Hydrological Sciences for Managing Water Resources in the Asian Developing World (Proceedings of a Symposium in Guangzhou, China, June 2006). IAHS Publ. 319, 2008.

Appropriate spatial scales to achieve model output uncertainty goals

MARTIJN J. BOOIJ[1] & CHARLES S. MELCHING[2]

1 *Water Engineering and Management, Faculty of Engineering Technology, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands*
m.j.booij@utwente.nl

2 *Department of Civil and Environmental Engineering, Marquette University, PO Box 1881, Milwaukee, Wisconsin 53201, USA*

Abstract Appropriate spatial scales of hydrological variables were determined using an existing methodology based on a balance in uncertainties from model inputs and parameters extended with a criterion based on a maximum model output uncertainty. The original methodology uses different relationships between scales and variable statistics. It is extended with two different uncertainty propagation methods, the mean-value first-order second-moment (MFOSM) method and Monte Carlo analysis, and backward uncertainty propagation to obtain appropriate scales based on two uncertainty criteria. The methodology is applied to three flood estimation methods. The application to the River Meuse basin in western Europe revealed that the methodology can be used for the considered flood estimation methods under similar climatological and geographical conditions. The results showed different relative input and parameter uncertainties for the different flood estimation methods (3–6%) for a specific maximum output uncertainty (25%) and different appropriate spatial scales for the dominant variables.

Key words appropriate scale; backward uncertainty propagation; flood estimation; HBV model; MFOSM method; Monte Carlo analysis; SCS method; unit hydrograph

INTRODUCTION

A broad palette of models is available for river flood estimation, ranging from simple, lumped empirical models to complex, distributed models which include lots of physics and mathematics. The complexity of models depends on the processes incorporated, the process formulations used, and the different space and time scales applied. In general, models should be sufficiently detailed to capture the dominant processes and natural variability, but not unnecessarily refined such that computation time is wasted, or that the model requirements exceed available data. The optimum model complexity depends on the objectives and area of study and results in a so-called appropriate model. Booij (2002a) has developed a model appropriateness procedure in which dominant processes, appropriate scales, and associated appropriate process formulations are determined. Booij (2003) applied this procedure to determine the appropriate spatial scales of four dominant variables (elevation, soil type, land use, and precipitation) and integrate them into an appropriate model scale. An appropriate scale of a variable is intuitively defined as a scale which is sufficiently detailed to capture the variability of that variable, but not more than that. This definition was partly embedded in the criterion used for appropriate scales requiring a balance in uncertainties from inputs and parameters.

Results from Booij (2003) revealed appropriate spatial scales for elevation, soil, and land use of 0.1, 5.3, and 3.3 km, respectively, for modelling climate change impacts on flooding in the River Meuse in western Europe. The appropriate spatial scale for daily precipitation of about 20 km had been assessed in an earlier study (Booij, 2002b). This resulted in an appropriate spatial model scale of about 10 km. Several other studies show that the "appropriate" scales for elevation (e.g. Wolock & McCabe, 2000), soil moisture (e.g. Western *et al.*, 1998), and land use type (e.g. Moody & Woodcock, 1995) vary considerably, depending on geographical and climatological conditions, extent of the area, and support scale of the data, and can hardly be mutually compared. Furthermore, most studies assessed the appropriate scale of individual variables and did not integrate them into an appropriate model scale. This study extends and improves the methodology of Booij (2003) for the determination of appropriate spatial scales by including a criterion for the output uncertainty as well. Ideally, a model user would like to prescribe a specific maximum value for the output uncertainty to ensure that the model results for situations other than the current situation (past, future) are still reliable.

The objective of this paper is, therefore, to determine the appropriate spatial scales of the dominant variables based on a balance in uncertainties from inputs and parameters and a maximum output uncertainty. Dominant refers to the relative importance of variables with respect to river flooding in the River Meuse basin in France, Belgium, and The Netherlands. The objective is achieved by introducing an extended methodology for the determination of appropriate scales incorporating uncertainty criteria. This methodology is applied to three flood estimation methods using different uncertainty propagation methods, backward uncertainty propagation and different relationships between scales and variable statistics are used, and, finally, conclusions are drawn.

