

Construction of Meta Models for Aquatic Ecosystems, Derived via Discrete Event Simulation and a Two-Phase Design of Experiment Technique

Jorge Luis Romeu, Ph.D.

Dept. Mech. & Aerosp. Eng.; Syracuse University
Email: romeu@cortland.edu; Web: <http://web.cortland.edu/romeu>

ABSTRACT

This paper discusses how discrete event simulation can be used to model environmental problems, and the three key advantages it brings to environmental modeling. One is the possibility of including, in the simulation model, all the problem characteristics (e.g. rain, distribution policies, mass sizes, etc. and their socioeconomic and ecological impacts in different constituencies). The second advantage is the possibility of modeling without resorting to oversimplification of the problem. The third advantage is that, using Design of Experiment (DOE) techniques we are able to derive simpler Meta Models that can then be used as a legitimate proxies of the more complex original simulation model, for system design, study and optimization. To demonstrate this approach, we implement a prototype simulation model for an aquatic ecosystem. We discuss model requirements and constraints and develop a full factorial experiment. We then present different uses, including examples of how the simulation can be used for arbitration, to help different constituencies to arrive to a consensus. Finally, we extend the approach by analyzing an eleven-factor example, with and without factor interaction, introducing a two-phase DOE approach that handles such interaction case. The proposed two-phase DOE method is illustrated via two numerical examples.

RESUMEN

Este trabajo discute como la simulacion discreta puede ser utilizada para modelar problemas medioambientales, y las tres ventajas principales que trae a tal area. Una es la posibilidad de incluir, en el modelo de simulación, todas las características del problema (e.g. lluvia, políticas de distribución de aguas, tamaño de los subsistemas, etc. y sus impactos socioeconomicos y ecologicos en las distintas esferas. La segunda ventaja es la posibilidad de modelar, sin tener que simplificar el problema. La tercera es que, utilizando técnicas de Diseño de Experimentos (DOE) estamos en condiciones de derivar Meta Modelos mas sencillos, que pueden ser utilizados como substitutos legitimos de la simulación original, para estudiar, diseñar u optimizar. Para demostrar este metodo, implementamos un prototipo de simulación de un ecosistema acuatico. Discutimos los requerimientos y restricciones del modelo e implementamos un diseño factorial completo. Seguido presentamos diferentes usos, incluyendo ejemplos de cómo la simulación puede ser utilizada para arbitraje, para mover hacia un consenso distintos grupos socioeconomicos. Finalmente, extendemos el enfoque analizando once factores, con y sin interaccione entre estos factores, e introducimos una metodología que analiza, en dos fases, aplicando DOE fraccionados, la situación. Se dan ejemplos numericos.

INTRODUCTION

Environmental issues continue to be of great importance in the XXI Century. For they affect health, drive production costs up (short and long term) and are felt across borders, making them a global problem. In addition, ecological problems are fairly complex, involving many physical, human and socioeconomic factors that often interact with each other. Therefore, building tractable mathematical models for ecological problems is difficult and often leads to oversimplification of model assumptions. As a result, these mathematical models bear little resemblance with the real problem entities being modeled.

In this paper we propose the use of a combination of discrete event simulation, via specialized languages such as GPSS (General Purpose Simulation System), and of Design of Experiments (DOE), to derive Meta Models (functional models of the simulation). This approach presents several advantages over purely mathematical ones, for environmental studies. For example, in a simulation, the modeler can use broader distribution assumptions, in addition to Exponential or Normal. The modeler can also consider all problem factors and their interactions. There is no longer need to keep the model formulation simple, in order to obtain a closed form solution. These two gains more than justify pursuing a simulation modeling approach.

In addition, by implementing DOEs on such simulation models, the experimenter can (1) identify the Key problem factors that impact the response under study, and (2) derive a class of Meta Models (yielding similar indices of fit, with different variables) that can be used, in lieu of the original and more complex simulation, to optimize, design, assess etc. Then, when a specific region of the variable space has been identified, the entire simulation model can be used again in it, to refine the Meta Model results.

In the rest of this paper we overview modeling efforts using discrete event simulation in ecosystems, then develop an example of simulation/DOE approach for a simple, three-factor case. Finally, we extend it to an eleven factor model, developing an approach that implements fractional factorial DOEs in two phases, to derive useful Meta Models.

LITERATURE REVIEW

From the above, it seems apparent that simulation is an appropriate and useful tool in environmental studies and, therefore, should have been frequently used. However, as we will see in this section, it has seldom been applied in this area, and the question is "why not"? To try to answer it, we overviewed, among other sources, the last decade of the Winter Simulation Conference (WSC) an important forum for research on discrete event simulation. We found that (i) not many simulationists and operations research (O.R.) engineers have worked in modeling environmental problems, and (ii) few environmentalists use simulation to model ecological problems. This is interesting, since other quantitative researchers such as statisticians, are actively applying their models and tools in ecological studies.

