# Appropriate Models In Decision Support Systems For River Basin Management

*Yueping Xu, Martijn J. Booij*
*University of Twente*
*Enschede, the Netherlands*


## *Abstract*

*In recent years, new ideas and techniques appear very quickly, like sustainability, adaptive management, Geographic Information System, Remote Sensing and participations of new stakeholders, which contribute a lot to the development of decision support systems in river basin management. However, the role of models still needs to be emphasized, especially for model-based decision support systems. This paper aims to find appropriate models for decision support systems. An appropriate system is defined as 'the system can produce final outputs which enable the decision makers to distinguish different river engineering measures according to the current problem'. An appropriateness framework is proposed mainly based on uncertainty and sensitivity analysis. A flood risk model is used, as a part of the Dutch River Meuse DSS to investigate whether the appropriate framework works. The results showed that the proposed approach is applicable and helpful to find appropriate models.*


*Key words Appropriate models; Latin Hypercube Simulation; Morris' method; Flood risk model; River Meuse.*

## Introduction

In recent years, there is increasing interest in the development of decision support systems (DSSs) for river basin management. On the one hand, new ideas and techniques appear very quickly, like sustainability, adaptive management, Geographic Information System (GIS), Remote Sensing (RS) and participations of new stakeholders (Smits et al. 2000), which contribute a lot to the development of DSSs. On the other hand, the role of models in DSSs has been emphasized (Wunderlich 1989; Jolma et al. 1997). A decision support system for river basin management often encompasses a number of sub-models, including models for flood risk, ecology, tourism, recreation and navigation. These models are fundamental in supporting the whole decision-making process. However, often very complicated and sophisticated models are used which are difficult to understand and operate for decision makers. These models need extensive computation time and are difficult to interpret and maintain. They are discouraging to decision makers seeking support. Moreover complex models are often not suitable, for example for preliminary planning purpose DSSs or insufficient data. Sometimes simple models are more appropriate than complex ones (Perrin et al. 2001; Vreugdenhil 2002). That is why the idea of appropriateness is suggested in this paper.

According to European Commission (2000), "Decision makers need to be aware of the degree of uncertainty attached to the results of the evaluation of the available scientific information." The main reason for considering uncertainty is to avoid wrong decisions associated with large losses. While uncertainty in decision - making process is undesirable, it can still be useful in river basin management as it provides a basis for the need for further investigation and research to improve the scientific basis for decisions. In this paper, an approach is proposed combining the idea of appropriateness and uncertainty. This approach is used to determine whether the models in the DSSs are appropriate by investigating the role of uncertainty in the model outputs and how the uncertainty affects the ranking of the measures. As we know, the difficulty of ranking measures is caused by the high uncertainty in the models that produce a lot of overlap between different outputs. An appropriate system is thus defined as 'the system can produce final outputs which enable the decision makers to distinguish or rank different river engineering measures under uncertainty'. A flood risk model, as a part of the Dutch River Meuse DSS will be used in this preliminary study to see whether the proposed approach works.

## Appropriateness framework

Fig. 1 shows the proposed appropriateness framework used in this paper. Suppose there are *k* possible measures $\Theta \equiv \{\theta_1, \theta_2, ..., \theta_k\}$ that we want to evaluate, where $\theta$ represents different measures. Assume a performance criterion $L(\theta)$ representing average performance in the DSSs. Fig. 1 shows that after quantitative modelling, uncertainty analysis will be done for each of the *k* measures, which produces noisy outputs of the performance criterion. We compare measures on the performance criterion using multiple comparison methods. If the measures are distinguishable (it is possible to rank the measures), the models are considered appropriate. When the models are inappropriate (indistinguishable measures), sensitivity analysis will be used to determine the most important inputs and parameters in the models. After that, uncertainty will be reduced with respect to these most important inputs and parameters to obtain appropriate models.