METHODOLOGY

Appropriate scales based on balanced input and output uncertainties

The hydrological model output Y (random) is a function of an n-dimensional independent random vector $\mathbf{X} = (x_1, x_2, \ldots, x_n)$, $Y = f(\mathbf{X})$, see left-hand side of Fig. 1. The functional form of this relationship depends on the problem, study area, and level of accuracy required, but in general conceptualizes the input–output relation of a hydrological system with different parameters. Examples range from simple conceptualizations such as the SCS curve number and dimensionless unit hydrograph methods (Maidment, 1992) to more complex ones embedded in hydrological models such as HBV (Bergström, 1995) and SHE (Abbot *et al.*, 1986). Often, the calculated output is subject to large uncertainties arising from natural uncertainty, data uncertainty, model parameter uncertainty, and model structure uncertainty. In the original approach for the determination of appropriate spatial scales of Booij (2003), only a criterion based on a balance in these uncertainty sources originating from dominant variables $U(\mathbf{X})$ has been used. These sources $U(\mathbf{X})$ were used to estimate the appropriate spatial scales for different dominant variables $\lambda(\mathbf{X})$ at a specific temporal

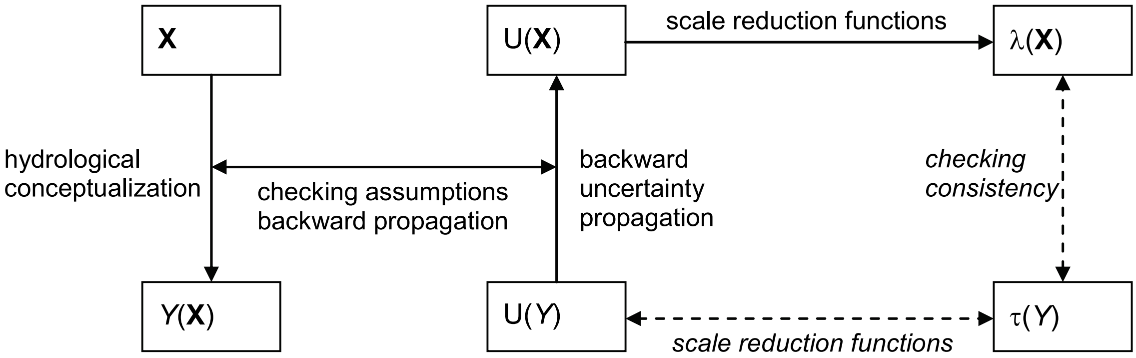

Fig. 1 Functional relationships between input **X** and output Y, related uncertainties $U(\mathbf{X})$ and $U(Y)$ and appropriate spatial scales of input $\lambda(\mathbf{X})$ and appropriate temporal scale of output $\tau(Y)$.

scale, see upper-right side of Fig. 1. This approach could result in arbitrarily large uncertainties in the model output $U(Y)$. Moreover, the temporal scale was somewhat arbitrarily chosen.

The first limitation is addressed in an extended approach for the determination of appropriate scales, which is illustrated in Fig. 1. The main improvement is the addition of a criterion prescribing a specific maximum value for the output uncertainty. This uncertainty $U(Y)$ is backward propagated to the uncertainty sources originating from dominant variables $U(\mathbf{X})$ and then, appropriate spatial scales are determined. Backward uncertainty propagation and its assumptions and the determination of appropriate scales using scale reduction functions are considered at the end of this section. First, three flood estimation methods are described and two methods for (forward) uncertainty propagation are considered.

The second limitation is not addressed here and an arbitrary, but reasonable, temporal scale of one day was chosen. Based on a prescribed maximum value for the output uncertainty $U(Y)$, the appropriate temporal scale for the output $\tau(Y)$ can be assessed in a similar way to the appropriate spatial scale for the dominant input variables and parameters $\lambda(\mathbf{X})$. The $\tau(Y)$ found can also reasonably be used as the appropriate temporal scale for the inputs and parameters (as far as necessary). An illustration of this approach applied to the Red River in Vietnam and China can be found in Booij & Tran (2005).

Hydrological flood estimation methods

The (extended) methodology for the determination of appropriate scales incorporating uncertainty criteria is applied to three flood estimation methods: the SCS curve number method combined with: (a) the SCS dimensionless unit hydrograph (Maidment, 1992, p. 9.21–9.26); and (b) a dimensionless unit hydrograph derived in the UK (UH method; Shaw, 1994, p. 433–435) and a modified version of the hydrological model HBV (Bergström, 1995). For HBV, regionalisation relations from Booij (2005) linking the HBV key parameters with river basin characteristics (e.g. slope and surface area) in a linear way are applied. This enables comparison of output uncertainties from different flood estimation methods resulting from similar uncertainty sources. This way, the general applicability of the methodology described in the previous section can be tested.

