the flow and transport properties of (i) the COx formation and (ii) the engineered system, plus the hydraulic gradient across the COx formation (for vertical upward flow). The COx formation's parameters include flow and transport parameters, such as permeability, kinematic porosity, and diffusion coefficient. The engineered system parameters include permeability, kinematic porosity, and pore diffusion coefficient associated with drift and shaft backfill, concrete, seals, fractured EDZ, and micro-fissured EDZ. The COx model is the focus of the demonstrations in this paper.

## III. DEMONSTRATION OF SAMPLING METHOD

We provide an example below for a simple, but realistic case based on the results from the COx model of MELODIE. In this demonstration, we develop a simplified regression model for site performance, which was then exercised by three experimental design methods: two-level FD, BBD, and LHD, employing similar numbers of samples for each method. The results from exercising the regression model with the samples were then examined statistically to determine which factors are most sensitive, and how well a stepwise-regression model fits to the sample results.

*Regression Model Derived from MC-MELODIE Runs* - We executed the MC-MELODIE model to generate 100 realizations of results, which are referred to as "data" in the subsequent analyses. The computer runs produced two types of outputs: (i) combined flux from the COx formation, ramps, and shafts and (ii) flux from ramps and shafts only. Outputs used were for either the peak obtained within 1,000,000 years of simulation or at 1,000,000 years if still rising. Peak values from 99 realizations (one run was unusable) were considered adequate for the purposes of demonstrating the sampling methods. We were most interested in a relatively simple example that exhibited properties of the actual system being studied. We chose only the Ramp/Shaft flux result for demonstration because it showed more dependence on a wider range of parameters than total flux, which was dominated by the host-rock diffusion coefficient.

We then fitted a stepwise-regression equation to the logarithms of all parameters and peak ramp/shaft flux for the 99 MC-MELODIE runs. The resulting regression relationship is shown in Figure 3 and has an $R^2$ of 0.89. This original regression equation consisted of the sum of an intercept and six regression coefficients multiplying each parameter. In other words, the model is linear in each of the parameters in the logarithmic space; therefore, the surface is a flat hyperplane. As expected, any reasonable sampling scheme using this model would be able to produce a nearly perfect fit with only a small number of samples. Therefore, it was not possible to use the logarithmic regression model to discriminate among the three sampling methods. Sampling in Cartesian space is a more-realistic test of the three experimental designs for the following reasons: (i) regression model implies that factors multiply each other in Cartesian space, therefore the response surface is no longer flat; (ii) logarithmic model places too much emphasis on the (presumed less-important) smaller fluxes, while the Cartesian model places proper emphasis on the large fluxes; and (iii) BBD or other three-level sampling designs, would show no special benefit unless the response surface has curvature.

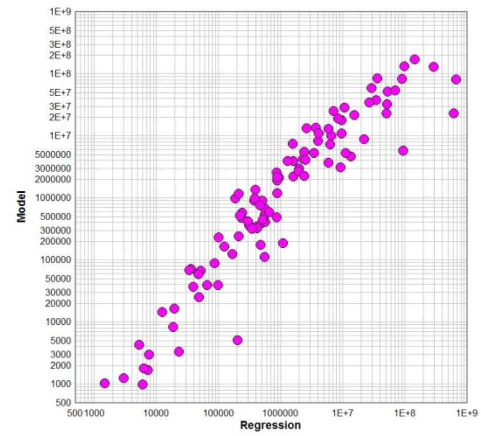


Fig 3. Regression model for ramp/shaft flux

Parameters for the regression ($X_1$ through $X_6$) are shown in Table 1. These correspond to K2, Ke, Kb, and Ks (i.e., the hydraulic conductivities of the COx, EDZ, backfill and seal, respectively in m/yr), Dp2 (i.e., the COx diffusion coefficient, $m^2$/yr), and H1mH3 (i.e., the difference in head vertically across the COx formation, m) in MC-MELODIE.

We demonstrated the relative effectiveness of LHD, FD 2-level design and BBD. The sampling range for all designs was taken as the 5$^{th}$ and 96$^{th}$ percentile of the original data used to fit the regression model (i.e., 5 values in from the high and low ends of the range). The middle value for the BBD used the 51$^{st}$ percentile. We used a similar number of samples for each experimental design: 60 samples for LHD, 64 samples for two-level FD, and 54 samples for BBD. Note the number of runs necessary for the FD and BBD is pre-determined by the number of parameters and cannot be changed. For LHD, the number of runs is flexible, but was chosen in this case to be about the average of the other two methods. Results produced from the samples were statistically analyzed to determine how well they agree with the data. One of the statistical procedures used, stepwise regression, analyzes how much of the fitted result is due to each parameter or parameter combination as they are added to the regression equation in terms of the coefficient of determination, $R^2$. In all cases, we used the software package JMP version 5 and the suggested statistical interpretations for each sampling routine[11].