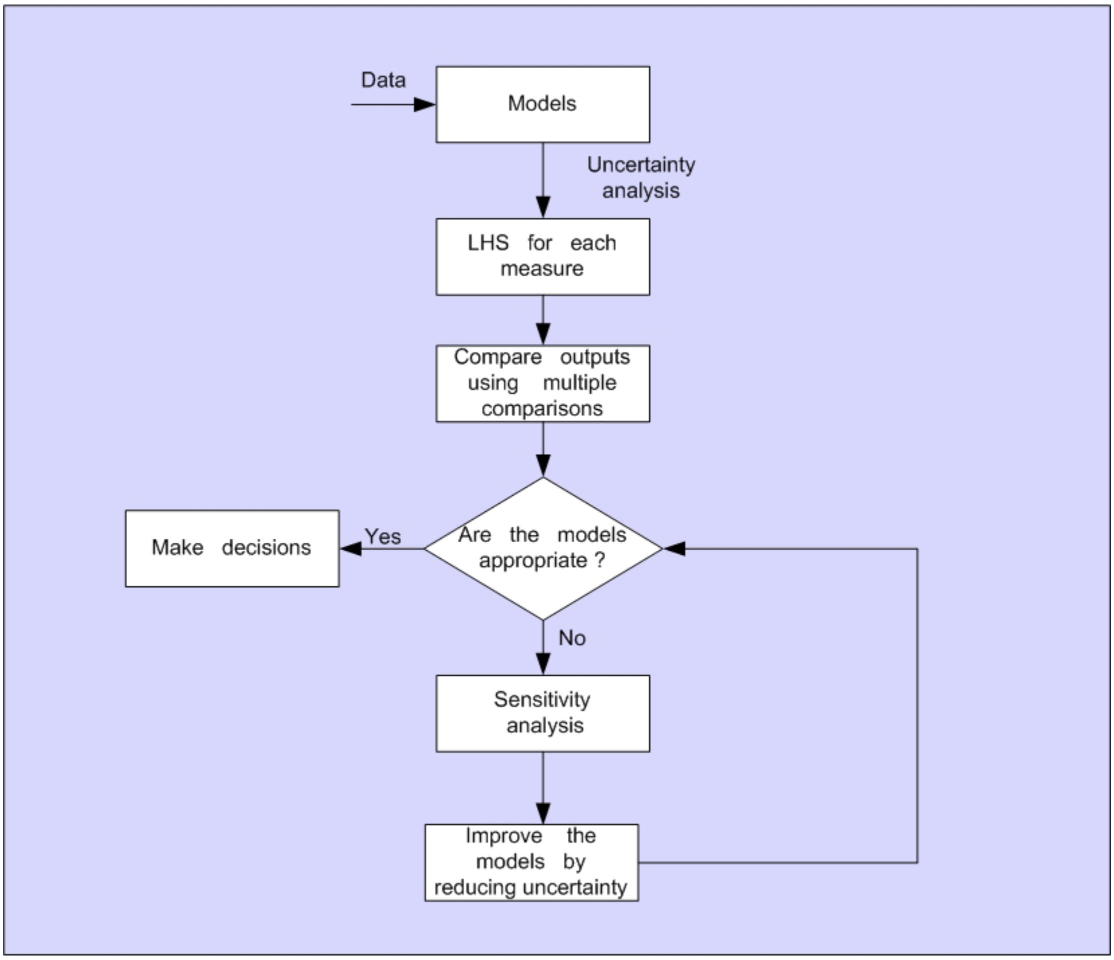


**Figure 1** The appropriateness framework (LHS represents Latin Hypercube Simulation)

In this approach, there are three important aspects which need to be described in detail in this section. They are uncertainty analysis, multiple comparison procedures and sensitivity analysis.

**Uncertainty analysis**

As shown in Fig. 1, uncertainty analysis is the step after the quantitative modelling. According to Morgan & Henrion (1990), there are several types of uncertainty. In this preliminary study only uncertainty in input data and parameters will be considered. The uncertainty caused by the model structure will not be studied although it has been known to be important (Devooght 1998; Perrin et al. 2001). The Latin Hypercube Simulation (LHS) method will be used to investigate the propagation of uncertainty in the inputs and parameters to the model outputs (Saltelli et al. 2000).

The Latin Hypercube Simulation is a stratified sampling method that efficiently estimates the uncertainty in the simulation results. The probability distribution of each input or parameter is subdivided into *N* intervals with an equal probability. The models in the DSSs will be run *N* times with a random combination of input or parameter from each interval. In this way, low probability outcomes are also well represented.

### Multiple comparison procedures

Multiple comparison procedures are used to compare the simulation results from different river engineering measures. These comparisons may involve group means, medians and variances. The multiple comparison methods include the Scheffe method, the Tukey-Kramer method, the Fisher-Hayer method and the GH procedure (Toothaker 1993; Rafter et al. 2002). Most available multiple comparison methods are only used to compare the means for a performance criterion with a consideration of data variability.

Multiple comparison procedures are generally completed after the One-way Analysis of Variance (ANOVA) results. ANOVA is used before multiple comparison procedures to check if there are differences among the simulation results. The reason is that ANOVA is more robust and reliable than multiple comparisons. However, ANOVA can only show whether there are significant differences among measures and it cannot tell where the differences exist. So after ANOVA, multiple comparisons can be used to check where the differences exist. The ANOVA test assumes that all simulation results are normally distributed, all simulation results have equal variances, and all simulation results are mutually independent. The ANOVA test is known to be robust to modest violations of the first two assumptions (Rafter et al. 2002).