Since 1990 we found few WSC papers dealing with environmental topics. For example, in 1992 one WSC session was dedicated to natural resource models. Three papers were presented, dealing with ecological issues. Knisel et al. (1992) describe the development of a mathematical model with three basic components (hydrology, erosion and chemistry) to evaluate pollution from agricultural and forestry sources. The model

simulates impact of management on response of field-size areas to climatic input. Wiles et al. (1992) conducted simulation experiments to evaluate scouting plans for use with weed control decision model for soybeans. As in the present paper, the authors were concerned with demonstrating the simulation methodology as a valid tool for evaluation of alternatives. Chang and Odeh (1992) presented a simulation model to generate monthly data for water quality using a first order Markovian process.

Clymer (1993) in a limited review of the simulation literature discusses the merits "of discrete event model applications in the field of global simulation". Clymer discovers that "very little use has been made of discrete event models in world simulation." He suggests a list of eight global problems for discrete event simulation (including land, fresh water, atmosphere and ocean physics and chemistry, including pollutants and thermal effects). And he presents a short bibliography of papers about modeling ecological problems or related subjects.

In 1994, WSC dedicated a session to waste management and environmental modeling problems and were six papers were presented. Boomer et al. (1994) develop a dynamic model that simulates expected activities occurring between underground waste storage tanks and define activities and functions for planned systems. Hand and Barr (1994) present two flexible simulation models for trans ceramic waste and perform trade-offs between different system design alternatives. Shaver (1994) describes a simulation model for the development of a solid waste management system and evaluates alternative system configurations, throughput and capacity for the proposed solutions.

Hoefler et al. (1994) simulate a safety subsystem at an operating nuclear plant and compare different models on the basis of cost. Ross et al. (1994) simulate environmental restoration activities and remedial strategies for contaminated water and storage facilities, predicts system behavior, and compares alternative solutions. Lehr et al (1994) discuss the sensitivity analysis for a simulation model that investigates the oil weathering process. Ioannou (1999) uses simulation in dam construction, but only to optimize the use of construction equipment. Finally, only Romeu (1995 and 1997) briefly outlines a GPSS simulation model of an aquatic ecosystem, within several educational applications of simulation, and then proposes its use for applied environmental work. And we have found little if anything in the area, after that.

Modeling work done by ecologists, using simulation, is not much broader. For example, Hall, Jourdonnais et al. (1989), among selected few, implement simulation models (in FORTRAN) for the Flathead River Basin, Montana, an ecological system including a lake and stream fish reserve. They showed how simulation provides a neutral context in which to negotiate the trade-offs between economic and ecological advantages of a given resource management strategy, and arrive to a satisfactory solution.

Ecologists using simulation in their research, usually employ HOL languages (e.g. FORTRAN). The development of HOL programs can take several months and hundreds of lines of code. Using simulation with a specialized language such as GPSS, the development of these same programs take significantly less effort.

Finally, some statisticians actively work in ecological problems. For example, the American Statistical Association (ASA) has an Environmental Section, publishes a newsletter and offers yearly modeling and paper awards. Several other statistical societies also work in this area and regularly publish their work in refereed journals.

MODELING METHODS

Simulation modeling has been characterized, and rightly so, as a method of last resort. The first modeling choice is always a theoretical model, when it exists. Unfortunately, it often doesn't, or the required assumptions are so stringent that one is forced to oversimplify them in order to obtain a closed form solution. Secondly, we can turn to an empirical model, most often of the regression type. However, due to the many problem factors and their (first, second and higher order) interactions, the amount of data required to develop the model is beyond availability. In addition, the functional form selected for the regression model is usually arbitrary. Hence, we are left with the simulation approach as our "last resort tool", since all other available modeling techniques are no longer applicable or are very difficult to implement.

The Conceptual Model

For implementing the simulation model we need (i) knowledge of, and relationships between, system components and (ii) data to estimate the distributions and parameters that describe their functions. Subject matter experts can provide the first. And the data for modeling and validation (the second requirement) can be obtained from the census as well as hydrological, industrial, business and agricultural databases of the local, state and federal governments. From the different system components we obtain the simulation model factors. From their inter-relationships, we obtain the factor interactions. In the simulation model, these elements can be defined in minute detail, something more difficult (or even impossible) to do in the theoretical or regression models.

The use of special purpose simulation languages, such as GPSS, make the actual programming and debugging efforts less onerous than if HOL languages (e.g. FORTRAN, C) were to be used. Finally, the simulation model can yield a host of responses and performance measures of interest. Moreover, these model responses can then be recombined, via the use of adequate weights, into broader meta-responses that allow a more objective comparison of the conflicting and competing interests of the various aquatic ecosystem constituents (e.g. ecologists vs. business developers).

Once the simulation program is written and verified (programming part is correct) it needs to be validated. That is, we need to show that the simulation actually mimics the real system it purports to model. When modeling existing aquatic ecosystems, the real data also exists. Hence, we can use it for validation (to compare it with the simulation model output). Once the validity of the model is established, we can start experimenting with it, which is the final objective of the entire simulation exercise.