MFOSM and Monte Carlo uncertainty propagation

The essence of uncertainty propagation analysis is to explore the statistical properties of the model output Y based on the statistical properties of input and parameters $\mathbf{X}$. Generally, the complete probability density function of Y cannot be derived *a priori* from the probability density functions of $\mathbf{X}$ (if these are known); however, often first and second moments of Y can be obtained from statistical properties of $\mathbf{X}$. Two uncertainty propagation methods are used: the mean-value first-order second-moment (MFOSM) method and Monte Carlo analysis.

MFOSM method The MFOSM method is derived from Taylor's linear approximation of Y around the mean of $\mathbf{X}$, $\mu(\mathbf{X})$, in which the nonlinear components are truncated:

$$Y = f(\mathbf{X}) \cong f[\mu(\mathbf{X})] + \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}[\mathbf{X} - \mu(\mathbf{X})] \tag{1}$$

and the first and second moments for independent variables are, respectively:

$$\mu(Y) \cong f[\mu(\mathbf{X})] \tag{2}$$

$$\sigma^2(Y) \cong \sum_{i=1}^{n}\left[\frac{\partial f(\mathbf{X})}{\partial x_i}\right]^2 \sigma^2(x_i) \tag{3}$$

The main advantages of MFOSM are its simplicity, the partitioning of the model output uncertainty $\sigma^2(Y)$ into its various contributions, and the two moments of Y expressed as functions of the moments of $\mathbf{X}$. Disadvantages are the assumptions of linearity in the region around $\mu(\mathbf{X})$ and the independence of the different components x_i. Melching (1995) summarized the results of many applications of MFOSM to hydrological modelling.

Monte Carlo analysis Monte Carlo analysis involves the random sampling of components x_i from the input vector $\mathbf{X}$ and the determination of the model output Y. Advantages of Monte Carlo analysis are its general applicability for estimation of response statistics of any nonlinear and/or discontinuous model, and the possibility to obtain probability distributions of Y given the probability distributions of $\mathbf{X}$. Disadvantages of Monte-Carlo analysis are the computational demand and the inability to directly show the uncertainty contributions of each component. Examples of hydrological applications can be found in Harlin & Kung (1992) and Uhlenbrook *et al*. (1999).

Backward uncertainty propagation and appropriate scale assessment

Backward uncertainty propagation The two uncertainty analysis methods described above consider forward uncertainty propagation. However, backward uncertainty propagation is rarely performed. One of the exceptions is Abusam *et al.* (2003), who illustrated a procedure for backward uncertainty propagation applied to a water quality example. They started from the statistical distribution function of the output (obtained through Monte Carlo analysis), and moved backward towards finding the set of parameter vectors that had resulted in a class or classes of interest (e.g. extreme values) in the given distribution function. Here, the MFOSM method will be used for backward uncertainty propagation. Once it has been shown that this method can be used to determine the output uncertainty, it will be used to propagate backward the output uncertainty $U(Y)$ to the uncertainty sources originating from dominant variables $U(\mathbf{X})$. Assuming a specific maximum value for the output uncertainty σ_{max} (Y) and a balance in uncertainties from inputs and parameters through the requirement of the same relative $U(x_i) = \sigma(x_i)/x_i$, the required relative input uncertainty can be derived from equation (4):

$$\overline{\left(\frac{\sigma(x)}{x}\right)} = \frac{\sigma_{\max}(Y)}{\sqrt{\sum_{i=1}^{n}\left[\frac{\partial f(\mathbf{X})}{\partial x_i}\right]^2 x_i^{\,2}}} \tag{4}$$

The assumptions of linearity and independence of inputs and parameters in the MFOSM method should be checked, although comparison of the forward uncertainty propagation results of the MFOSM method with the Monte Carlo results should give sufficient confidence in applying the backward propagation approach using MFOSM.