The null hypothesis of multiple comparisons is: the means of the model outputs from all measures in $\Theta \equiv \{\theta_1, \theta_2, ..., \theta_k\}$ are effectively the same. The standard way to compare means of a certain performance criterion is to calculate multiple acceptance intervals for *k (k-1)/2* differences $\delta_{ij} = \overline{L}_i - \overline{L}_j$. A confidence interval is formed using a point estimate (Toothaker 1993), which is the best guess for $\delta_{ij}$, based on the LHS results. The margin of error reflects the accuracy of the guess based on the variability of the data. If the interval calculated does not contain zero, then the null hypothesis is rejected and $\overline{L}_i$ and $\overline{L}_j$ are declared different at given levels of significance.

### Sensitivity analysis

Sensitivity analysis (SA) can be conducted to determine whether a model resembles the system, the inputs and parameters that most contribute to the model output variability and if and which (group of) inputs and parameters interact with each other (Saltelli et al. 2000). According to Saltelli, there are three types of SA: local, screening and global. Local SA is used where the emphasis is on the local impacts, a screening method is used to identify the most important inputs or parameters and global SA is used to apportion the uncertainty in the simulation results to the input and parameter uncertainty. In this paper we use a screening SA method — the Morris' Method to determine the most important inputs and parameters in the models because of its global characteristics and efficiency.

The basic idea of the Morris' method is to determine, within reasonable uncertainty, which input or parameter may be considered to have effects which are (a) negligible, (b) linear and additive, and (c) nonlinear or involved in interactions with other inputs. This method uses the concept of elementary effect $F_i$ for each input or parameter. The sample mean $E_i$ and variance $S_i^2$ of the observed elementary effects for input *i* are unbiased estimators of the mean and variance of $F_i$. They are used to determine the effects of each input and parameter (Saltelli et al. 2000).

The order of importance of the inputs and parameters will be determined by calculating the Morris distances, which are the Euclidean distances from (*Ei*, $S_i^2$) to the origin (0, 0). The idea in this paper is that by reducing the uncertainty in the most important inputs and parameters identified by the Morris' method, the system will be improved to get distinguishable results.

## Model description

Fig. 2 gives the system diagram for the flood risk model, which is a part of the developed DSS for the Dutch Meuse River. On the left-hand side of this figure are the inputs to the models and on the right-hand side of this figure is the expected annual damage (EAD). The EAD is regarded as one of the main performance criteria to compare the effects of different measures. There are several sub-models shown in Fig. 2, namely flood frequency model, hydraulic model, inundation model, damage model and risk model. They will be described briefly in this section.

