# NISS

# Sensitivity Analysis of Computer Models: The World Bank HDM-III Model

Donath Mrawira, William J. Welch,
Matthias Schonlau, and Ralph Haas

Technical Report Number 75
March, 1998

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

# Sensitivity Analysis of Computer Models: The World Bank HDM-III Model

Donath Mrawira[1], William J. Welch[2], Matthias Schonlau[3], and Ralph Haas[4]

Key words: Computer experiment, data needs, life-cycle cost, maintenance and rehabilitation.

[1]Donath Mrawira is a Postdoctoral Fellow in the Department of Civil Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

[2]William J. Welch is a Professor in the Department of Statistics and Actuarial Science and The Institute for Improvement in Quality and Productivity, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

[3]Matthias Schonlau was a graduate student in the Department of Statistics and Actuarial Science and The Institute for Improvement in Quality and Productivity, University of Waterloo, when this research was conducted. He is now a Postdoctoral Fellow at the National Institute for Statistical Sciences, P.O. Box 14006, Research Triangle Park, NC 27709-4006, USA.

[4]Ralph Haas is a Professor in the Department of Civil Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

ABSTRACT

A problem often arising in engineering applications of computer models is to determine the importance of each data item in the large pool of required input factors. This paper explores a statistical approach for investigating factor sensitivities. The methodology is demonstrated with the HDM-III highway life-cycle cost analysis model. Specifically, the net present value (NPV) of life-cycle costs predicted by the HDM-III model is analyzed, and sensitivities of NPV to the link characterization input factors are investigated. In the statistical designed experiment, combinations of the input factors are chosen using Latin hypercube sampling, a method well suited to the deterministic HDM-III model. Two analyses of the output data are performed, based on a first-order regression approximation and on a Gaussian stochastic-process model. For NPV, the factor rankings are similar, but the sensitivities obtained from the two techniques show some marked differences. This demonstrates the greater flexibility of the stochastic-process model in dealing with nonlinearities and factor interactions in complex input-output relationships.

# INTRODUCTION

Use of mathematical (computer-based) models has become very popular in engineering. Good examples in transportation engineering and road management include the World Bank HDM-III (Watanatada 1987), the British Transport Research Laboratory RTIM3 (Cundill and Withnall 1995), the Ministry of Transportation of Ontario OPAC-2000 (He et al. 1996), etc.

These models tend to be complex and computationally intensive with a large number of input variables. Sensitivity analysis of the model factors has, as a result, evolved as an indispensable tool, both for evaluating the reliability of the decisions based on the model outputs and for determining the most influential factors. The latter information is useful in subsequent model applications, for example, to prioritize expenditure on data collection (Mrawira 1996), to generate approximate/preliminary analyses, to develop simplified model versions, etc.

Traditionally, sensitivity analysis has been used to assess whether some input factors to a decision making process require further careful examination so as to reduce uncertainties associated with the decision taken (Little and Mirrlees 1974). Ashley (1980), in an article investigating the influence of factor errors in traffic forecasting models, points out why sensitivity analysis by computer simulation is sometimes inevitable. Computer based models are often too complex to analyze analytically.

The traditional approach to sensitivity analysis is to change one factor at a time, the so-called *ceteris paribus* method. More efficient, alternative methods have been developed in the framework of statistical design of experiments. In a designed experiment, levels of one factor are combined with levels of the other factors in a planned fashion. They are more efficient since they yield more reliable estimates of factor sensitivities and, more importantly, they can identify factor interactions. The differences between *ceteris paribus* and designed experiments are summarized in Table 1.

The objective of the paper is to demonstrate a comprehensive methodology for investigating factor sensitivities of complex computer models in engineering applications.

The HDM-III model, commonly used by road agencies in managing road networks, is used as a case study. Specifically, we analyze the economic net present value (NPV) predicted by HDM-III.

The next section outlines the data requirements for the HDM-III model. We then describe the Latin hypercube design used for the sensitivity analysis experiment. Latin hypercubes (McKay et al. 1979) are a class of experimental plans proposed specifically for deterministic computer codes like HDM-III. They are easily adapted to deal with inequality constraints on the input factors and allow estimation of interaction effects, the critical limitation of *ceteris paribus* experiments. Latin hypercubes can have many levels for each input, and are thus well suited to the detection of highly nonlinear effects. Statistical modeling of the nonlinear input-output relationship is then taken up. We describe two methods. A first-order regression model is straightforward to implement and interpret but cannot deal with nonlinear behavior or interaction effects. The second method, based on a stochastic-process model for the input-output relationship, can model nonlinearities and interactions without explicit specification of the functional forms. After describing the results of the sensitivity analysis for the HDM-III model, we conclude with a summary of the specific findings and some discussion of the methodology in general.

## THE HDM-III MODEL

The World Bank Highway Design and Maintenance Standards Model (HDM-III) is popular in highway agencies as a tool for evaluating and analyzing maintenance and rehabilitation options. The model allows comparison of policies or standards and can play an important role in supporting decision-making in the road investment sector in general (Watanatada et al. 1987, World Bank 1989). The model estimates detailed life-cycle pavement deterioration, agency costs, and road users' costs for different design and maintenance alternatives, and hence provides rational and consistent economic decision criteria for technical planners and policy makers.

Priority programming constitutes one of the most important functions of pavement management analysis. The purpose, at both the project and network levels, is to compare project alternatives for implementation. A detailed treatment is given by Haas et al. (1994); see also the reviews by, for example, Haas et al. (1985) and Liebman (1985). The shaded area in Figure 1 shows the major components of priority programming. The priority analysis component is the subject of this paper.

Priority analysis is in general a tedious and computationally intensive procedure. It requires an explicit consideration of the primary effects of traffic, link characterization attributes and pavement standards, as well as the effects of maintenance intervention upon the cost streams arising throughout the life-cycle of a road facility (Chesher and Harrison 1987; Paterson 1987). This complexity is the motivation behind software tools like HDM-III, RTIM3, etc. These models are primarily used as analysis engines to provide the function of economic evaluation either independently or within larger pavement management processes.

The concept of applying HDM-III as an analysis engine in a network-level pavement management system has been demonstrated widely: for example in Brazil (Queiroz et al. 1992) and in Queensland, Australia (Robertson and Charmala 1994; Howard et al. 1994). In Brazil, HDM-III was used in conjunction with the Expenditure Budgeting Model (EBM) which uses a heuristic technique to solve a multi-year budget constraint problem (World Bank 1989) providing the optimization analysis function.