Appropriate scale assessment The determination of appropriate spatial scales has been extensively described by Booij (2002b, 2003). Here, only a brief summary of the approach is given, emphasizing the coupling with the required relative input uncertainty from equation (4). The appropriate spatial scale for a variable depends on its correlation structure and the application area studied. This can be expressed through relations between relevant statistics (i.e. variance, extreme values) at the point scale (subscript "p") and those at the really averaged scale (subscript "A"), e.g. for the variance:

$$\left(\sigma^2\right)_A = \left(\sigma^2\right)_p \kappa^2 \tag{5}$$

where κ^2 is the variance reduction function decreasing with increasing area A. Its magnitude depends on the spatial correlation structure of the variable, and the size and shape of the area. Similar relations are available for other statistics such as extreme value statistics.

The appropriate spatial scale can be determined by means of relations such as equation (5), given a specific appropriateness criterion. This criterion is based on the bias allowed in estimating the statistics of areally averaged variables from the statistics of point variables. This bias is assumed to be equal to the required relative input uncertainty from equation (4), enabling a direct coupling between the output uncertainty (and temporal scale) and the appropriate spatial scales of input variables and parameters. For example, for the variance this gives:

$$\overline{\left(\frac{\sigma(x)}{x}\right)} = 1 - \kappa^2 \tag{6}$$

Thus, once the maximum output uncertainty is known, the required relative input uncertainty equal to the bias can be determined, and, finally, the appropriate spatial scales can be assessed.

MFOSM AND MONTE CARLO UNCERTAINTY PROPAGATION RESULTS

Different hydrological flood estimation methods

The effect of forward propagation of uncertainties in inputs and parameters towards the output using Monte Carlo analysis on the three flood estimation methods has been investigated using 10 000 simulations for each method with normally distributed uncertainty in inputs and parameters, assuming a standard deviation of 5% of the mean. Thus, $U(x_i) = \sigma(x_i)/x_i$ from equations (4) and (6) is equal to 0.05. This assumption is made for the purpose of illustration of the methods proposed. Actual uncertainty analyses of hydrological models have applied much higher relative standard deviations of inputs and parameters. All inputs and parameters are assumed to be independent. The means of the peak discharges for the SCS method, UH method and HBV model are, respectively, 2530, 2500 and 2420 m^3/s and their corresponding standard deviations are, respectively, 980, 930, and 560 m^3/s. Thus, the relative output uncertainty $\sigma(Y)/Y$ is 39, 37 and 23%, respectively.

Differences between the SCS and UH methods are small, probably due to the similar equations used to derive peak discharges. Therefore, in the remainder of this paper only the SCS method is analysed and is assumed to represent the behaviour of the UH method as well. Differences between the SCS method and the HBV model are considerable, despite the fact that similar parameters are used (partly through regression relations). The main causes of these differences may be the land use related parameters and obviously the way peak discharges are derived using the parameters.

Different uncertainty propagation methods

Figure 2 shows the uncertainty in the peak discharge for the SCS method and HBV model, using the MFOSM method and the Monte Carlo (MC) approach for uncertainty propagation, resulting from 5% uncertainty in inputs and parameters. Results from MFOSM and MC for each flood estimation method are quite similar, suggesting that the more cumbersome MC approach could be replaced by the more straightforward MFOSM method. This would enable application of the backward uncertainty propagation method.

This is further investigated in Fig. 3, where the relative output uncertainty as a function of the relative input uncertainty for the SCS method and HBV model using MFOSM and MC uncertainty propagation is shown. The results from Fig. 2 are confirmed, i.e. the differences between MFOSM and MC for both the SCS method and HBV model are negligible, in particular when compared with the differences between the two flood estimation methods. It thus can be concluded that MFOSM can be used for the SCS method and HBV model under similar climatological and geographical conditions, and assuming independency of all inputs and parameters. When certain inputs and/or parameters show significant correlation, this may be included in the MFOSM method through the use of correlation coefficients between these inputs and parameters.

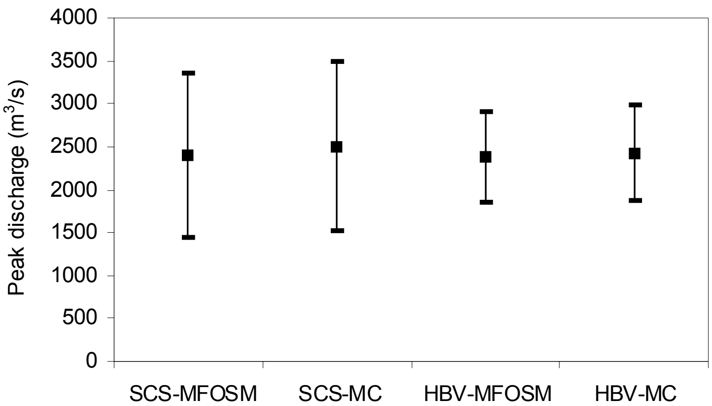

Fig. 2 Uncertainty in peak discharge expressed as mean +/– standard deviation for the SCS method and HBV model using MFOSM and MC (Monte Carlo) uncertainty propagation methods resulting from 5% uncertainty in inputs and parameters.