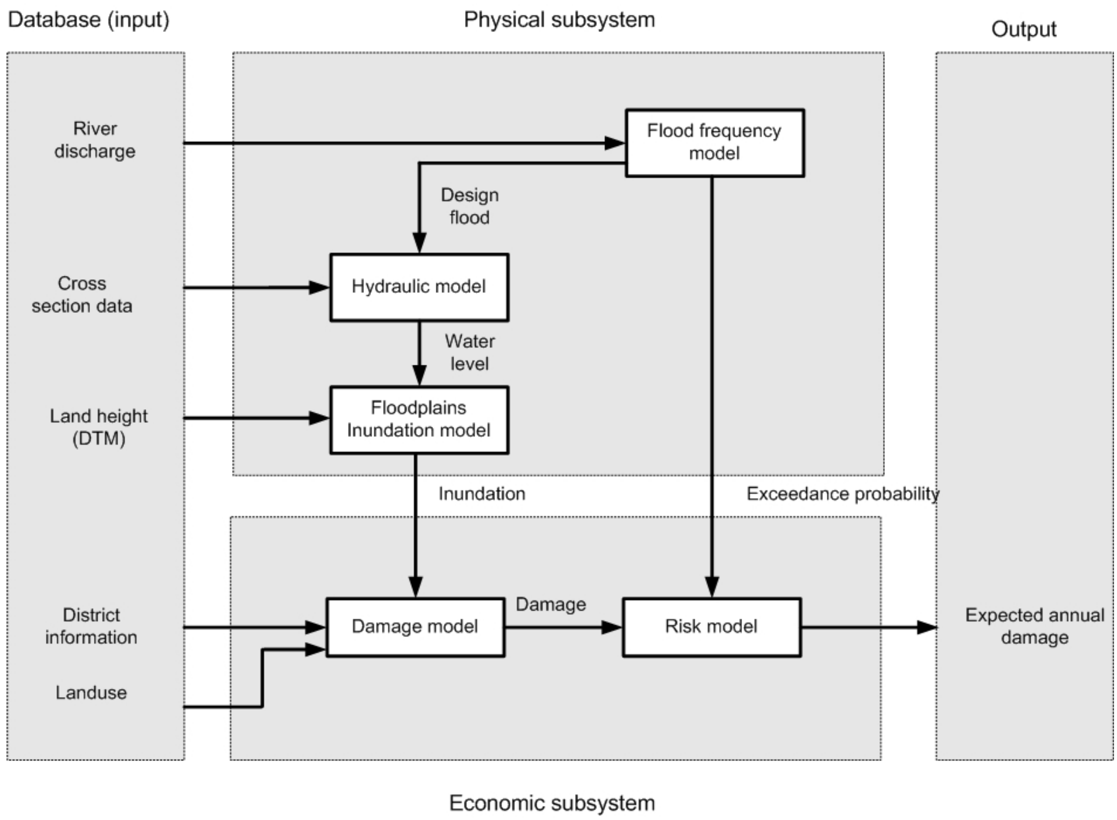


**Figure 2** A system diagram for the flood risk model

**Flood frequency model**

Flood frequency analysis is an essential part of the risk model. The primary objective of flood frequency analysis is to relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions (Chow et al.1988). In flood frequency analysis, an important aspect is to choose a certain distribution model that will be used to describe the flood flows. In this paper, the Gumbel Extreme Value distribution is used (Chow et al.1988).

**Hydraulic model**

The hydraulic model is used to calculate water levels in the river channel for certain flood flows. The measurements of water levels in the Dutch Meuse River indicate a steady and even uniform flow for different discharges (Rijkswaterstaat, afd. ANW 1998). Stepwise steady non-uniform flow simulation is used here for flood routing (Van Rijn 1994). In this model, we assume there are no lateral flows.

**Inundation model**

The inundation model is used to calculate the inundation depths in the flood plains. The inundation depths are the differences between water levels and land heights. Here the assumption is that water levels are the same from west to east (approximately perpendicular to the river) in the floodplains.

**Flood damage model**

The damage model is developed to assess the damage for floods of different probabilities. The economic damage in the floodplains is determined by the inundation depths, land use types and the number of units of that land use type. The damage is given in monetary values per unit (in euro).

**Risk model**

The objective of the flood risk model is to calculate the expected annual damage (EAD) for each measure. For floods of different probabilities, corresponding values of flood damage can be calculated based on the flood damage model. The EAD is the expected annual value of these damages.

## Results

In this example, four measures will be used beside the base situation. For convenience, the base situation is indicated as 'Measure 1'. Measure 2 is to broaden the summer bed by 25 meters. Measure 3 is to deepen the summer bed by 1 meter. Measure 4 is the construction of embankments along the river and Measure 5 is the spatial planning measure in the floodplains.

### LHS results

As mentioned before, model uncertainty will not be considered in this study. Only the uncertainty in the inputs and parameters is considered. In this example, there are totally 112 inputs and parameters in the models. For the hydraulic parameters, a questionnaire has been completed to investigate how uncertain these parameters are. The parameters in the flood frequency model are assumed to be normally distributed. The distributions of other inputs and parameters are set uniform in shape, because there are insufficient data to infer any particular type of distribution. For these inputs and parameters, ranges of variability have been selected either according to the information available, or, in absence of such information, assuming 20% percent of uncertainty involved in inputs and parameters (nominal value ).

The Lilliefors test for goodness of fit is used to check whether the simulation results from each measure follow a normal distribution (Conover 1980). According to this test, the simulation results from the five measures are log-normally distributed. For the convenience of comparison, the natural logarithm of data (normally distributed) will be used. The fitted normal distributions are shown in Fig. 3. This figure shows that there are large areas of overlap between the measures, especially for Measure 3 and Measure 5. The means of these two measures are so close that they may produce an undistinguishable situation.