**TABLE I. Increase in $R^2$ as Parameter is Added to Stepwise Regression. Rank is the Relative Sensitivity of the Factor.**

| | Latin Hypercube | | Full Factorial 2-Level | | Box-Behnken | |
|---|---|---|---|---|---|---|
| **Parameter** | **Increase in $R^2$** | **Rank** | **Increase in $R^2$** | **Rank** | **Increase in $R^2$** | **Rank** |
| $X_1$ (K2) | | | 0.0367 | 7 | — | — |
| $X_2$ (Ke) | 0.2022 | 2 | 0.0108 | 12 | — | — |
| $X_3$ (Kb) | 0.0227 | 5 | 0.0109 | 11 | — | — |
| $X_4$ (Ks) | 0.0116 | 8 | 0.0644 | 1 | — | — |
| $X_5$ (Dp2) | — | — | 0.064 | 2 | 0.106 | 2 |
| $X_6$ (H1mH3) | — | — | 0.0606 | 5 | — | — |
| $X_1^2$ | — | — | — | — | — | — |
| $X_2$ | — | — | — | — | — | — |
| $X_3^2$ | — | — | — | — | — | — |
| $X_4^2$ | — | — | — | — | — | — |
| $X_5^2$ | 0.0127 | 9 | — | — | — | — |
| $X_6^2$ | — | — | — | — | 0.0116 | 8 |
| $X_1X_2$ | — | — | — | — | — | — |
| $X_1X_3$ | — | — | — | — | — | — |
| $X_1X_4$ | — | — | 0.0365 | 8 | 0.016 | 7 |
| $X_1X_5$ | 0.0111 | 11 | 0.0364 | 9 | 0.0422 | 4 |
| $X_1X_6$ | — | — | 0.0343 | 10 | — | — |
| $X_2X_3$ | 0.0202 | 7 | — | — | — | — |
| $X_2X_4$ | — | — | — | — | — | — |
| $X_2X_5$ | — | — | — | — | 0.0239 | 6 |
| $X_2X_6$ | 0.3459 | 1 | — | — | 0.0403 | 5 |
| $X_3X_4$ | 0.0125 | 10 | — | — | — | — |
| $X_3X_5$ | 0.0084 | — | — | — | — | — |
| $X_3X_6$ | 0.027 | 12 | — | — | — | — |
| $X_4X_5$ | — | | 0.0638 | 4 | 0.0743 | 3 |
| $X_4X_6$ | 0.0843 | 3 | 0.064 | 3 | — | — |
| $X_5X_6$ | 0.0176 | 6 | 0.06 | 6 | 0.3881 | 1 |

*LHD Experiment:* The LHD used 60 samples from a uniform distribution for each parameter ranging from the 5th to 96th percentile of the original 100 MC-MELODIE realizations. Stepwise regression was performed for $X_1$ through $X_6$, their squares, and all cross-products, for a total of 27 parameters. The $R^2$ produced from the stepwise regression was 0.78 for the parameters $X_2$, $X_3$, $X_4$, $X_5^2$, $X_1X_5$, $X_2X_3$, $X_2X_6$, $X_3X_4$, $X_3X_5$, $X_3X_6$, $X_4X_6$, and $X_5X_6$. Keeping only the terms having the highest probability of significance based on an F statistic ($p < 0.05$), namely $X_2$, $X_4$, $X_5^2$, $X_1X_5$, $X_2X_6$, $X_3X_4$, $X_4X_6$, and $X_5X_6$, reduced $R^2$ to 0.75. The order of sensitivities from the stepwise regression to parameters for the LHS design is given in Table 1 and for both other experiments as well.

*FD experiment:* For the purposes of this demonstration, we chose a full-factorial approach. The full-factorial experiment used 64 (i.e., $2^6$) samples for all combinations of high and low parameters. The analysis considers all parameters and cross-products, but not squared terms, so these results cannot be directly compared to the LHD or BBD results. We note however that no squared parameters were shown to be of high significance in the LHD or BBD cases, as shown in Table 1. The $R^2$ from the stepwise regression was 0.54 for the parameters $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$, $X_4X_5$, $X_4X_6$,

$X_5X_6$, $X_1X_4$, $X_1X_5$, and $X_1X_6$. Order of parameter sensitivity is included in Table 1.

*BBD experiment:* The BBD for 6 parameters specifies 54 samples at the 5th, 51st, and 96th percentile of the original distributions from the MC-MELODIE run. The analysis routine in the statistics program[11] automatically considers all parameters, squares and cross-products; so the results can be compared directly to the LHS results. The $R^2$ from stepwise regression was 0.70 for the parameters $X_5^2$, $X_6^2$, $X_1X_4$, $X_1X_5$, $X_2X_5$, $X_2X_6$, $X_4X_5$, and $X_5X_6$. The order of parameter sensitivity is included in Table 1.

### III.A. Discussion

From this simple example, LHD appears to perform best based on the coefficient of determination, followed by BBD and FD. These results might not hold for larger problems with more sampled parameters and different distributions.