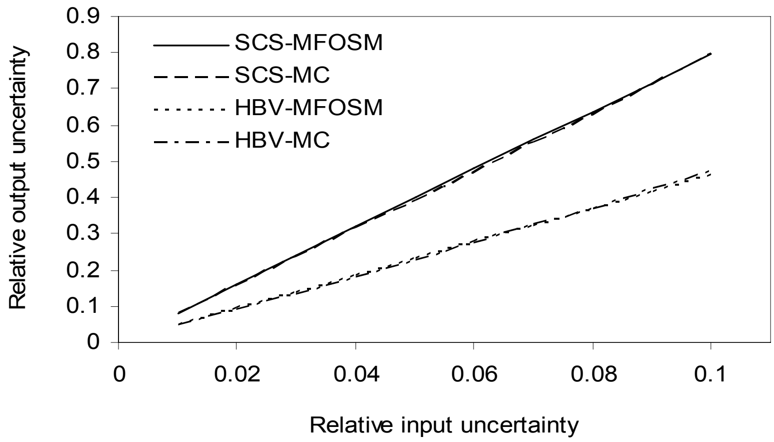

Fig. 3 Relative output uncertainty [$\sigma(Y)/Y$] as a function of relative input uncertainty [$\sigma(x_i)/x_i$] for the SCS method and HBV model using MFOSM and MC (Monte Carlo) uncertainty propagation methods.

BACKWARD UNCERTAINTY PROPAGATION AND APPROPRIATE SCALE RESULTS

Backward uncertainty propagation

The application of the backward uncertainty propagation method has been justified in the previous sub-section. This feature of the SCS method and HBV model can be used to estimate the required relative input uncertainty assuming a specific maximum value for the relative output uncertainty. This is illustrated in Fig. 4 for the SCS method and

HBV model with a relative output uncertainty of 25%. This results in a relative input uncertainty of about 3% for the SCS method, and of about 5.5% for the HBV model. Such input and parameter uncertainty targets are probably not achievable in practical, real world modelling. For example in Melching *et al*. (1991), the relative parameter uncertainties in an uncertainty analysis applied to two hydrological models varied from 15 to 110%. Thus, if models such as the SCS method and HBV model are to be used, a higher uncertainty must be accepted. Otherwise, one should evaluate more complex models that "might" yield higher accuracy, and evaluate these models using the methodology proposed in this paper.

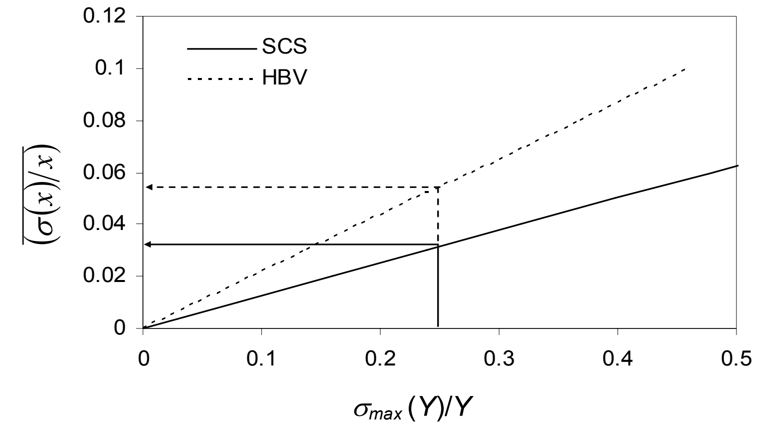

Fig. 4 Illustration of determination of the required relative input uncertainty for a variable [$\overline{\sigma(x)/x}$] accepting a specific maximum value for the relative output uncertainty [$\sigma_{\max}(Y)/Y$] of 25% for the SCS method and HBV model.