Figure 2 shows the HDM-III data requirements relevant to the rehabilitation and maintenance (R&M) priority-programming problem. The investigation in this paper considers the link characterization factors. They are fairly detailed and numerous; link characterization for a paved road, for instance, requires about 34 input attributes, plus an option for pavement deterioration calibration parameters. These requirements are especially daunting for low-income road agencies, and may constitute a major disincentive for adopting the model. Although some of these input requirements are optional (with default values supplied internally by the model) no formal guide exists in the literature identifying the most sensitive input factors on which the user could focus to achieve a reasonable precision in the life-cycle predictions.

The current literature on sensitivity analysis of HDM-III is very limited. Notable contributions on this subject include Mrawira and Haas (1996), Kerali et al. (1991), Queiroz et al. (1991), and World Bank (1988). However, all of these studies were based on the *ceteris paribus* approach. Furthermore, most of them focused on one investment strategy, upgrading a gravel road to paved standard.

## LATIN HYPERCUBE EXPERIMENTAL DESIGNS

Conventional factorial designs were developed for experiments where the response is measured with random error. In contrast, Latin hypercube designs (McKay et al. 1979) were proposed specifically for deterministic computer codes like HDM-III. They allow estimation of complex nonlinearities and are particularly well suited to the flexible modeling approach in the subsection titled "The Gaussian Stochastic Process Model".

The 35 link-characterization input parameters and their ranges are listed in Table 2. A total of 500 runs are performed to explore the input space. Some input variables are constrained as follows:

$$\begin{aligned} ACRA + ARAV + APOT &\leq 100 \\ ACRW &\leq ACRA \\ RDS &\leq RDM \\ AGE1 \leq AGE2 &\leq AGE3 \\ ACRWb &\leq ACRAb. \end{aligned}$$

Consider first the unconstrained variables. Each is given a grid of possible values. For example, we give *MMP* 26 equally spaced levels between 5 and 300, i.e. a spacing of 1/25 of the range. This number of levels allows detection of nonlinearities if they exist, whereas a traditional two- or three-level design (e.g., Box, Hunter and Hunter 1978) cannot detect highly nonlinear behavior. Each of the 26 levels of *MMP* will be repeated 19 or 20 times in the 500 runs of the code. The remaining variables are put on similar grids.

For a completely random Latin hypercube the 500 values for *MMP* would be in random order. Similarly, the second input, *A*, etc. Combining, for example, the *MMP* and *A* values in this way hopefully fills out the two-dimensional *MMP*-*A* space, representing all combinations of these two parameters. Similarly, there will be many combinations of any three parameters, etc. Randomly combining the columns, however, would produce correlation between variables from chance, which would make it difficult to separate their effects on the outputs. Iman and Conover (1982) described how to transform a starting, completely random Latin hypercube to one with very good correlation properties. Iterating their procedure produces near-zero correlations here.

To illustrate how we deal with the constraints, consider the constraint $ACRW \leq ACRA$. In the Latin hypercube design we have $ACRA$ on the range [0, 60] (see Table 2) and a pseudo parameter, $ACRW'$ on the range [0, 2/3]. The variable $ACRW'$ from the Latin hypercube is transformed to

$$ACRW = ACRW' \times ACRA.$$

This guarantees that $ACRW$ is in the range [0,40] and that $ACRW \leq ACRA$. The other constraints are dealt with similarly.

Designing in this way gives plenty of levels to many parameters but requires relatively few runs.

The HDM-III model is run for each of the 500 parameter combinations in the experimental plan. The output chosen for the case study is the NPV of life-cycle costs for a roughness-dependent overlay strategy.

## STATISTICAL MODELING

### First-Order Regression Approximation

We first try fitting a first-order regression model linear in the 35 input variables to the NPV values from the 500 HDM-III runs. The regression model is

$$Y = \beta_0 + \beta_1 \frac{x_1}{b_1 - a_1} + \beta_2 \frac{x_2}{b_2 - a_2} + \ldots + \beta_j \frac{x_j}{b_j - a_j} + \ldots + \beta_{35} \frac{x_{35}}{b_{35} - a_{35}} + \epsilon, \qquad (1)$$

where $Y$ is the NPV response variable of interest, $x_j$ is the $j$th input variable with range $[a_j, b_j]$ from Table 2 for $j = 1, \ldots, 35$, and $\beta_0, \ldots, \beta_{35}$ are coefficients to be estimated by least squares. In such models, $\epsilon$ is usually regarded as random error with mean zero and constant variance, but the error is actually deterministic model bias in our context.

The advantage of this model is that $\beta_j$ provides a direct estimate of factor sensitivity for $x_j$. Because we divide each $x_j$ by its range, $\beta_j$ is the effect on NPV of changing $x_j$ from its minimum to its maximum value, when the other factors are kept fixed, up to error from $\epsilon$. The major disadvantage is that this model does not deal with nonlinearities or interactions between input variables and, hence, may be a poor approximation.

### The Gaussian Stochastic Process Model

As the first-order regression model does, indeed, provide a poor fit to the data (see the subsection titled "Assessing the Approximating Models"), we also employ a more flexible modeling approach.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_{500}$ denote the input vectors for the 500 runs in the experimental design. Each vector $\mathbf{x}$ is 35-dimensional for the 35 inputs $x_1, \ldots, x_{35}$. The corresponding NPV output values are denoted $\mathbf{y} = (y_1, \ldots, y_{500})^T$, where "$T$" indicates vector or matrix transpose. Then, following the approach of, e.g., Welch et al. (1992), the response $y(\mathbf{x})$ is treated as a random function or as a realization of a stochastic process,

$$Y(\mathbf{x}) = \mu + Z(\mathbf{x}), \qquad (2)$$

where $\mu$ is the mean of $Y(\mathbf{x})$ and the stochastic process $Z(\mathbf{x})$ is assumed to be Gaussian, and to have zero mean and covariance $\sigma^2 R(\mathbf{x}, \mathbf{x}')$ between the values of $Y$ at input vectors $\mathbf{x}$ and $\mathbf{x}'$.

This model can be motivated as follows. Assuming continuity of the response and some smoothness, when two input vectors $\mathbf{x}$ and $\mathbf{x}'$ are close together the outputs are expected to be similar. As the distance between the two input vectors increases, the similarity of the outputs is expected to decrease. Mathematically, we express similarity as the correlation function $R(\cdot, \cdot)$ of the stochastic process, which can be tuned to the data. Here it takes the form

$$R(\mathbf{x}, \mathbf{x}') = \prod_{j=1}^{35} \exp(-\theta_j |x_j - x_j'|^{2-\alpha_j}), \tag{3}$$

where $\theta_j \geq 0$ and $0 \leq \alpha_j < 2$ for $j = 1, \ldots, 35$. The parameter $\alpha_j$ controls the smoothness of the response as a function of $x_j$ (smoother as $\alpha_j$ decreases to zero), while $\theta_j$ controls the nature of the variability in the response (more local as $\theta_j$ increases).