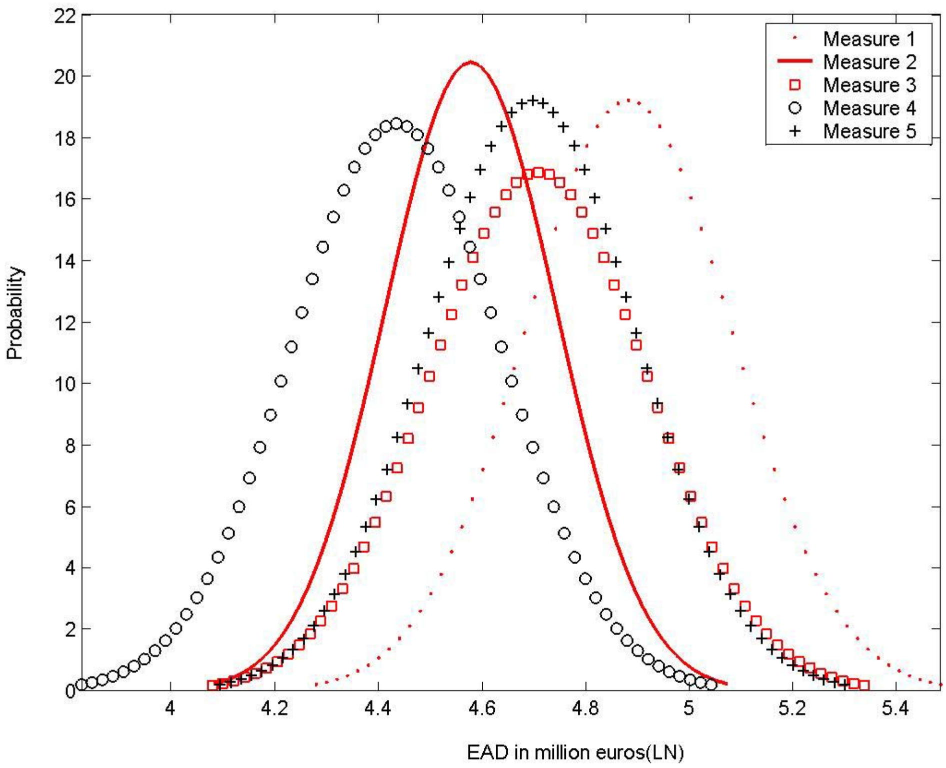


**Figure 3** The fitted normal distributions for five LHS results

### Multiple comparison results

The ANOVA test tells that the differences between the means are highly significant. The test therefore strongly supports the alternative hypothesis, that one or more of the simulation results are drawn from populations with different means. Fig. 4 shows the box plots for the five simulation results, which gives a qualitative picture of the differences. The box has lines at the lower quartile, median, and upper quartile values (95%). The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers.

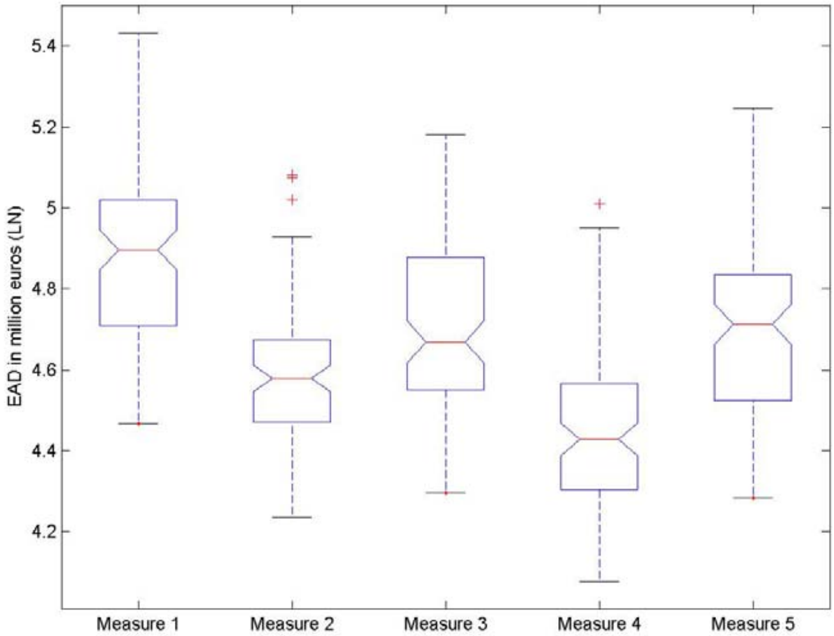


**Figure 4** The box plots for five LHS results

In order to see where the differences exist, one of the multiple comparisons methods, the Tukey-Kramer method, is used after the ANOVA test (Toothaker 1993). Fig. 5 shows the Tukey-Kramer results. The X-axis is the 95% confidence intervals for the means of five LHS results and the Y-axis are the measures. This figure shows that except Measure 3 and Measure 5, all the other three measures are significantly different from each other. The significance level used is 0.05.