Simulation is a numerical procedure. Hence, running it once and obtaining a point estimator is not very useful. We want to run a series of similar cases (a sample of pseudo-independent and identically distributed observations) and use them to derive confidence intervals (CI) for the system responses or performance measures of interest. We can also implement what-if analyses and hypotheses tests for specific system conjectures of interest. In addition, we can investigate whether some specific mode factor, or a combination of them, effect the model response and, if so, to estimate the magnitude and direction of these effects. Such results are obtained via the implementation of statistical experiments, usually factorial or (if many effects are under

scrutiny) fractional factorial designs. To illustrate this approach we present, in the next section, a simulation model for an aquatic ecosystem.

The Simulation Model

The prototype simulation model described here has been successfully presented, among others, to the US Environmental Protection Agency (EPA), the Great Lakes Research Consortium (GLRC), and the Mexican Institute of Water Technology (IMTA).

The simulator models an aquatic ecosystem common along the coast of the Gulf of Mexico. It is composed of estuaries, wetlands, lagoons and river-lake systems. Two specific examples of these ecosystems are the river ports of Tampico and the Papaluapan-Alvarado estuary, both in Mexico. Such ecosystems are very important for the preservation of species and their habitats (wetlands, lagoons) as well as for their socioeconomic development (river commerce, industry, agriculture, urban uses, energy and recreation). But these two activities define two constituencies (environmentalists and business developers) that often compete for, and disagree about, the ecosystem resources and how to better use them. Hence, there is need for an analysis tool (the simulation model) to help them better resolve their differences and work together.

We built a GPSS simulation model of a system of two inter-connected water environments: the Panuco (the river port of Tampico, in Mexico) and its lagoon. These two water environments (s) constitute our real ecosystem. In our simulation model, each has its own parameters: maximum and minimum water capacity levels, physical configuration and water distribution policies (Figure 1). These parameters reflect the socioeconomic and political requirements, constraints and characteristics of the two above-mentioned constituencies. The simulation model assesses the overall effects that constituent-driven changes in the ecosystem factors produce in the operation levels and the services rendered by the ecosystem (e.g. port activities, agricultural yields, urban quality of life). The simulation model also assesses the corresponding environmental effects (e.g. deterioration of wildlife habitats).

Figure 1 shows a schematic of the above-described aquatic ecosystem, which has several deterministic and stochastic water delivery obligations.

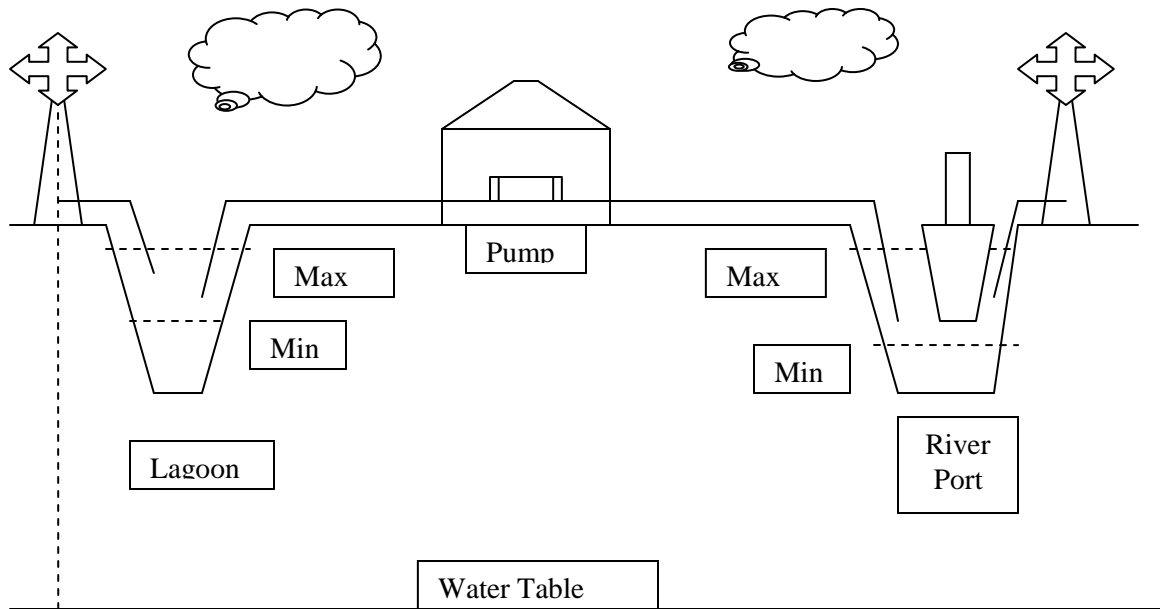


Figure 1. Schematic of the River Port and Lagoon aquatic ecosystem.

The lagoon provides water for drinking and other personal uses to the population, as well as for agriculture, animal husbandry, etc. It also provides humid environments (wetlands) for wildlife reproduction as well as recreational facilities for the population.

