PERFORMANCE ASSESSMENT OF HYDROLOGIC MODELS BASED ON THE UNCERTAINTY OF MEASUREMENTS

By

Amin Elshorbagy

Centre for Advanced Numerical Simulation (CANSIM), Department of Civil & Geological Engineering, University of Saskatchewan, Saskatoon, SK, Canada

CANSIM SERIES REPORT NO. CAN-06-02

Centre for Advanced Numerical Simulation (CANSIM)

Department of Civil & Geological Engineering,

University of Saskatchewan, Saskatoon, SK, CANADA

August, 2006

Performance Assessment of Hydrologic Models Based on the Uncertainty of Measurements

Amin Elshorbagy
Centre for Advanced Numerical Simulation (CANSIM), Department of Civil and Geological Engineering, University of Saskatchewan, Saskatoon, SK, Canada

Abstract

The predictive uncertainty of hydrologic models has attracted the attention of hydrologists during the last decade. However, an often overlooked type of uncertainty is the uncertainty of hydrologic measurements. Hydrologic measurements are the reference against which simulated values are contrasted. This paper presents a methodology to account for the uncertainty of hydrologic measurements and to assess the model performance based on both predictive and measurement uncertainty. Accordingly, three error measures, namely the measurement uncertainty-based mean relative error (MU_MRE), measurement and predictive uncertainty-based credibility (MPU_CRE), and mean relative error (MPU_MRE) are defined to assess the performance of hydrologic models. The proposed approach and error measures are perceived to be more realistic in light of uncertain measurements. They can provide a different assessment from an assessment made based on traditional error measures. This approach could have further implications and impacts on the process of calibration and validation of hydrologic models.

1. Introduction

The importance and utility of hydrologic models to aid decisions in the design and planning of water resources cannot be overemphasized. Experienced users of such models used to rely on their subjective judgement and experience to select a model for a specific task and to judge the credibility of the model results. Water resources decision makers intuitively realize that there is a level of uncertainty associated with any model, but information on the level of uncertainty is frequently missing. Two different models may provide a similar value for a specific output based on a single-valued prediction (i.e., expected value); however, the results from a model with an uncertainty level of 10% are significantly different from the results from a model with 60-70% uncertainty, even if both models provide the same expected value. Using a hydrologic model without enough information about its uncertainty (i.e., the reliability of the expected value) could be misleading.

Uncertainty is a term that is used by different researchers to mean different things or to imply one of many possible types of uncertainty. In hydrologic modeling literature, uncertainty most often refers to the type of uncertainty caused by model parameter specification [Franks and Beven, 1997; Vrugt et al., 2003; Bates et al., 2004; Smith and Wheater, 2004]. However, others [e.g., Bastidas et al., 2003] acknowledged explicitly that uncertainty (expressed in their work as total error) is the result of data error, model structural error, and model specification error. The issue of the reliability of hydrologic models' output, already overshadowed by uncertainty, is further complicated by the uncertainty about the estimates of uncertainty themselves [Beven, 2004; Borsuk et al., 2002].

The ability of various promising methods, including those reported in the articles cited earlier, to estimate the bounds of uncertainty around the simulated values made it possible to reasonably estimate the predictive uncertainty of hydrologic models. However, the puzzle of uncertainty estimates is missing an important piece because the uncertainty of the measurements is frequently forgotten. The model-based results are evaluated against observed values, using the assumption that the observed values are certain. In uncertainty-related hydrologic literature, it is common to see a graph showing observed values (e.g., stremflow) falling within or outside the bounds of predictive uncertainty. This paper argues that the bounds of predictive uncertainty need to be evaluated against the bounds of measurement uncertainty. Accordingly, modelers and hydrologists should deal with two ranges of vlaues (the first is the range of possible simulated values and the second is the range of possible observed values). As expected, this concept of uncertainty of measurements leads to a change in the traditional method of assessing model performance and estimating model error. This paper will address the issue of measurement uncertainty, which will potentially lead to the modification of the definition of error measure, and develop an uncertainty-based erorr measure for hydrologic models.