Note that $R(\mathbf{x}, \mathbf{x}) = 1$, so the output from replicate runs would be perfectly correlated. Thus, although we are using a stochastic model to represent uncertainty about the function, the model respects the deterministic nature of the computer code.

The best linear unbiased predictor of $y$ at an untried $\mathbf{x}$, denoted by $\hat{y}(\mathbf{x})$, can be shown to be (see e.g., Sacks et al. 1989):

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{4}$$

where $\mathbf{r}(\mathbf{x})$ is the $n \times 1$ vector of correlations between $\mathbf{x}$ and each of the $n$ design points with element $i$ given by $R(\mathbf{x}, \mathbf{x}_i)$ in (3), $\mathbf{R}$ is an $n \times n$ correlation matrix with element $(i, i')$ given by $R(\mathbf{x}_i, \mathbf{x}_{i'})$ in (3), $\mathbf{1}$ is an $n \times 1$ vector with all elements equal to 1, and $\hat{\mu} = \mathbf{1}^T\mathbf{R}^{-1}\mathbf{y}/\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}$ is the generalized least squares estimator of $\mu$. The further parameters, $\sigma^2$ following (2), and $\theta_1, \ldots, \theta_{35}$ and $\alpha_1, \ldots, \alpha_{35}$ in (3), are estimated by maximum likelihood.

The predictor interpolates the observed response values, as it should for the deterministic HDM-III model. It has proven to be accurate for numerous applications, see e.g., Currin et al. (1991), Sacks et al. (1989), and Welch et al. (1992).

## Assessing the Approximating Models

Cross-validation is commonly used to assess the performance of a predictor. The response value $y_i$ for model run $i$ is predicted using all data except $y_i$. Figure 3 shows the actual NPV values from the HDM-III model plotted against their cross-validation predictions from (a) the first-order regression model (1) and (b) the stochastic process model (2). The stochastic-process model gives superior prediction accuracy for NPV. The cross validated root mean squared errors of prediction for the regression model and for the stochastic-process model are 5.23 and 2.87 percentage points, respectively.

# RESULTS

Table 3 summarizes the estimated regression coefficients in the first-order regression model (1) for NPV from the HDM-III model. It lists only the factors with estimated coefficients large enough that the null hypothesis $H_0 : \beta_j = 0$ is rejected at the 1% significance level in favor of the two-sided alternative $H_a : \beta_j \neq 0$ ($p$ value less than 0.01). As the output from HDM-III is deterministic, and hence the errors in the regression model (1) do not satisfy the usual probabilistic assumptions, statistical significance should be viewed here as descriptive rather than inferential. The factors are ordered in the table by the magnitudes of their estimated coefficients. Because the factor ranges are standardized in (1) the estimated coefficients estimate the change in average NPV when a factor changes from its minimum value to its maximum value, if all other factors are kept constant.

From Table 3 it appears that the most significant factors affecting NPV are the rutting calibration factor ($Krp$), the pavement width ($W$), the mean and the standard deviation of rutting ($RDM$ and $RDS$), the distress parameters ($ACRW$ and $APOT$), the pavement roughness ($QI$), the rise plus fall ($RF$), and cracking initiation ($Kci$).

The stochastic-process predictor (4) is a complex function of all 35 input parameters. To determine factor sensitivities, the predictor has to be separated into effects of individual parameters or small groups of parameters.

To isolate the effect of, say, *Krp* on NPV the predictor (4) is averaged with respect to all other input variables over the ranges in Table 2. The resulting *main effect* of *Krp* is shown in Figure 4(a). It is seen that *Krp* has a large estimated effect on NPV relative to the pointwise approximate 95% confidence intervals also shown in Figure 4(a) to provide an estimate of uncertainty. Moreover, the estimated effect of this important factor is highly nonlinear, which is one reason why the first-order regression approximation is poor here. Four further important main effects are shown in Figure 4; it is seen that the estimated $QI$, *Kci*, and $RF$ effects in plots (c), (d), and (e), respectively, are also nonlinear.

To quantify these estimated effects we can decompose the total variability in the predictor (4) over all 35 input parameters into contributions from individual factors, or groups of factors, and their interactions. Table 4 shows the breakdown. For example, the strong estimated effect of *Krp* in Figure 4(a) accounts for about 43% of the total variability in the predictor from varying all 35 input factors over the ranges in Table 2.

In Table 4 the combined contribution of the rutting factors $RDM$ and $RDS$ is reported. These two factors could not be varied independently in the experimental design because of the constraint $RDS \leq RDM$. It is difficult, therefore, to estimate their separate effects. The 6.7% for their joint contribution reported in Table 4 is calculated in the following way. First, the remaining 33 factors are averaged out from the predictor, leaving a function of $RDM$ and $RDS$. This estimated joint effect of $RDM$ and $RDS$ is shown in the contour plot in Figure 5. The variability in the predicted NPV across the $RDM$-$RDS$ region shown in Figure 5 accounts for 6.7% of total predictor variability, as reported.

It is seen in Figure 5 that there are large changes in the estimated effect as $RDS$ varies for fixed $RDM$, but changing $RDM$ for fixed $RDS$ has relatively little effect. This suggests that $RDS$ is a much more important factor than $RDM$. The pointwise standard errors attached to the estimated NPV values in Figure 5 are all around 0.5 percentage points, i.e., fairly small relative to the changes in Figure 5. Thus, there is little doubt that $RDS$ is the more important factor. In contrast, the regression model ranked *RDM* as more important than *RDS*. One problem inherent with the regression model is that it estimates the effect of changing *RDM*, *when all other factors are kept fixed.* As *RDM* increases, the constraint $RDS \leq RDM$ does not restrict *RDS* as much and *RDS* also tends to be larger in the experimental design. The inaccurate first-order regression model probably has difficulty separating the effects of these two related inputs.

Some interaction effects also show up as important. For example, Table 4 reports a 2.5% contribution to total NPV predictor variability from the interaction between *Krp* and *QI*. This interaction contribution is computed in the following way. First, the joint effect of *Krp* and *QI* together is estimated by averaging out the remaining 33 factors from the predictor. The resulting estimated joint effect on NPV is shown in Figure 6. Much of the variability in this estimated joint effect is due to the strong estimated main effects of *Krp* and *QI* in Figure 4. If we subtract the two estimated main effects from the estimated joint effect, the result is the estimated interaction (nonadditive behavior) of *Krp* and *QI* on NPV. The variability in the estimated interaction across the rectangular region of *Krp* and *QI* values accounts for 2.5% of total predictor variability (in addition to the 42.6% and 5.6% main effect contributions).