The river port provides the means of transportation to bring in raw materials and food, and to export regional products. Its derived economic activity supports agriculture, local industry, commerce and general transportation needs of the population, therefore generating badly needed jobs. In our model we do not consider the effects of evaporation or possible water theft which, if need arises, can also be incorporated into the simulation. All the above-described activities and services constitute outputs of our aquatic ecosystem model.

On the other hand, water for all system obligations come from only three ecosystem sources: rain, the water table and “borrowing” from the other water mass, either lagoon or the river port itself. Let’s explain.

When the rainfall (which is a stochastic input, whose statistical distribution is specified in the simulation model) is insufficient, the lagoon or the river levels drop below acceptable levels (Min, in Figure 1). Then, the other (river or lagoon) entity can “lend” some of its water via a water transfer. Alternatively, and as a last resource, it can also obtain water from the water table. However, there is a high economic and environmental cost associated with pumping water out from the water table, as it is associated with many future ecological problems. All the aquatic ecosystem ecological and economic variables or factors defined in Figure 1, and their trade-offs, are summarized in Table 1.

Table 1. Aquatic Ecosystem Trade-Off Design/Decision Variables and Costs

Controlled Variables

Economic

- * EcoSite Replenishing Levels (MIN)
- * EcoSite Capacities (MAX)
- * EcoSite Ordering Schedule
- * Water Transfer Policy
- * Water Usage Policy
- * Water Shortage Policy
- * Water Table Level
- * EcoSite Initial Conditions

Ecologic

- * Wetland size
- * Water Temp
- * Transfer Speed
- * Water Salinity
- * Pollution
- * Area/Depth
- * Fish/Foul Populat
- * Fish/Foul Reprod

Associated Costs:

- * of pumping water from Water Table
- * of pumping water from EcoSite to EcoSite
- * of sale of water to users
- * of water shortages
- * of lack of water
- * indirect costs
- * total costs

- * of foul/fish extinction
- * of biodegradation
- * of habitat Damage
- * of habitat clean up
- * of habitat recondit
- * of species relocalizat
- * total costs

Uncontrolled Variables:

- * Water Table capacity
- * Weather conditions
- * Rain schedules
- * Water Theft

- * Salinity
- * Oxygen
- * Temperature
- * Evaporation

Types of Problems:

- * New system design
- * Existing system optimization
- * Comparison of strategies
- * Education of the population

- * Regulation trade-off
- * Impact evaluation
- * Conflict mediation
- * Risk Analysis

The river port capacity and configuration design (Max, in Figure 1) depends on the available water and the commerce needs. If there were enough water, the river canal could be widened to accommodate more ships, enhancing its commercial capacity. But as water falters during the dry season, the river level drops below acceptable levels (Min) and ships cannot operate unless additional water is transferred from the lagoon or the water table to fill the gap. On the other hand, a narrower river canal would ensure the uninterrupted port operation at lower water levels, without need for drawing excessive water from the lagoon, or draining the water table. But this would also let fewer ships access the river port, inducing a slower economic activity.

The total system operation cost, as well as the costs of distributing water among competing user constituencies (agriculture, industry, commerce, general population) constitute the economic objective, or response function of interest. We include the corresponding linear, non-linear, deterministic and stochastic problem constraints (Max, Min, rainfall distribution, etc). The simulation model can also consider other performance measures such as the number of water transfers between *s*, the number of (and time to first) water stock-outs, changes in agricultural, industrial and commercial outputs, in levels of employment, etc. all of which are a direct consequence of the levels of water available in the two *s* (or water masses). Other problem constraints include operating budgets, available land, ecological mandates, water distribution priorities, etc.

The response variable used in our example is total cost of water transfers. This included, in addition to the cost of physically transferring water, also the indirect costs induced by the ensuing lack of water (affecting the economy, ecology, population lifestyle, etc). Other ecological responses (e.g. size of the wetland areas, changes in flora and fauna reproduction levels) may also be used. The possibility of trading off real life conflicting (economic vs. ecological) management policies (e.g. rationing versus selling water) adds another important feature to the simulation modeling capabilities.

Model Operation

Our GPSS simulation model was used to assess and compare the economic and ecological effects of diverse water management strategies and river and canal designs, on the entire ecosystem. The simulation was also used to assess the joint impact of these strategies and designs on the ecological, social and economic resources of the entire ecosystem.

In particular, we wanted to assess the joint effect of three factors. One, was the use of the water from the water table versus the use of several inter- water transfer policies, to balance water deficits within the ecosystem. This factor is important because the water table depletion produces serious and long-term adverse effects on the flora and fauna, as well as on agriculture and animal husbandry. In addition, the amount and speed of the ensuing water transfers affect water temperature and salinity, which, in turn, impacts fish and fowl reproduction.