Uncertainty of measurements

The uncertainty about the measurements is a result of misrepresentation (the focus of this study is on the data used as observed values to contrast with the simulated results). For example, a lumped watershed model predicts depth-and-area-averaged soil moisture values that cannot be accurately contrasted with soil moisture observations monitored at a single point. Another, and possibly more important, source of measurement uncertainty is the level of accuracy of the properly calibrated measurement instrument reported by the manufacturer. For example, measurements of soil moisture taken using time domain reflectrometry (TDR) sensors are subject to errors due to air gaps around the access tubes, the relatively small volume of soil sampled by the probes (their sphere of influence), and the vulnerability to local anomalies within the soil such as stones and other spatial variations [Strangeways, 2003]. Other sources, such as human error, may contribute to the uncertainty, but they are extremely difficult to quantify.

Methodology: quantification of uncertainty

In this section the methodology adopted in this paper is briefly outlined, followed by an illustrative example in the following section. First, a hydrologic model was executed to simulate a hydrologic system and produce simulated daily values of a specific hydrologic process over the period of one year. Second, the predictive uncertainty of the model was estimated using Monte Carlo simulation of the model residuals. The estimated level of uncertainty was used to delineate the upper and lower boundaries of the predictive uncertainty. Third, the uncertainty about the measurements of the hydrologic process under consideration was quantified using the level of accuracy of the measuring instrument (e.g., +/- 2.5%); the hydrologic variable used in this study is volumetric soil moisture content. The estimated level of uncertainty was used to delineate the upper and lower boundaries of the measurement uncertainty. Fourth, a measurement uncertainty-based relative error measure (MU-MRE) was defined and estimated to assess the performance of the model as follows:

Practically, it makes no sense to train or calibrate a model such that the simulated values get closer to the observed values than the pre-determined bounds of measurement error. Simulated values that fall within the bounds of measurement uncertainty were considered as having zero residuals. The error is evaluated using the following equations:

$$MU\_RE = \left| \frac{\theta_{si} - \theta_{bo}}{\theta_{bo}} \right| x100 \quad (1)$$

$$MU\_MRE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\theta_{si} - \theta_{bo}}{\theta_{bo}} \right| x100 \quad (2)$$

where MU_RE is the measurement uncertainty-based relative error, MU_MRE is the measurement uncertainty-based mean relative error, N is the total number of observations, θ_{si} is the simulated values, and θ_{bo} is the nearest boundary of uncertain measurements (observed values).

A more comprehensive measure of the model performance should be based on both measurement and predictive uncertainties. In this case, the error measure contrasts the range of uncertain simulated values against the range of uncertain measurements. Figure 1 presents an illustrative example of four possible cases: (a) the measurement and predictive uncertainties match, which leads to zero error, and the model is perceived as a perfect model; (b) the predictive uncertainty is contained within the measurement uncertainty. In this case, the error is considered to be zero, but the model is imperfect since underestimation and overestimation are encountered but are difficult to quantify; and (c) & (d) the measurement uncertainty is contained within the predictive uncertainty.

In this case, the error is the portion of the simulated range that falls outside the measurement range (shaded areas in Figure 5c and 5d). Two uncertainty-based error measures are developed to account for both predictive and measurement uncertainty. The first one measures the relative area of the simulated range that falls outside the observed range, which is named the measurement and predictive uncertainty-based credibility (MPU_CRE). This could measure the credibility of the model, but it does not measure how far the simulated range falls outside the observed range. Therefore, another measure, which estimates the same relative area assessed by the MPU_CRE using the relative ordinates calculated by Eq (1), is calculated and named the measurement and predictive uncertainty-based mean relative error (MPU_MRE). For example, the MPU_CRE could be the same (e.g., 50%) in both cases (Figure 5c) and (5d) but the MPU_MRE is different (e.g., 10% and 30%).

Illustrative example: A case study of simulation of reconstructed watersheds

A system dynamics watershed (SDW) model for reconstructed watersheds has been developed by Elshorbagy et al. [2005], and was used in this study. The model was constructed to simulate watersheds that are reconstructed as part of the reclamation strategy adopted by the oil sands mining industry in Northern Alberta, Canada. The mining of oil sands near Fort McMurray, Alberta, involves the stripping of saline/sodic overburden to gain access to the oil-bearing formation. The overburden is placed in large mined-out pits and surface dumps and is re-contoured before being capped with a mandated 1 m soil cover.