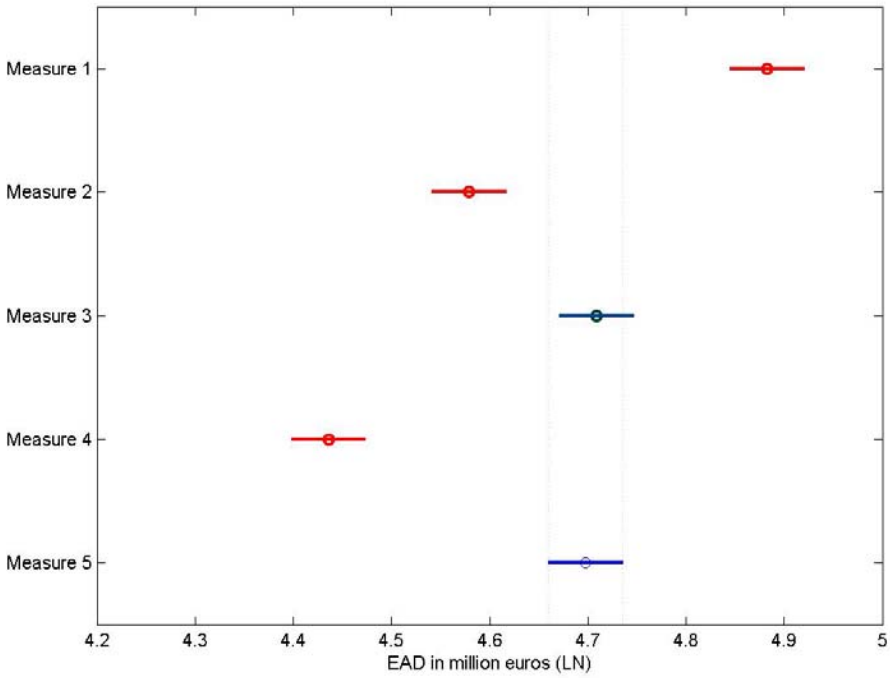


**Figure 5** The Tukey- Kramer results before model improvement

Table 1 gives the 95% confidence intervals for the differences of the means for the five simulation results. There are totally 5*(5-1)/2=10 pairs of differences in this table. Each row of this table represents one test, and there is one row for each pair of measures. The entries in the row indicate the means being compared, the estimated difference in means, and a confidence interval for the difference.

**Table 1** Confidence intervals for the means (million euros in natural logarithm).

| Row No. | Measure No. | Measure No. | Lower limit | Mean of the difference of means | Upper limit |
|---|---|---|---|---|---|
| 1 | 1 | 2 | 0.23 | 0.30 | 0.38 |
| 2 | 1 | 3 | 0.10 | 0.17 | 0.25 |
| 3 | 1 | 4 | 0.37 | 0.45 | 0.52 |
| 4 | 1 | 5 | 0.11 | 0.18 | 0.26 |
| 5 | 2 | 3 | -0.21 | -0.13 | -0.05 |
| 6 | 2 | 4 | 0.07 | 0.14 | 0.22 |
| 7 | 2 | 5 | -0.19 | -0.12 | -0.04 |
| 8 | 3 | 4 | 0.20 | 0.27 | 0.35 |
| 9 | 3 | 5 | -0.06 | 0.01 | 0.09 |
| 10 | 4 | 5 | -0.34 | -0.26 | -0.19 |

For example, take the sixth row that contains [2 4 0.07 0.14 0.22]. These numbers indicate that the mean of Measure 2 minus the mean of Measure 4 is estimated to be 0.14, and a 95% confidence interval for the true mean of the differences of means is [0.07, 0.22]. In this example, the confidence interval does not contain 0.0. Therefore, the difference is significant at the 0.05 significance level.

Another example is the ninth row that contains [3 5 -0.06 0.01 0.09]. These numbers show that the mean of the difference of Measure 3 and Measure 5 is estimated to be 0.01, and a 95% confidence interval for the mean of the differences of means is [-0.06, 0.09]. This confidence interval does contain 0.0, which means the difference between Measure 3 and Measure 5 is not significant at the 0.05 significance levels.

Table 1 show that nine out of ten pairs of the confidence intervals don’t contain the value of zero. Only the confidence interval in the ninth row (the pair of Measure 3 and Measure 5) contains zero and indicates that these two measures are not significantly different at the 0.05 significance level. This means Measure 3 and Measure 5 are indistinguishable according to the appropriateness definition.

**Appropriateness analysis and improvements**

From the multiple comparison results, we know that not all measures are distinguishable. Measure 3 and Measure 5 are indistinguishable. This situation means that it is difficult to rank the measures due to the high uncertainty involved in the LHS results (see Fig. 3). Therefore we determine that the models used are inappropriate in this example because of the failure to distinguish the measures.

In order to improve the models, the Morris’ method is used to investigate the order of importance of all inputs and parameters in the models. The calculated Morris distances for inputs and parameters are presented in Table 2 (order of magnitude). The subscripts g and z in this table represent two river sections. The order of importance based on the Morris distances is shown in the last column of Table 2. The Morris results show that the most important 14 inputs and parameters are related to the hydraulic model. The uncertainty in these inputs and parameters is reduced by assuming these 14 most important inputs and parameters in the hydraulic model are deterministic.