Another factor of interest was the *Ecosite configuration* and design. Port and canal design configurations (e.g. their sizes and capacities) impact the available water in an ecosystem. The size of the river port (capacity for servicing one or two ships) determines the level of commerce, thus impacting the number of jobs in industry, agriculture, etc. Therefore, for economic reasons, we want the larger port. On the other hand, maintaining such larger capacity during the dry season requires draining water from the lagoon and perhaps even from the water table. This, in turn, affects agriculture and the environment. Lagoon water draining reduces crop yields and levels of species reproduction (smaller wetlands). In addition, tampering with the water table may produce irreversible damage to the ecosystem. Therefore, for ecological reasons, we want a small port. This is a classical example of two different constituencies that conflict, due to requiring the use of the same, constrained (water) resources, while espousing very different value systems.

The third factor of interest, seasonality, depends entirely on nature and mainly serves to complicate the above-described situation by its interaction with the first two factors.

When rain is scarce and the port design is too large, there is a water shortage. Maintaining the port open during the dry season then requires draining the water table or the lagoon, thus endangering the health of the ecosystem as well as the agriculture and animal husbandry activities. And there is also less water to serve the general needs of the population (drink, hygiene, etc.).

Thus, the three factors included in our statistical model (and their levels) were the following. First, River Port capacity (denoted α), which had two levels: high (two ships) or low (only one). The second factor was Seasonality (denoted β) with two levels, defined by the rain schedules for the rainy and dry seasons. The third factor (denoted γ) was the (inter-site) Water Transfer Policy. The first priority was always pumping from the water remaining above the alternative respective minimum level (transfers were 1/3 or 1/2 of such availability). Only when the alternative did not have enough water, was it allowed to pump water from the water table. The factor interactions were denoted with double letters (e.g. $\alpha\beta$). Each factor had two levels, denoted with letters $i=j=k=2$. Finally, we used $l=9$ replications for each run setting or factor combination.

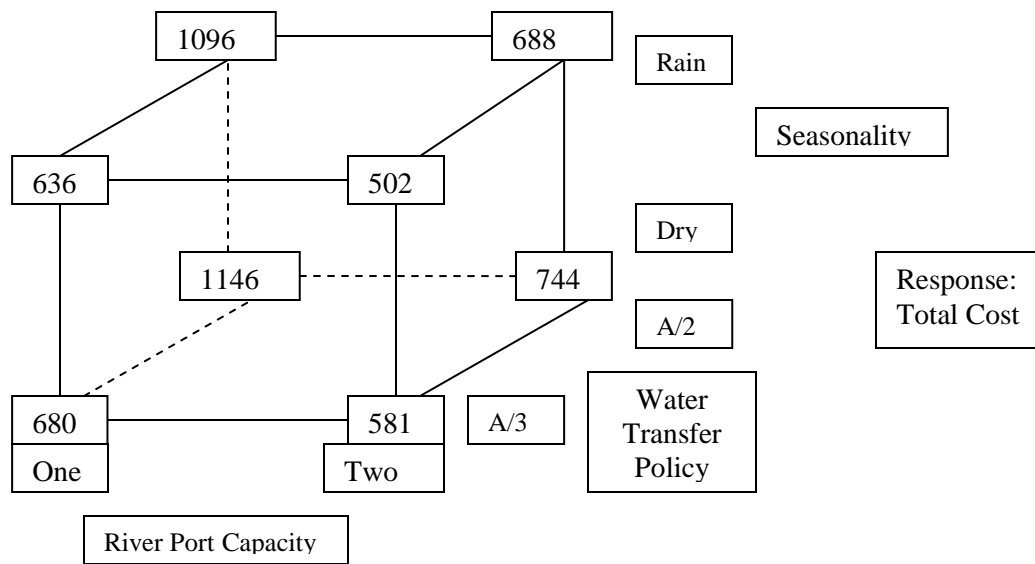
The mathematical expression of the 2^3 complete DOE factorial design (e.g. Montgomery, (2001), Box, Hunter and Hunter, (1986), Cox and Cochran (1992), implemented in this case is given by:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl} \quad (1)$$

The response variable (Y) selected was the Total Operating Cost, considered in the following manner. The port capacity (one or two ships) influenced, through the amounts of goods moved in and out of the region, the levels of employment in agriculture, industry and commerce. On the other hand, capacity also determined the water shortage levels. Water transfer policies defined the amounts (and sources) of water "loaned" from one to the other, during a water shortage. Smaller amounts implied more frequent transfers -but allowed more water for use in its original , e.g. transportation, wetlands or agriculture. On the other hand, larger but less frequent transfers decreased the levels of the lagoon and river port, also effecting the transportation and agriculture as well as the wetland sizes. These, in turn, affected employment and species reproduction levels. Finally, seasonality (rainy and dry seasons) aggravated the above-mentioned situations, directly as well as via its interaction with the two other factors.

Therefore, the response function Y included the cost of the actual water transfers as well as the social, economic and ecologic "costs" that such transfers induced. Individual costs were defined by placing a dollar value on their impact on each of the aforementioned elements, by unit of water level lost, in each of the two s. In our simulation, costs were defined by the "business developer". Similarly, experiments using our model but with costs defined by "ecologists" or other interested parties can also be run and their results can be compared. This is one more advantage of simulation in these types of studies.