Inspection of Figure 6 indicates that *Krp* is a very important factor, but its effect is dependent on the level of *QI*. When *QI* is at its lower limit, 1.5 IRI, NPV is estimated to change from about -2% to around 3% then back down to about -6% as *Krp* changes over the range [0.2, 4]. For *QI* at the midpoint of its range, 5.75 IRI, the estimated peak NPV is about 11%, at about *Krp*= 1.5. When *QI* is at its upper limit, 10 IRI, the estimated peak NPV is about 10%, but the peak occurs when *Krp* is near its lowest value, 0.2. Thus, there can be no single number representing *Krp* sensitivity, which is another reason why the linear regression model was inaccurate. (The same arguments could be repeated for *QI*: its estimated effect depends markedly on the level of *Krp*.)

Similarly, Table 4 reports some interaction between *Krp* and the rutting factors, *RDM* and *RDS*, and these interactions should also be explored in the same way to understand the *Krp* effect. Overall, the effect of *Krp* in the HDM-III model seems to be highly complex, depending on *QI*, *RDM*, and *RDS*.

To confirm that the *Krp* effect is highly nonlinear and, moreover, that it depends on *QI*, for example, the HDM-III model was run for all combinations of *Krp* at the 11 levels 0.2, 0.58, 0.96,..., 4.00 and *QI* at the three levels 1.5, 5.75, and 10 IRI. The other 33 factors were fixed at the midpoints of their ranges in Table 2. The plots in Figure 7 show the actual NPV values from the HDM-III model as a function of *Krp* for the three levels of *QI*. They confirm the features of the estimated joint effect of *Krp* and *QI* in Figure 6.

In summary, according to the stochastic-process model, the most important factors affecting NPV are the rutting progression (*Krp*), the carriageway width (*W*), road roughness (*QI*), and the pavement distress parameters (*RDS*, *RDM*, *ACRA*, *ACRW*, *APOT*, and *ARAV*). NPV is also sensitive to the pavement strength (*SN*) and the alignment grade (*RF*) for the overlay strategy used in the analysis.

It was noted above that the sensitivity of NPV to *Krp* is strongly dependent on the levels of other factors, in particular, the roughness (*QI*) and deformation distresses (*RDS* and *RDM*). If we average over these other factors, Figure 4(a) shows that for *Krp* less than about 1.5, an overlay is more beneficial for pavements as the rate of rutting progression increases. After the maximum at a *Krp* of about 1.5, the benefit decreases sharply with increasing rate of rutting. This is as expected in practice since a thin overlay (50 mm is assumed in this study) does not correct the rutting problem.

The shape of the estimated main effects in Figure 4 of the roughness (*QI*), the carriageway width (*W*), and the alignment grade (*RF*) are similar to earlier findings (Mrawira 1996). The effect of roughness on NPV was shown in the earlier study to flatten out for roughness beyond 7 IRI. This is again logical since an overlay treatment (or other comparable strategies) will not improve the serviceability for extremely damaged pavements. In other words, it is inefficient to overlay a road which has exceeded 7 IRI roughness. Other treatments (e.g., major rehabilitation) should be prescribed. The sensitivity of carriageway width (*W*) to NPV is linear and negative since it is a direct multiplier in agency annual costs but without significant savings on user vehicle operating costs. The slow increase in NPV with RF reflects the impact of alignment grade on truck and heavier vehicle speeds, and hence increased fuel and tire consumption. Figure 4(e) suggests that it is marginally more economical to overlay a steeper road section than a flat one.

In the context of prioritization of maintenance and rehabilitation works, the alignment and road width are normally fixed. Thus, the primary sensitive factors in R&M programming for weak pavements under moderate traffic volumes are the progression of rutting, the pavement distress parameters (particularly roughness and rut depth and its variability, but also the surface distresses), the structural strength, and the rates of cracking initiation and roughness progression.

The regression model and the stochastic-process model agree fairly well qualitatively: They give roughly the same rankings for the important input parameters. Quantitatively, though, there are some marked differences. Increasing the rutting calibration factor, *Krp*, from 0.2 to 4.0 is estimated to reduce NPV by nearly 16 percentage points according to the regression model (see Table 3). The more accurate predictor from the stochastic process model estimates a reduction of 11–12 percentage points [see Figure 4(a)]. This latter estimate averages the effect over the other factors. As noted above, the effect of *Krp* on NPV is greatly modified by other factors, though. For a pavement with a low value for *QI* (around 1.5 IRI), *Krp* has less effect on NPV over the range [0.2, 4.0] (see Figures 6 or 7). Another difference is that the regression model estimates a reduction in NPV of over four percentage points when *RDM* changes from 0 to 50 mm. Yet, the stochastic process model and Figure 5 suggest that *RDM* has a much smaller effect on average NPV for any fixed value of *RDS*.

# CONCLUSIONS

The paper has demonstrated that Latin hypercube experimental designs offer a sound methodology for investigating factor sensitivities of computer models. Such designs insure that the physical/practical factor ranges of the input space are fully explored.

They are also flexible enough to be easily modified to deal with constraints between factors.

Statistical approximating functions were found useful in analyzing the NPV output from the HDM-III model to determine factor sensitivities. The main advantage of the regression approximation is that it is straightforward and the estimated coefficients are immediately interpretable. The first-order model cannot identify interactions or nonlinearities in factor effects, however. If we had chosen the range [0.2, 2.0] instead of [0.2, 4.0] for *Krp* in the HDM-III runs, then *Krp* would have had a very small estimated effect in the regression model, missing much of the rise and sharp drop apparent in Figure 4(a). The regression model could be improved by introducing quadratic nonlinearities and bilinear interactions. Factor sensitivities would then be less immediate, however, and there is the practical difficulty of choosing significant coefficients from a second-order model with over 600 terms.

The stochastic-process predictor approach after Sacks et al. (1989) is computationally intensive but provides more comprehensive estimates of factor effects. Nonlinearities and interactions are identified in a fairly automatic way without the analyst needing to specify functional forms. Visualization of the predictor provides insight into the complexities of a model like HDM-III.

For the NPV output from the HDM-III model, the most significant factor was found to be the rutting calibration (*Krp*). The dominance of *Krp* in NPV predictions for weak older pavements under moderate traffic loading (a departure from the current literature) was confirmed by a further investigation of the response as a function of *Krp* at three levels of initial roughness ($QI$). The confirmation showed the behavior of the *Krp* effect is greatly modified by the level of $QI$. NPV was also very sensitive to the carriageway width ($W$), the initial deformation parameters ($RDM$ and $RDS$), and the distress parameters ($ACRA$, $ACRW$, $APOT$, $ARAV$).

## ACKNOWLEDGMENTS

Welch's research was supported by NSERC of Canada and by the National Science Foundation through grant NSF-DMS 9313013.