**Table 2** Different groups of the Morris distances (order of magnitude).

| Inputs and parameters | Descriptions of inputs and parameters | Morris distances | Order of importance |
|---|---|---|---|
| $i_z$ | Slope of the river | $10^{14}$ | 1 |
| $i_g$ | Slope of the river | $10^{13}$ | 2 |
| $K_z$<br>$K_g$ | Nikuradses for the winter bed (floodplains) | $10^9$ | 3<br>4 |
| $bp_z$<br>$h_g$<br>$h_z$<br>$bp_g$ | Bed level coefficient<br>Depths of the summer bed<br>Depths of the summer bed<br>Bed level coefficient | $10^8$ | 5<br>6<br>7<br>8 |
| $B_{g1}$<br>$C_z$<br>$B_{g2}$<br>$B_{z2}$<br>$C_g$<br>$B_{z1}$ | Width of the summer bed<br>Chezy coefficient for the summer bed<br>Width of the winter bed<br>Width of the winter bed<br>Chezy coefficient for the summer bed<br>Width of the summer bed for Zandmaas | $10^7$ | 9<br>10<br>11<br>12<br>13<br>14 |
| Mu<br>Others | Sample mean in flood frequency model<br>Other coefficients | $<10^7$ | 15<br>16~112 |

The results of the Tukey-Kramer method based on the improved models are shown in Table 3. Smaller confidence intervals for all pairs of measures can be observed compared to Table 1. For example, for the pair of Measure 1 and Measure 5, the confidence interval changes from (0.11, 0.26) in Table 2 to (0.12, 0.20) in Table 3. At the same time, the means of the differences also decrease with the uncertainty reduction, except Measure 3 and Measure 5. This is possibly due to the non-linearity of the models and insufficient simulation runs. In Table 3, we also see that all the confidence intervals in the rows don't contain the value of zero, even for Measure 3 and Measure 5. Fig. 6 shows the 95% confidence intervals for the means of the LHS results for the five measures based on the improved models. A bigger difference between Measure 3 and Measure 5 can be observed. Based on this result, we determine that all the measures are significantly different from each other.

**Table 3** Confidence intervals for the means (million euros in natural logarithm).

| Row No. | Measure No. | Measure No. | Lower limit | Mean of the difference of means | Upper limit |
|---|---|---|---|---|---|
| 1 | 1 | 2 | 0.16 | 0.23 | 0.30 |
| 2 | 1 | 3 | 0.05 | 0.11 | 0.18 |
| 3 | 1 | 4 | 0.28 | 0.35 | 0.42 |
| 4 | 1 | 5 | 0.12 | 0.18 | 0.20 |
| 5 | 2 | 3 | -0.18 | -0.11 | -0.05 |
| 6 | 2 | 4 | 0.05 | 0.12 | 0.19 |
| 7 | 2 | 5 | -0.17 | -0.10 | -0.02 |
| 8 | 3 | 4 | 0.17 | 0.24 | 0.30 |
| 9 | 3 | 5 | 0.01 | 0.07 | 0.14 |
| 10 | 4 | 5 | -0.23 | -0.16 | -0.10 |