Figure 2: Complete Factorial Experiment for the Simulation



RESULTS

The GPSS simulation model was run for the 2³ complete factorial experiment. A batch means output analysis method was used. Each batch consisted of one month of daily operation, starting at “average level”. The first ten batches were discarded to account for the transient or warm up period. After this, we took every fifth batch, to avoid inter-batch correlation, for a total of nine replications. The analysis of variance results (Table 2) show how every main effect (as well as the interaction between capacity and water transfer policy) were all statistically significant. The other interactions were not statistically significant.

Table 2: Analysis of Variance Table for the Simulation Experiment

Source	D. F.	Mean Square	F Value	P-Value
River Port Capacity	1	1219401	588.71	0.000
Water Transfer Policy	1	1828892	882.96	0.000
Seasonality (Rain/Dry)	1	58186	28.09	0.000
Enviro-Site x Policy	1	373104	180.33	0.000

The capacity factor (design of the river port) was highly significant. The amount of water (and hence of transfers) required to maintain the river port open (as well as adequate wetlands for reproduction needs) depend on the river port size. But river port capacity also affects the need for water transfers, whose costs include the (environmental and economic) effects produced by such transfers. The statistical analysis showed that the single most significant factor was Water Transfer Policy. Assessing water transfer sizes and their induced costs was an important analysis objective. Assessing the effects of establishing different water prices versus establishing water distribution policies (i.e. rationing) is now possible with the simulation model.

Another important problem element, that can now be analyzed, is water user profile (e.g. different agricultural, industrial, commercial and household water requirements). Water usage impacts directly and indirectly the quality of the life of both, the population and the ecosystem. It produces low level, short-term, effects (e.g. lack of water for lawns, for swimming pools). It can also reduce the size of wetlands for wildlife reproduction, for habitats, etc. inducing long-term effects. Stochastic user profiles can be included in the simulation model, something much more difficult to implement either in the theoretical or empirical models. Simulation can also consider the joint effect of several of these factors (Table 1), via complete and fractional factorial experimental designs, obtaining through them, valuable quantitative information.

We can see in Table 3 how the two largest changes in the Cost response function (in respectively inverse direction) correspond to factors Size of River Port and Water Transfer Policy. These factors modify the total cost (response) in Equation (1) by the amount and the direction indicated Table 3, (adding or subtracting from average cost).

Table 3: Examples of model-derived quantitative information (CI of changes in Total Cost)

Factor Response	Change Effectuated	95% CI Factor Effect on Response
Size of the river port	One to two ships capacity	Cost decrease in \$232 to \$287
Water Transfer Policy	Transfer 1/3 to 1/2 of availability	Cost increase in \$293 to \$344
Seasonality	Dry into Rainy Season	Cost decrease in \$33 to \$80
Size x Transfer Policy	Dry/one ship to Rain/Two ships	Cost decreases in \$122 to \$165

Table 3 is based on results of the 2^3 complete factorial design implemented (Figure 2). For example, the most expensive (total cost) factor combination (\$1146) corresponds to operating a port designed for one ship, during dry season and with 1/2 availability water transfer policy. And the least expensive combination (\$502) corresponds to operating a port designed for two ships, during rainy season and with a transfer policy of 1/3 of the available water. The best of the two water transfer policies analyzed is moving only one-third of the available water in the basin. Under such policy, there is not a very large difference between a one or two ship port design –but two ships means more trade.

Again, the simulation results presented are based upon “cost values” defined by the “business developers”. The simulation can also be run with different cost values (of agricultural production losses, wetland deterioration, species reproduction reduction, etc.) defined by other constituencies (ecologists, farmers, etc.). Such other cost definition may yield totally different results. These can then be traded-off, compared, etc. We can also run alternative “what if” scenarios. The simulation now also becomes a decision-support and an arbitration tool.

Finally, multi-criteria (ecological, social, economic, etc.) system responses (using the elements in Table 1) can now be obtained, by combining (say k) contrasting and competing individual responses (say Y_1, \dots, Y_k) into a single, complex one. This is done by defining a set of weights (say $\alpha_i; 1 \leq i \leq k; \sum \alpha_i = 1$) that respond to the different priorities or costs, defined by each different constituency interested in the ecosystem. Thence, linear combinations can be formed, using such weights ($\sum \alpha_i Y_i$), to quantify

The contrasting policies and philosophies of these different constituencies. This way, a more objective comparison of such competing and contrasting policies, produced by the (unbiased) simulation model results, can now provide a more rational environment where these diverse constituencies can discuss their differences and better reach a consensus.

EXTENSIONS TO HIGHER DIMENSIONS

The practical use of the combination of discrete event simulation and DOE, to derive Meta Models for complex environmental problems, will now be illustrated by applying it to solve an eleven-factor study. The eleven factors considered are given below:

1. A=Replenishing Levels (Minimum water)
2. B=Reservoir Capacity (Maximum water))
3. C=Ordering Schedule (from which reservoir)
4. D=Transfer Policy (how much water to take)
5. E=Allocation to each sector (e.g. industry, agriculture, population)
6. F=Size of the Reservoirs (extension)
7. G=Generation of electricity
8. H=Hospitals and schools support
9. I=Wetland size for reproduction
10. J=Water Table maximum usage
11. K=Fish/Foul Population to support

The response of interest is again represented by “total cost” of operation (to society as a whole, including costs to the environment, etc.). Such a cost can be established by the different constituencies (e.g. industrialists, ecologists, the population at large) and the different solutions can then be contrasted and traded-off.