## APPENDIX I. REFERENCES

Ashley, D.J. (1980). "Uncertainty in the context of highway appraisal." *Transportation*, 9 (3), 249–267.

Box, G.E.P., Hunter W.G., and Hunter J.S. (1978). *Statistics for experimenters.* Wiley, New York.

Chesher, A. and Harrison, R. (1987). *Vehicle operating costs: Evidence from developing countries.* Johns Hopkins University Press, Baltimore.

Cundill, M. A. and Withnall, S.J. (1995). "Road Transport Investment Model RTIM3." *Proc., Sixth International Conference on Low-Volume Roads, Vol. 1.* Transp. Res. Board, Washington, D.C., 187–190.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments." *Journal of the American Statistical Association*, 86 (Dec), 953–963.

Haas, R.C.G., Karan, M.A., Cheetam, A., and Khalil S. (1985). "Pavement rehabilitation programming: A range of options." *Proc., North American Pavement Management Conference, Vol. 2.* Toronto, March 1985, 6.41–6.54.

Haas, R., Hudson, W.R., and Zaniewski, J. (1994). *Modern pavement management.* Krieger Publishing Co., Malabar, Florida.

He, Z., Kennepohl G., Haas, R., and Cai, Y. (1996). "OPAC 2000: A new pavement design system." Paper presented at the Transportation Association of Canada Annual Conference, Charlottetown, October 1996.

Howard, R.H., Robertson N.F., and Francisco, R. (1994). "Introduction of investment analysis into pavement management practices in the Philippines." *Proc., Third International Conference on Managing Pavements, Vol. 1.* Transp. Res. Board, Washington, D.C., 267–277.

Iman, R.L. and Conover, W.J. (1982). "A distribution-free approach to inducing rank correlation among input variables." *Communications in Statistics Part B: Simulation and Computation*, 11 (3), 311–334.

Kerali, H.R., Snaith, M.S., and Koole, R.C. (1991). "Economic viability of upgrading low-volume roads." *Transp. Res. Rec. 1291, Vol. 1.* Transp. Res. Board, Washington, D.C., 34–40.

Liebman, J.S. (1985). "Optimization tools for pavement management." *Proc., North American Pavement Management Conference, Vol. 2.* Toronto, March 1985, 6.5–6.15.

Little, I.M.D. and Mirrlees, J.A. (1974). *Project appraisal and planning for developing countries.* Heinemann, London.

McKay, M.D., Conover, W.J., and Beckman, R.J. (1979). "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code." *Technometrics*, 21(2), 239–245.

Mrawira, D. (1996). "Streamlining the World Bank's highway design and maintenance standards model for network level application." Ph.D. Thesis, University of Waterloo, Waterloo, Canada.

Mrawira, D. and Haas, R. (1996). "Towards adapting HDM-III model to local conditions: Lessons from a case study in Tanzania." *Transp. Res. Rec. 1524.* Transp. Res. Board, Washington, D.C., 58–66.

Paterson, W.D.O. (1987). *Road deterioration and maintenance effects: Models for planning and management.* Johns Hopkins University Press, Baltimore.

Queiroz, C., Fonseca, V.M.R., and dos Santos, A.B. (1991). "Economic appraisal of Guinea-Bissau roads program using HDM-III and EBM models." *Transp. Res. Rec. 1291, Vol. 1.* Transp. Res. Board, Washington, D.C, 105–115.

Queiroz, C., Hudson, W.R., and Haas, R. (1992). "Standardization of pavement management systems in Brazil and other developing countries." *Transp. Res. Rec. 1344*. Transp. Res. Board, Washington, D.C, 31–37.

Robertson, N.F. and Charmala, R. (1994). "Applications of HDM-III to road upkeep investment studies in Queensland." *Proc., International Workshop on HDM-4, 1, 95-112.* Kuala Lumpur, Malaysia.

Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. (1989). "Design and analysis of computer experiments" (with discussion). *Statistical Science*, 4 (4), 409–435.

Watanatada, T., Harral, C.G., Paterson, W.D.O, Dhareshwar, A.M., Bhandari, A., and Tsunokawa, K. (1987). *The highway design and maintenance standards model, Vol. 1: Description of the HDM-III model.* Johns Hopkins University Press, Baltimore.

Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., and Morris, M.D. (1992). "Screening, predicting, and computer experiments." *Technometrics*, 34 (1), 15–25.

World Bank (1988). *Road deterioration in developing countries: causes and remedies. A World Bank Policy Study.* World Bank, Washington D.C.

World Bank (1989). *HDM-PC: The highway design and maintenance standards model: User's Guide and PC Disks, Version 2.0.* World Bank, Washington, D.C.

# APPENDIX II. NOTATION

*The following symbols are used in this paper:*

| | |
|---|---|
| $n$ | Number of runs in the experimental design |
| $R$ | Correlation function of $Z$ |
| $\mathbf{r}$ | Vector of correlations in (4) |
| $\mathbf{R}$ | Matrix of correlations in (4) |
| $x_j$ | Value of input factor $j$ |
| $y$ | NPV value from HDM-III model |
| $Y$ | Statistical model for $y$ |
| $\mathbf{y}$ | Vector of HDM-III NPV values |
| $Z$ | Stochastic process in model (2) |
| $\alpha_j$ | Correlation parameter for input factor $j$ in (3) |
| $\beta_0$ | Intercept in regression model (1) |
| $\beta_j$ | (Linear) effect of input factor $j$ in model (1) |
| $\epsilon$ | Error term in model (1) |
| $\mu$ | Mean of $Z$ |
| $\sigma^2$ | Variance of $Z$ |
| $\theta_j$ | Correlation parameter for input factor $j$ in (3) |
| $\mathbf{1}$ | $n \times 1$ vector of 1's in (4) |

Further symbols, for the input factors, are defined in Table 2.

| Ceteris paribus experiments | Designed experiments |
|---|---|
| Experiments in which only one factor at a time is varied; all other factor are kept constant | Experiments which may change the levels of all factors from one run to the next |
| • Can be inefficient (large number of runs) | • Efficient (better accuracy with fewer runs) |
| • Not capable of detecting factor interactions | • Interaction effects can be easily analyzed |
| • Requires detailed knowledge of system for an informed search | • Experimental runs give global coverage of input space |
| • Useful if the model has few factors and is approximately additive | • Can handle a large number of factors even with nonadditivity (interaction) |