Because of the linear, monotonic relation between the input and output uncertainty, relations based on the Monte Carlo approach such as in Fig. 3 could eventually also be used to determine the required relative input uncertainty. However, this latter approach is less efficient and straightforward than the backward approach illustrated here, because of the cumbersomeness of the Monte Carlo approach itself, and the necessity to evaluate the output uncertainty for a number of input uncertainties.

Appropriate scales

Analogously to Booij (2003) the required input uncertainties can be translated to appropriate spatial scales (Δx) expressed as a fraction of the spatial correlation length (L). This is shown in Fig. 5, where the determination of the appropriate spatial scale for a variable ($\Delta x/L$) for a specific relative input uncertainty [$\overline{\sigma(x)/x}$] for the HBV model is illustrated. The appropriate scale for a relative input uncertainty of 5.5% is about 13% of the spatial correlation length, when the variance of the variable is of interest. This could be somewhat more when extreme values are of interest. This results in smaller appropriate scales for dominant variables, such as precipitation, elevation, soil, and land use, than those determined in Booij (2002b) and Booij (2003), assuming a relative input uncertainty of 10% (resulting in appropriate scales as a fraction of the correlation length of 20% for the variance and 25% for extreme values). When requiring a relative output uncertainty of 25% and a balance in uncertainties from inputs and parameters, the appropriate scales for precipitation, elevation, soil, and land use become about 13 km, 65 m, 3.4 km, and 2.1 km, respectively. Using the methodology from Booij (2003), these dominant variables can be integrated into a (smaller) appropriate model scale. This is not shown here, but obviously the final result depends on the required *output* uncertainty of 25% (compared with the required *input* uncertainty of 10% in the earlier study).

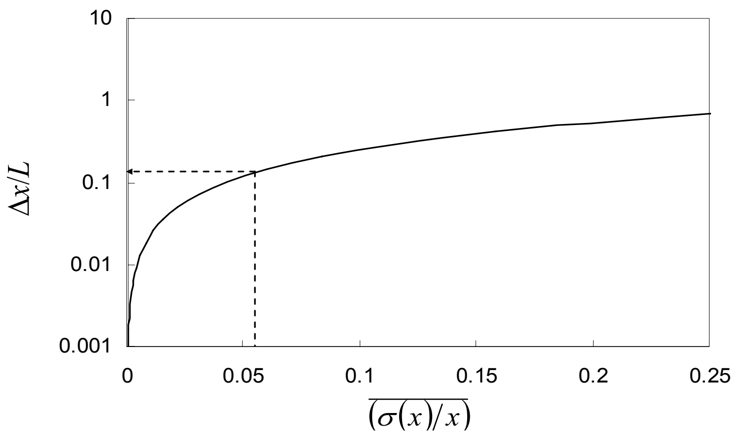

Fig. 5 Illustration of determination of appropriate spatial scale for a variable ($\Delta x/L$) for a specific relative input uncertainty [$\overline{\sigma(x)/x}$] (see Fig. 4) for the HBV model.

CONCLUSIONS

A methodology is introduced to determine the appropriate spatial scales of dominant variables in the context of river flooding based on a balance in uncertainties from inputs and parameters and a maximum output uncertainty. It has been applied to three flood estimation methods and checked using two uncertainty propagation methods. The application revealed that the methodology can be used for the flood estimation methods considered under similar climatological and geographical conditions and assuming independence of all inputs and parameters. The results showed for a specific maximum output uncertainty different relative input and parameter uncertainties for the different flood estimation methods resulting in different appropriate spatial scales of dominant variables.

The assumed balance in uncertainties (equal coefficients of variation) among the uncertain model inputs and parameters is unlikely to be achieved in actual modelling cases. Further, the contribution of the various uncertain inputs and parameters to the model output uncertainty is also uneven (i.e. not balanced). Thus, future research should be directed toward determining different levels of required accuracy/uncertainty for different inputs and parameters, on the basis of the sensitivity of the output uncertainty to these inputs and parameters. These different levels of required accuracy/uncertainty should also be compared to the ability to reduce the uncertainty of the various inputs and parameters by increasing the data available for their determination, decreasing the spatial scale of their evaluation, and/or other means. In this way, a methodology for the determination of appropriate spatial scales based on a balance in contributions of inputs and parameters to the output uncertainty, and a maximum output uncertainty could be obtained. Furthermore, the feasibility of uncertainty contributions of inputs and parameters might also be included.