If a Complete 2^k DOE were used (here for $k=11$ factors), and only two replications per run were made, we would have to conduct a total of 4096 experiments, something out of the possibility of most researchers. Instead, we will first use a Plackett-Burnham (1946) fractional factorial design. Three replications per factor settings now yield only $3 \times 12 = 36$ total runs, a much smaller effort, by far than with the complete design.

In our simulation, we generated the data using, as the true model, the following:

$$y_i = \mu + 6A + 2B - 4D - 6E - 2G + 4H + 8I - 8J + \varepsilon_i(2)$$

Factors	A	B	C	D	E	F	G	H	I	J	K	Bo
RegCoef	6	2	0	-4	-6	0	-2	4	8	-8	0	12

The levels of each factor were ± 1 , and there were no interactions assumed (we will analyze interactions later in the paper, using an adaptive procedure we are currently exploring, as the Plackett-Burnam approach has problems handling such problems). The downside of this approach is that interactions are confounded with Main Factors. Having no interactions, therefore, solve the problem. However, in reality there are second and higher order interactions present, in most cases.

The reader can compare the estimated regression coefficients with the true ones, as well as the Main Factor effects, which are the double of the corresponding regression coefficients. Main effects C and F were correctly detected as Non Significant. Main Effect K, which was Non Significant, was erroneously detected as Significant (at $\alpha=0.05$). Finally, all Significant Main Effects, except G, were detected as such.

The Plackett-Burnam DOE analyses results are given below:

Plackett-Burnam Design for Eleven Factors

Run	A	B	C	D	E	F	G	H	I	J	K	Avg
1	1	-1	1	-1	-1	-1	1	1	1	-1	1	36.14
2	1	1	-1	1	-1	-1	-1	1	1	1	-1	24.39
3	-1	1	1	-1	1	-1	-1	-1	1	1	1	0.5
4	1	-1	1	1	-1	1	-1	-1	-1	1	1	-5.96
5	1	1	-1	1	1	-1	1	-1	-1	-1	1	2.62
6	1	1	1	-1	1	1	-1	1	-1	-1	-1	31.26
7	-1	1	1	1	-1	1	1	-1	1	-1	-1	21.12
8	-1	-1	1	1	1	-1	1	1	-1	1	-1	-10.54
9	-1	-1	-1	1	1	1	-1	1	1	-1	1	15.92
10	1	-1	-1	-1	1	1	1	-1	1	1	-1	12.02
11	-1	1	-1	-1	-1	1	1	1	-1	1	1	7.33
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	11.66
Factors	A	B	C	D	E	F	G	H	I	J	K	Bo
RegCoef	6	2	0	-4	-6	0	-2	4	8	-8	0	12
RegEst.	4.5	2.3	-0.1	-4.3	-3.6	1.4	-0.8	5.2	6.1	-7.6	-2.8	12.2
MainEff	12	4	0	-8	-12	0	-4	8	16	-16	0	n/a
EstimEff	9.1	4.7	-0.2	-8.6	-7.2	2.8	-1.5	10.4	12.3	-15.2	-5.6	12.2
Signific.	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes

To explore the situation where there is interaction between factors, we used the same simulation model described above, with the following more complex functional form:

$$y_i = \mu + 6A + 2B - 4D - 6E - 2G + 4H + 8I - 8J + 2AB - 4HI + GJ + DE + \varepsilon_i(3)$$

But now, we used a two-phase approach. First, we implemented a traditional Plackett-Burnam to all factors, and used its results only to separate the factors into (1) those with positive coefficients, and (2) those with negative coefficients. Then, we proceeded to implement a Resolution IV Fractional Factorial design (Box et al. (1086), Cochran and Cox, (1992), Montgomery, (2001)) on each group.

We have implemented this approach, our modification of the Controlled Sequential Bifurcation method (Wan et al., (2003), Sanchez and Wang, (2005)) with success, in our recent investigations consisting in the derivation of a collection of alternative but similarly efficient (per model R^2) Meta Models for very large simulation systems.

Results of the above-described first phase, the implementation of the eleven-factor Plackett-Burnam, are given below:

Plackett Burnam results on the Interaction example

Factors	A	B	C	D	E	F	G	H	I	J	K
MainEff	12	4	0	-8	-12	0	-4	8	16	-16	0
FacEstim	-98.6	61.1	41.3	-86.5	98.4	66.4	79.7	51.8	-26.6	37.6	-96
RegPar.	6	2	0	-4	-6	0	-2	4	8	-8	0
RegEstim	-49.3	30.5	20.6	-43.2	49.2	33.2	39.8	25.9	-13.3	18.8	-48
Signific.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

It is evident, from the analysis of the above table, that (1) the determination of the factor significance and non-significance, from the Plackett-Burnam results, are incorrect, and (2) that the numerical estimations of the coefficients are very far off.