Table 1: Motivation for Designed Experiments

| Symbol | Factor description | Range |
|---|---|---|
| *MMP* | Average monthly rainfall | 5–300 mm |
| *A* | Altitude | 0–2500 m |
| *RF* | Rise plus fall | 0–120 m/km |
| *C* | Horizontal curvature | 0–700°/km |
| *W* | Carriage-way width | 2.5–10 m |
| *WS* | Shoulder width | 0–2 m |
| *HSNEW* | New surface layer thickness | 10–200 mm |
| *HSOLD* | Old surface layer thickness | 10–200 mm |
| *CMOD* | Strength of soil cement | 0.5–30 GPa |
| *HBASE* | Total base layers thickness | 100–700 mm |
| *COMP* | Relative compaction | 85–100 % |
| *SNSG* | Subgrade CBR | 2–50 % |
| *SN* | Structural number | 0.5–6 |
| *Kci* | Cracking initiation | 0.2–4 |
| *Kcp* | Cracking progression | 0.5–3 |
| *Kvi* | Raveling initiation | 0.2–3 |
| *Kge* | Roughness-age term | 0.8–1.4 |
| *Kpp* | Pothole progression | 0.2–3 |
| *Krp* | Rut depth progression | 0.2–4 |
| *Kgp* | Roughness progression | 0.8–2 |
| *ACRA* | Area of all cracks | 0–60 % |
| *ACRW* | Area of wide cracks | 0–40 % |
| *ARAV* | Area raveled | 0–40 % |
| *APOT* | Area of potholes | 0–5 % |
| *RDM* | Mean rut depth | 0–50 mm |
| *RDS* | Rut depth standard deviation | 0–40 mm |
| *QI* | Roughness | 1.5–10 IRI |
| *CQ* | Construction quality | 0, 1 |
| *AGE1* | Age of preventive treatment | 0–30 years |
| *AGE2* | Age of surfacing | 0–30 years |
| *AGE3* | Age of last construction | 4–30 years |
| *CRP* | Cracking retardation time | 0–3 years |
| *RRF* | Raveling retardation factor | 1–4 |
| *ACRAb* | Previous area of all cracks | 0–60 % |
| *ACRWb* | Previous area of wide cracks | 0–40 % |

Table 2: Factors Investigated and Their Ranges (Source: Mrawira 1996)

| Input factor | Estimated coefficient, $\hat{\beta}$ | Standard error | $t$ statistic | $p$ value |
|---|---|---|---|---|
| $Krp$ | -15.65 | 0.75 | -20.88 | 0.0000 |
| $W$ | -8.12 | 0.75 | -10.79 | 0.0000 |
| $RDM$ | -4.27 | 1.13 | -3.78 | 0.0002 |
| $ACRW$ | -3.97 | 1.32 | -3.02 | 0.0027 |
| $QI$ | 3.73 | 0.75 | 4.95 | 0.0000 |
| $APOT$ | -3.43 | 0.75 | -4.59 | 0.0000 |
| $RDS$ | -2.91 | 0.97 | -2.99 | 0.0029 |
| $RF$ | 2.85 | 0.75 | 3.80 | 0.0002 |
| $Kci$ | 2.56 | 0.75 | 3.41 | 0.0007 |

Table 3: Factors Significant at the 1% Level in the Linear Regression Model (1), Ordered by the Magnitudes of Their Estimated Coefficients

| Input factor(s) | Contribution to predictor variability (%) |
|---|---|
| *Krp* | 42.6 |
| *W* | 13.5 |
| (*RDM*, *RDS*) | 6.7 |
| *QI* | 5.6 |
| *Krp*× (*RDM*, *RDS*) | 4.6 |
| (*ACRA*, *ACRW*, *APOT*, *ARAV*) | 3.3 |
| *SN*× *QI* | 2.9 |
| *Krp*× *QI* | 2.5 |
| *Kci* | 2.5 |
| *RF* | 1.8 |
| *Kgp*× *QI* | 1.2 |

Table 4: Factors or Interactions Accounting for at Least 1% of the Variablility in the NPV Predictor From the Stochastic-Process Model (2); Factors in Parentheses are Considered Together as a Group; $A \times B$ Denotes the Interaction Between Factors $A$ and $B$

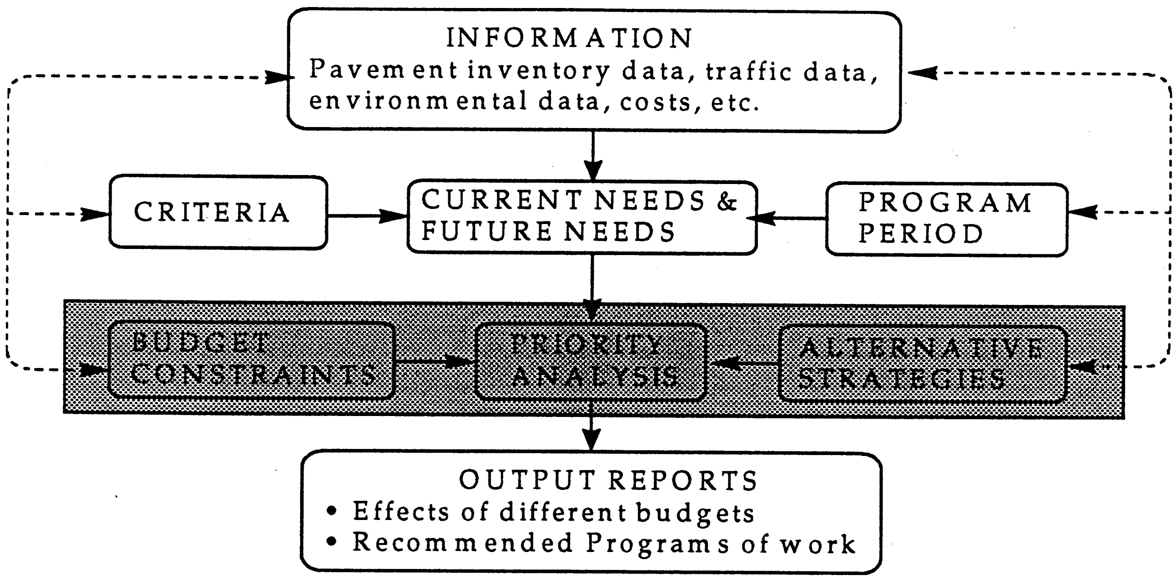


Figure 1: Major Steps in Priority Programming (Adapted From Haas et al, 1985)