Ranking of sensitivity coefficients (i.e., the increase in $R^2$ when a parameter is added to the regression equation), initially does not appear to compare well among the sampling methods. However, on closer inspection, it is clear that cross-products of parameters show the importance of the parameters in combination with other parameters, but not necessarily by themselves. In the case of the FD, sensitivities of the top six parameters are closely ranked, making it difficult to compare this result to the rankings from the other two methods.

## IV. CONCLUSIONS

Methods for analyzing the effects of hydro-dispersive parameter uncertainties on the mass transport were tested. These methods include response surface methodology (RSM)-based study using BBD and two other experimental design methods, FD and LHD. A simplified regression model was developed using results from the probabilistic model MC-MELODIE to serve as the test model for the experimental design methods. Results, based only on the coefficient of determination, show that LHD appears to perform best, followed by BBD sampling and full-factorial two-level sampling. The three sampling methods show a significant difference in ranking order from one another. On closer inspection, the importance of the cross-products of parameters becomes evident. In the case of the full-factorial design, sensitivities of the top parameters are closely ranked; therefore it is difficult to compare this result to the rankings from the other two methods. However, BBD and other three-level sampling methods show an advantage over LHD and two-level factorial methods because it can directly evaluate the curvature in the function from the sampled results. For experimental designs limited to few samples (such as million year MC-MELODIE runs), the three-level sampling designs allow planning of further experiments to take advantage of the additional information on the curvature of the response surface. This property of three-level designs, while desirable, is not necessarily important for characterizing the sensitivity of parameters for a natural system[5]. Given the results of this computational experiment, and our experience with using sampling methods for repository performance assessments, we see no particular advantage for using sampling techniques other than LHD. The examples of sensitivity analysis methods used in this study to identify influential parameters in this paper were primarily for demonstration purposes. More sophisticated approaches are available and may be necessary for future studies.

## ACKNOWLEDGMENTS

The authors would like to thank Amelie de Hoyos (IRSN) and Gordon Wittmeyer (SwRI) for their thoughtful comments on the paper and Michael Muller (SwRI) for carrying out several computationally-intensive MC-MELODIE runs.

## REFERENCES:

1. Andra. "Dossier 2005 Argile. Tome: Safety Evaluation of a Geological Repository." Châtenay Malabry, France: Agence Nationale pour la gestion des Déchets Radioactifs (2005)

2. G. Deman, J. Kerrou, H. Benabderrahmane, and P. Perrochet. "Parametric Uncertainty Analysis on a Groundwater Flow and Mass Transport Model Using Experimental Design and Response Surface Methodology: The Case of Andra Meuse/Haute-Marne Site." Montpellier, France: *The 5th International Meeting on Clays in Natural and Engineered Barriers for Radioactive Waste Confinement*, Montpellier, France, October 22–25, 2012. Paris, France: Andra (2012)

3. F. Cornaton. "Deterministic Models of Groundwater Age, Life Expectancy and Transit Time Distributions in Advective-Dispersive Systems." Ph.D. thesis. University of Neuchatel. Neuchatel, Switzerland (2004)

4. A. Khuri, and, S. Mukhopadhyay. "Response Surface Methodology." Wiley Interdisciplinary Reviews: *Computational Statistics.* Vol. 2, Issue 2. March/April. pp 128–149 (2010)

5.S. Bisgaard. "Quality Quadries, Why Three-Level Designs Are Not So Useful for Technological

Experiments." *Report 148*. Madison, Wisconsin: Center for Quality and Productivity Improvement. University of Wisconsin (August 1996)
6. G.E. P. Box and J.S. Hunter. "The 2k-p Fractional Factorial Design, Part 1." *Technometrics*. Vol. 42, No. 1. pp 28–47 (2000)
7. NIST. *NIST/SEMATECH e-Handbook of Statistical Methods.* Washington, DC: Department of Commerce, National Institute of Standards and Technology,
http://www.itl.nist.gov/div898/handbook/>
(17 November 2013).
8. R.L. Iman and M.J. Shortencarier. NUREG/CR–3624, SAND83-2365, "A FORTRAN Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models." Washington, DC: U.S. Nuclear Regulatory Commission (March 1984)
9. Wikipedia. "Factorial Experiment, Fractional Factorial Design, Central Composite Design" 2013 <http://en.wikipedia.org/wiki/Factorial_experiment#History> (2 December 2013)
10. S. Mohanty, S. Stothoff, M. Muller, M. Bourgeois, A. de Hoyos, G. Mathieu, D. Pellegrini. "Probabilistic Modeling of Radionuclide Transport Evaluation for Geological Disposal in Meuse/Haute-Marne, France." *Proceedings of the 12th International Probabilistic Workshop*,
4-5 November, 2014, Weimar, Germany. 2014.
11. SAS. "Design of Experiments Guide." 2013. Cary, North Carolina: SAS Institute, Inc. <http://www.jmp.com/support/help/Latin_Hypercube_Designs.shtml> (25 November 2013).