We use, however, this first stage exclusively to divide all factors into two groups: those with a positive, and those with a negative coefficient. We implement below a Resolution IV DOE for the positive group of variables: B, C, E, F, G, H, and J.

Resolution IV Design for the Positive Group

Factors	B	C	E	F	G	H	J	Bo
TRUE	12	4	0	-8	-12	0	-4	12
EffectEstim	12.14	2.53	1.17	-7.2	-11.82	0.39	-3.49	13.59
RegCoef	6	2	0	-4	-6	0	-2	12
RegEst.	6.07	1.26	0.59	-3.6	-5.91	0.19	-1.75	6.8
Signific.	Yes	Yes	No	Yes	Yes	No	Yes	

We can see how this second analysis yields both, correct significance/non significance determination in the factors, and closer estimations to the true values in the coefficients of the simulated model. Notice how the coefficient signs have changed to negative per the model, and how the Main Effects now come out clear.

CONCLUSIONS

We have shown how simulation can be effectively used to model the complex ecological problems in our countries, and how, by successive application of two types of Fractional Factorial DOEs (e.g. Plackett-Burnam and Resolution IV designs) we can extract, from complex simulations, simpler Meta Models that can then be used as proxies, to design, optimize or to search for regions of interest. The second is a new approach.

REFERENCES

- BOOMER, K., et al. (1994). "Hanford Tank Waste Remediation System Overview: a dynamic simulation model". *Proceedings of the Winter Simulation Conference*. 1167--1173.
- BOX, HUNTER and HUNTER. (1986). *Statistics for Experimenters*. Wiley
- CLYMER, B. (1993). "Applications of Discrete Event and Combined Modeling to Global Simulation". *Proceedings of the Winter Simulation Conference*. 1135--1138.
- COCHRAN and G. COX. (1992). *Design of Experiments*. Wiley.
- HALL, C. A., JOURDONNAIS, J. H. and J. A. STANFORD. (1989). "Assessing the Impacts of Stream Regulation in The Flathead River Basin, Montana, USA. A Simulation Modeling of System Water Balance". *Regulated Rivers: Research and Management*. 3:61--77.
- HAND, F. and G. BARR. (1994). "TRU Waste System Models". *Proceedings of the Winter Simulation Conference*. 1174--1179.
- HOEFER, P., C. MADU, E. DAGHER, S. V. DAVIS and J. M. DONNELLY. (1994). "Simulating One Dimension of Safety and Operational Efficiency at a Nuclear Power Plant". *Proceedings of the Winter Simulation Conference*. 1186-1189.
- IOANNOU, P.G. (1999). "Construction of a Dam Embankment with Non Stationary Queues". *Proceedings of the Winter Simulation Conference*.
- KNISEL, W. G., R. A. LEONARD and F. M. DAVIS. (1992). "Simulating Processes in Non Point Source Pollution". *Proceedings of the Winter Simulation Conference*. 1159—1165.
- LEHR, W. et al. (1994). "Model Sensitivity Analysis in Environmental Emergency Management: a case study in oil spill modeling". *Proceedings of the Winter Simulation Conference*. 1198--1205.
- MONTGOMERY, D. (2001). *Design and Analysis of Experiments*. Wiley.
- ODEH, R. Y and T. J. CHANG (1992) "Use of Simulation Method for Surface Water Quality Data." *Proceedings of the Winter Simulation Conference*, pp. 1172—1179.
- PLACKETT, R. L. and J. P. BURNMAN (1946). "The Design of Optimum Multifactorial Experiments". *Biometrics*.
- ROMEY, J. L. (1995). "Simulation and Statistical Education". *Proceedings of the Winter Simulation Conference*.
- ROMEY, J. L. (1997). "More on Simulation and Statistical Education". *American Journal of Mathematics and Management Sciences*, 17; 3 & 4 (Special issue on simulation).
- ROSS, T., D. A. KING, M. L. WILKINS and M. F. FORWARD. (1994). "Predicting Environmental Restoration Activities Through Static Simulation". *Proceedings of the Winter Simulation Conference*. 1190--1195.
- SANCHEZ, S. and WANG. (2005). "Two-Phase Screening Procedure for Simulation Experiments". *Proceedings of the Winter Simulation Conference*.
- SHAVER, S. (1994). "Hanford Solid Waste Management System Simulation". *Proceedings of the Winter Simulation Conference*. 1180--1185.
- WAN, ANKENMAN and NELSON. (2003). "Controlled Sequential Bifurcation: a new factor screening method for discrete event simulation". *Proceedings of the Winter Simulation Conference*.
- WILES, L. J., G. G. WILKERSON and H. J. GOLD. (1992). "Simulating Weed Scouting and Weed Control Decision Making to Evaluate Scouting Plans". *Proceedings of the Winter Simulation Conference*. 1166—1171.