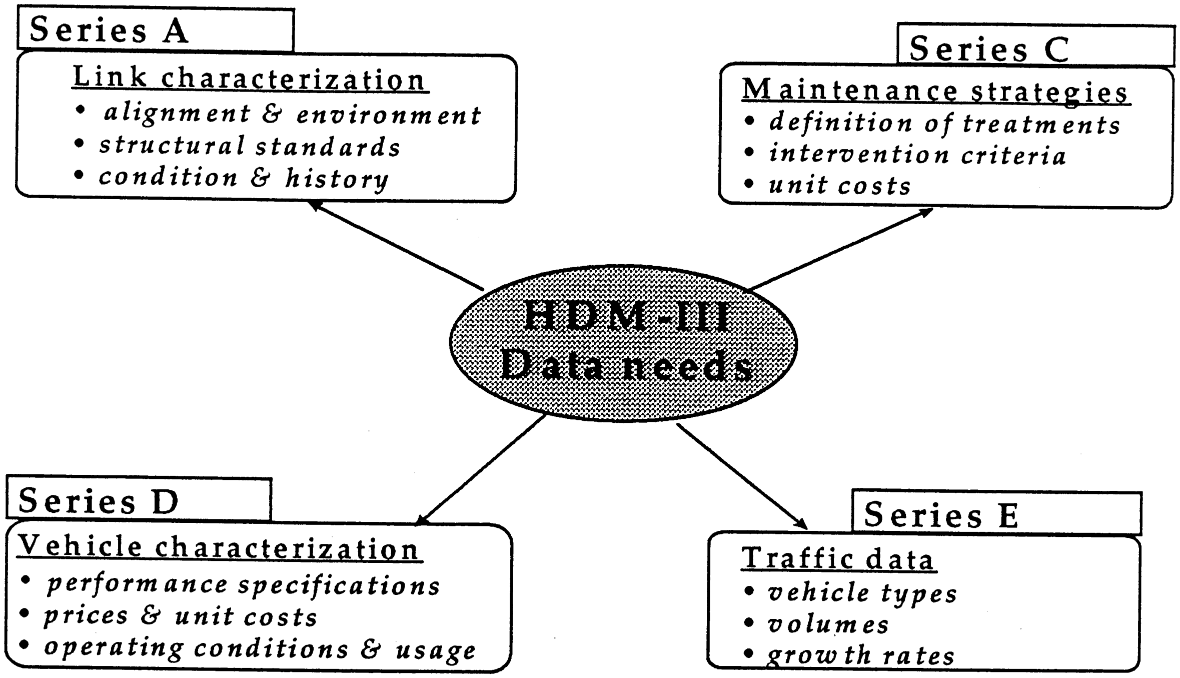


Figure 2: HDM-III Data Requirements for R&M Priority Programming

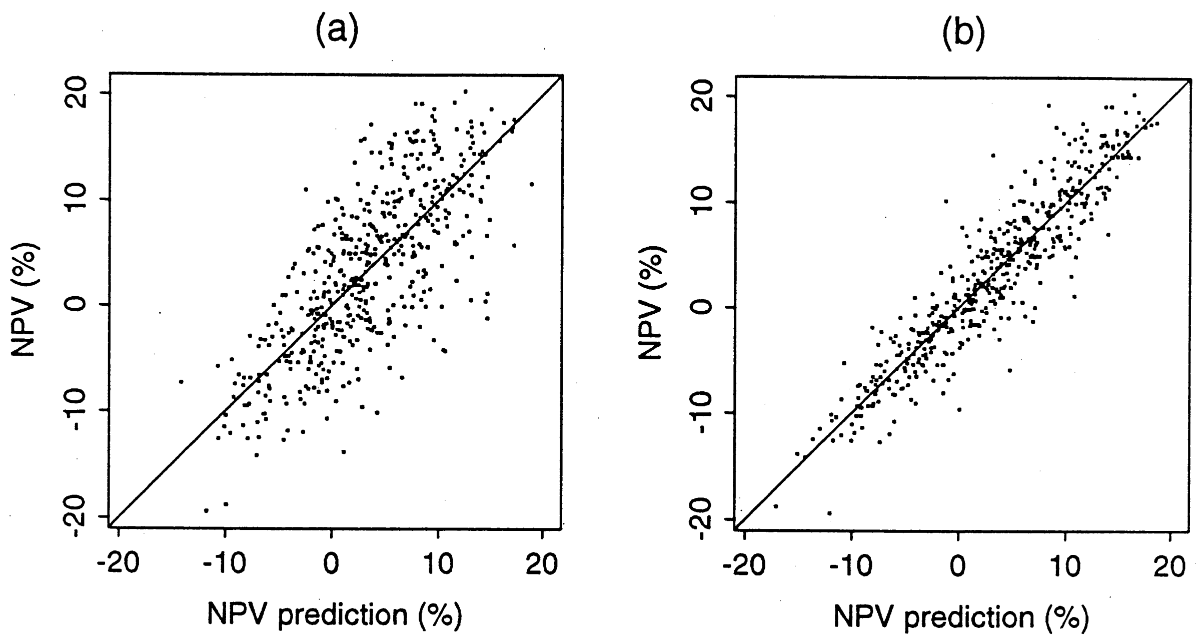


Figure 3: Actual NPV Values From HDM-III Versus Their Cross-Validation Predictions From (a) the Linear Regression Model and (b) the Stochastic-Process Model