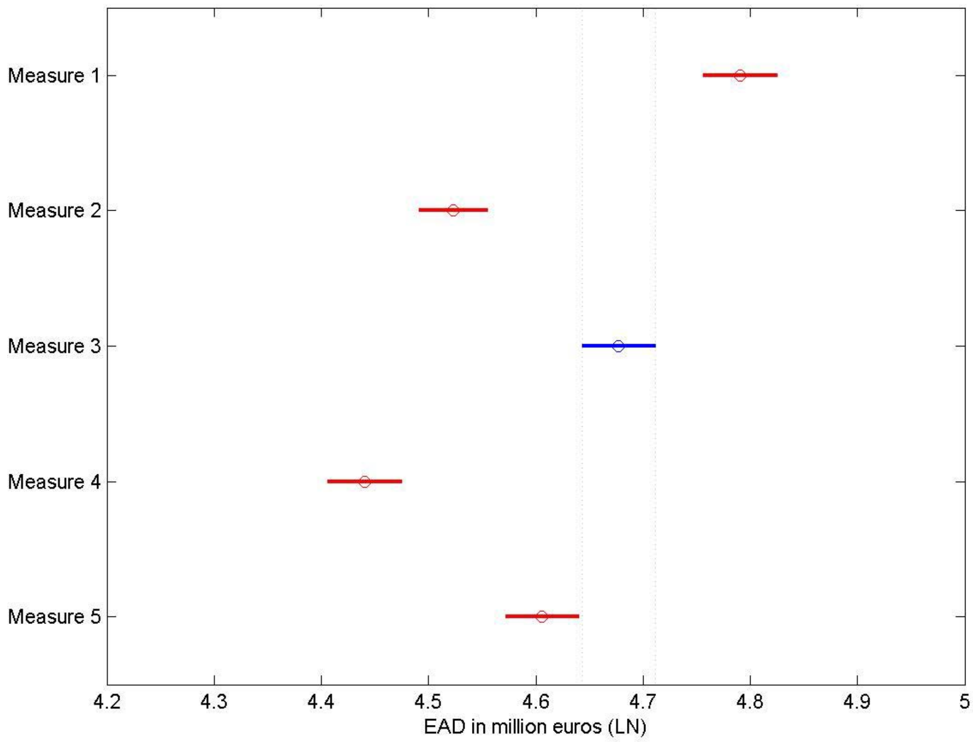


**Figure 6** The Tukey- Kramer results after model improvement

The performance criterion in this example is the EAD, which means that the higher EAD, the worse the measure. Therefore the following ranking of measures can be obtained:

Measure 1<Measure 3<Measure 5<Measure 2<Measure 4

We say the models are appropriate according to the definition of appropriateness.

## Conclusions

This paper proposed a framework for analyzing the appropriateness of the models in a DSS, using a flood risk model as an example. The point here is how to judge whether the models are appropriate with high uncertainties in the simulation results and how the reduction of uncertainty can help rank the measures.

In this paper, the comparisons of different measures are based on the means of the simulation results with a consideration of data variability. This will be satisfactory if the decision makers are only interested in the mean values. However, sometimes the decision makers are more interested in the risk of making a wrong decision. Then not only the means will be of interest but the uncertainty will also play an important role in making a sound decision than the example shown in this paper (Xu and Booij 2004).

In general reduction of the uncertainty should be expected to reduce the dispersion in the distribution, which makes it easier to distinguish the measures. In this paper, the reduction of uncertainty has been completed through assuming that the most important inputs and parameters were deterministic. However, in reality, it doesn't work in this way. In a future study, more realistic actions will be taken to reduce the uncertainty in the model structure or the uncertainty in inputs and parameters. Moreover, uncertainty reductions come at a cost. In case uncertainty reductions are quite expensive and still two or more pairs of measures are indistinguishable, it may be not worthwhile to improve the models any more. In this situation, the indistinguishable measures can be regarded to have the same effects on the performance criterion.

***Reference***

***Chow, V.T., Maidment, D. R. and Mays, L.W. (1988)*** *Handbook of Applied Hydrology, McGraw-Hill, New York.*

***Conover, W. J. (1980)*** *Practical Nonparametric Statistics. Wiley, New York.*

***Devooght, J. (1998)*** *Model uncertainty and model inaccuracy. In: Reliab. Eng. Syst. Safe. Vol. 59, 171-185.*

***European Commission, 2000*** *Communication on the precautionary principle, COM (200) 1.*

***Jolma, A, Marchi, C.D., Smith, M., Perera, B.J.C. and Somlyody, L. (1997)*** *StreamPlan: a support system for water quality management on a river basin scale. In: Environ. Modeling & Software, Vol. 12, 275-284.*

***Morgan, M.G., and Henrion, M. (1990)*** *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis, Cambridge Univ. Press, Cambridge, U.K.*

***Perrin, C., Michel, C. and Andréassian, V. (2001)*** *Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. In: J. of Hydrol. Vol. 242, 275-301.*

***Rafter, J.A., Abell, M.L. and Braselton, J.P. (2002)*** *Multiple comparisons methods for means. In: SIAM REV. Vol. 44, No. 2, 259-278.*

***Rijkswaterstaat, afd. ANW (1998)*** *Measuring data, reference station Borgharen.*

***Saltelli, A, Chan, K. and Scott, E.M. (2000)*** *Sensitivity analysis. John Wiley & Sons Ltd. England.*

***Smits, A.J.M., Nienhuis, P.H. and Leuven, R.S.E.W. (2000)*** *New approaches to river management. Backhuys Publishers, the Netherlands.*

***Toothaker, L.E. (1993)*** *Multiple Comparison Procedures. Series: Quantitative Applications in the Social Sciences; 89. SAGE Publications, London.*

***Van Rijn, L.C. (1994)*** *Principles of fluid flow and surface waves in rivers, estuaries, seas and oceans. AQUA publications, the Netherlands.*

***Vreugdenhil, C.B. (2002)*** *Accuracy and reliability of numerical river models. In: J. of Am. Water Resour. Assoc. Vol. 38. No. 4. 1083-1095.*

***Wunderlich, W.O. (1989)*** *Experiences with water management model development and use. In: Closing the gap between theory and practice (ed. by D.P. Loucks). IAHS Publication No.180. Galliard Ltd. Great Britain.*

***Xu, Y.P. and Booij, M.J. (2004)*** *Appropriate modeling in DSSs for river basin management. In: Proc. Second Biennial Meeting of the International Environmental Modelling and Software Society, Osnabrück, Germany. paper submitted.*