The assumptions underlying the backward uncertainty propagation used in the methodology are (semi-)linearity of the model behaviour and independence of all inputs and parameters. It has been shown that (semi-)linearity can be assumed for the flood estimation methods considered and the mean-value first-order second-moment (MFOSM) method can be used for backward uncertainty propagation. When certain inputs and/or parameters show significant correlation (dependency), this may be included in the MFOSM method through the use of correlation coefficients between inputs and parameters.

Depending on which flood estimation method is used, different combinations of appropriate spatial scales of dominant variables are found. Using flood estimation methods other than those described here, would probably result in yet more different combinations of scales. However, once a certain flood estimation method is chosen for a roughly similar climatological regime and geographical area, the methodology can be applied to determine, depending on the user's preferences with respect to output uncertainty, appropriate spatial scales based on a balance in input and parameter uncertainties. The case study presented here was primarily used to illustrate the usefulness of the methodology.

Acknowledgements Eric Sprokkereef of RIZA provided the essential Meuse basin data and Joop Gerretsen of Rijkswaterstaat Limburg made the discharge data available. Sten Bergström of the Swedish Meteorological and Hydrological Institute kindly provided the HBV model.

REFERENCES

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. L. (1986) An introduction to the European hydrological system – Système Hydrologique Europeen, "SHE". 2. Structure of a physically-based, distributed modelling system. *J. Hydrol.* **87**, 61–77.

Abusam, A., Keesman, K. J. & Van Straten, G. (2003) Forward and backward uncertainty propagation: an oxidation ditch modeling example. *Water Res.* **37**, 429–435.

Bergström, S. (1995) The HBV model. In: *Computer Models of Watershed Hydrology* (ed. by V. P. Singh), 443–476. Water Resour. Publ., Highlands Ranch, Colorado, USA.

Booij, M. J. (2002a) Appropriate modelling of climate change impacts on river flooding. PhD Thesis, University of Twente, Enschede, The Netherlands.

Booij, M. J. (2002b) Extreme daily precipitation in Western Europe with climate change at appropriate spatial scales. *Int. J. Climatol.* **22**, 69–85.

Booij, M. J. (2003) Determination and integration of appropriate spatial scales for river basin modelling. *Hydrol. Processes* **17**, 2581–2598.

Booij, M. J. (2005) Impact of climate change on river flooding assessed with different spatial model resolutions. *J. Hydrol.* **303**, 176–198.

Booij, M. J. & Tran, H. T. (2005) Appropriate scales for hydro-climatological variables in the Red River basin. In: *Regional Hydrological Impacts of Climatic Variability and Change – Impact Assessment and Decision Making* (ed. by T. Wagener, S. Franks, H. V. Gupta, E. Bøgh, L. Bastidas, C. Nobre & C. De Oliveira Galvão). (Symposium 6 at IAHS 2005, Foz do Iguaçu, Brazil, 3–8 April 2005), 325–332. IAHS Publ. 295, IAHS Press, Wallingford, UK.

Harlin, J. & Kung, C. -S. (1992) Parameter uncertainty and simulation of design floods in Sweden. *J. Hydrol.* **137**, 209–230.

Maidment, D. R. (1992) *Handbook of Hydrology*. McGraw-Hill, New York, USA.

Melching, C. S. (1995) Reliability estimation. In: *Computer Models of Watershed Hydrology* (ed. by V. P. Singh), 69–118. Water Resour. Publ., Highlands Ranch, Colorado, USA.

Melching, C. S., Yen, B. C. & Wenzel, H. G., Jr. (1991) Output reliability as guide for selection of rainfall-runoff models. *J. Water Resour. Plan. Manage. ASCE* **117**, 383–398.

Moody, A. & Woodcock, C. E. (1995) The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. *Landscape Ecology* **10**, 363–379.

Shaw, E. M. (1994) *Hydrology in Practice*. Chapman & Hall, London, UK.

Uhlenbrook, S., Seibert, J., Leibundgut, C. & Rodhe, A. (1999) Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrol. Sci. J.* **44**, 779–797.

Western, A. W., Blöschl, G. & Grayson, R. B. (1998) Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment. *J. Hydrol.* **205**, 20–37.

Wolock, D. M. & McCabe, G. J. (2000) Differences in topographic characteristics computed from 100- and 1000-m resolution digital elevation model data. *Hydrol. Processes* **14**, 987–1002.