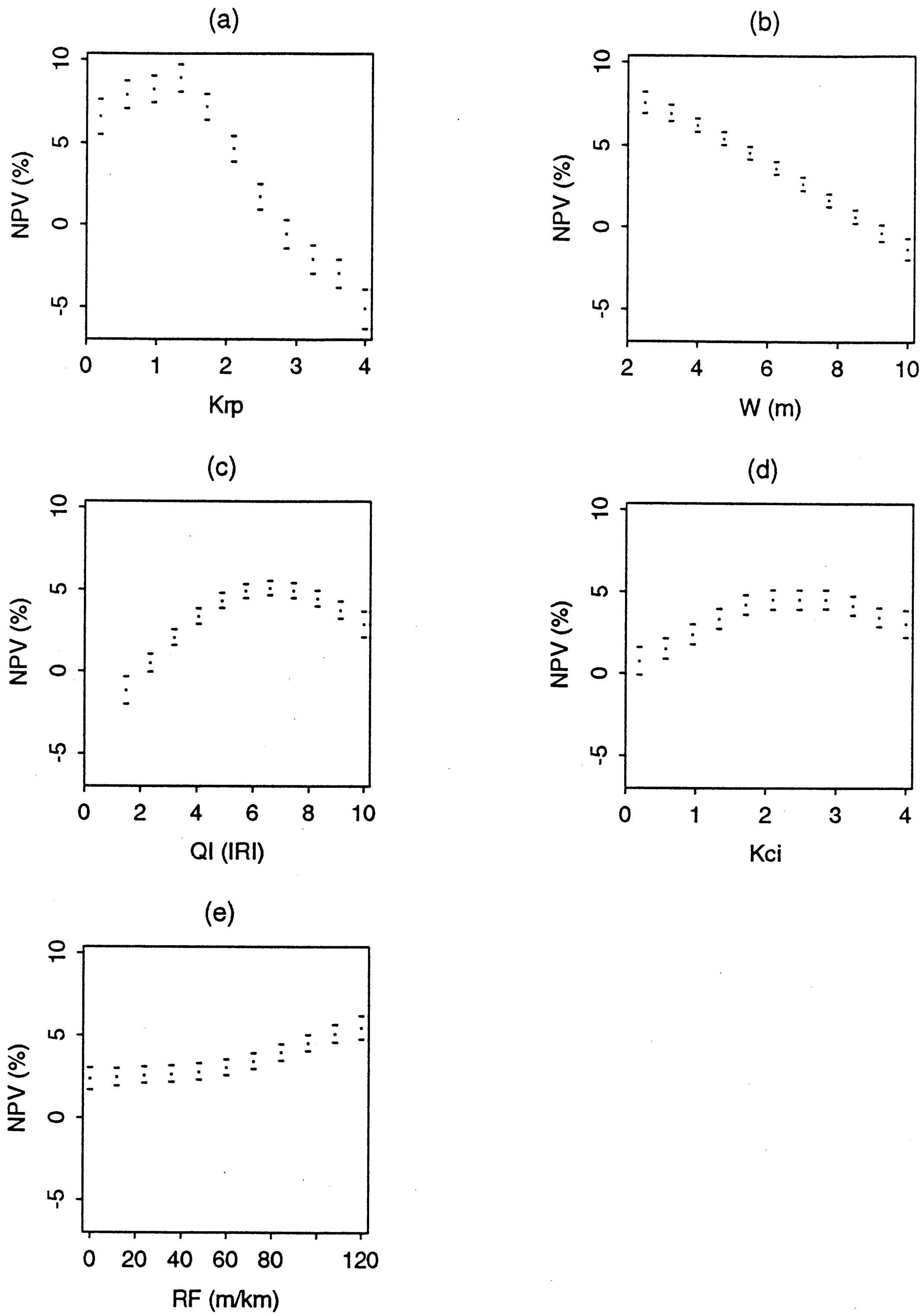


Figure 4: Estimated Main Effects (Central Dots) of (a) *Krp*, (b) *W*, (c) *QI*, (d) *Kci*, and (e) *RF* on NPV; the Bounds Are Approximate Pointwise 95% Confidence Intervals for the True Effect

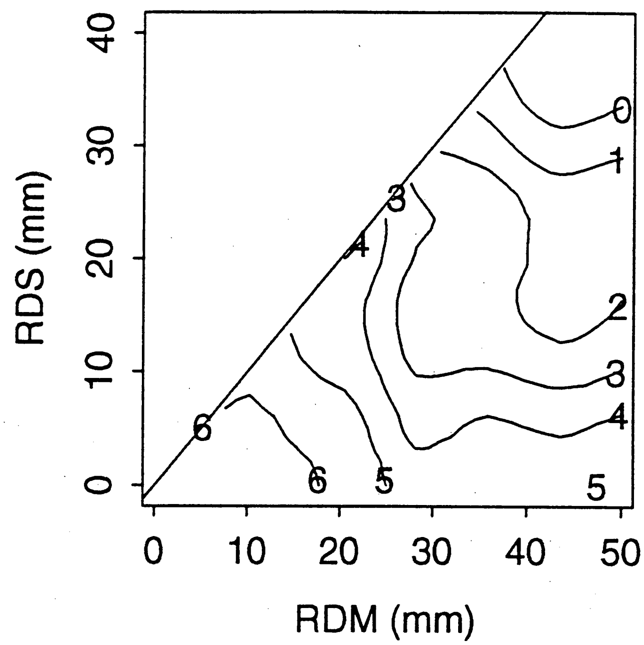


Figure 5: Estimated Joint Effect of *RDM* and *RDS* on NPV (%)

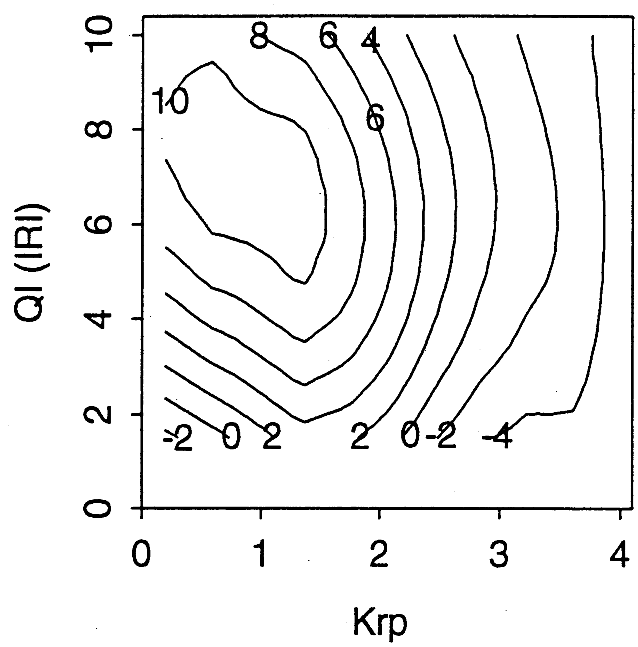


Figure 6: Estimated Joint Effect of *Krp* and *QI* on NPV (%)

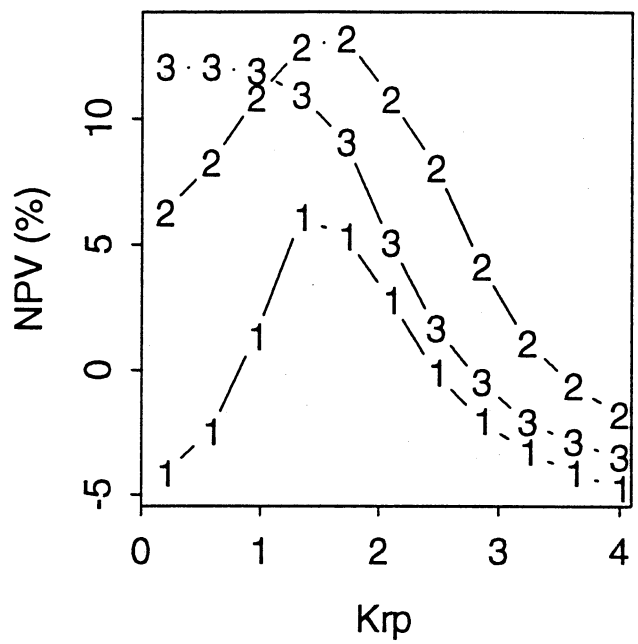


Figure 7: NPV From the HDM-III Model Versus *Krp* at *QI* Values of 1.5, 5.75, and 10 IRI, Denoted by Plotting Symbols 1, 2, and 3, Respectively