The oil sands mining industry is conducting large-scale cover experiments at the Mildred Lake mine in order to assess the performance of different reclamation soil covers. Three one-hectare prototype covers were placed on an area referred to as the South Hills Overburden (Figure 2) to study the basic mechanisms controlling moisture movement within the cover systems. The covers were constructed in 1999 with thicknesses of 1.0 m, 0.50 m, and 0.35 m comprised of a thin layer of peat (15-20 cm) overlying varying thickness of secondary soil. A field instrumentation program was carried out consisting of detailed monitoring of matric suction, volumetric water content, and soil temperature within the different soil profiles, as well as measurements of runoff, interflow, and site-specific meteorological conditions. For more details on the instrumentation program and site description, the readers are referred to Boese [2003] and Meier and Barbour [2002].

Evaluation of the hydrologic performance of the soil covers to assess their ability to maintain sufficient soil moisture during growing seasons has been conducted and detailed in Elshorbagy et al. [2005] and Jutla et al. [2005]. Such performance is an important indicator of the efficacy of the soil cover in restoring the production function of the watershed. The uncertainty analysis conducted in this paper relies mainly on contrasting measured (observed) soil moisture content with those simulated using the SDW model. Soil moisture was selected for this study because it is one of the most important indicators of the soil cover performance; the possibility of quantifying the uncertainty about the measurements taken using the TDR sensors was another important factor.

Results and analysis

The SDW model was executed to simulate depth-averaged soil moisture content in both the peat and till layers in soil cover D3, which was used in this study. The simulated and observed soil moisture content levels are shown in Figure 3. Only soil moisture values during the growing season are reported here. The soil is frozen during the remaining portion of the year. A satisfactory performance of the model is achieved with mean relative error (MRE) values of 11% and 2% for the peat and till layers, respectively [Jutla et al., 2005]. The residuals (simulated values minus observed values) of the SDW model were used to quantify the predictive uncertainty of the model. Both Weibull and Normal distributions, and Logistic and Normal distributions were found to provide a close fit to the residual values of peat and till layers, respectively. Accordingly, Normal distributions were used to generate probability distributions of residuals using the following calculated values of mean and standard deviation of residuals measured as relative error in percent units: (0.06, 0.13) and (-0.02, 0.03) for both peat and till layers, respectively. The upper and lower predictive uncertainty bounds are selected such that they contain 95% of the residuals (Fig. 4).

Walker et al. [2004] have shown that volumetric soil moisture sensors used in the field have various levels of accuracies when compared to soil moisture estimated using the thermo gravimetric method. They suggest that the highest level of +/- 2.5% v/v accuracy could be attained using calibrated connector-type TDR sensors. Other types of sensors could result in lower levels of accuracy (i.e., higher uncertainty). In this study, a minimum level of measurement uncertainty of +/- 2.5 % v/v is assumed. This simply means that as much as +/- 5 mm and +/- 20 mm of water could overshadow every soil moisture measurement within the peat and till layers, respectively. In light of this understanding, the simulated soil moisture values are plotted against the upper and lower bounds of uncertain measurements (Figure 5). Assessment of the SDW model performance based on the measurement uncertainty is significantly different where residuals (i.e., errors) are considered only for simulated values that fall outside the bounds of measurement uncertainty. The MU-MRE values of the SDW model are 2% and zero for the peat and till layer, respectively.

Figure 6 shows the results of the SDW model with regard to the simulation of soil moisture in light of both predictive and measurement uncertainties. The performance of the SDW model based on both deterministic and uncertainty-based approaches is summarized in Table 1. About 45% of the range of simulated soil moisture values in the peat layer falls outside the range of uncertain measurements (i.e., the credibility is 55%). The MPU_MRE value of 10% measures how far the 45% is from the boundaries of uncertain measurements. It can be also interpreted that the MPU_CRE measures the accuracy of the model, whereas the MPU_MRE measures the precision.

Discussion

This paper argues that the developed MU_MRE, MPU_CRE, and MPU_MRE are more realistic than the traditional error measures, such as the MRE, because the former measures take into consideration the inevitable uncertainty of measurements as well as the predictive uncertainty. Predictive uncertainty could be given more attention by breaking it down into parameter uncertainty, as addressed earlier in literature [Freer et al., 1996], and structure uncertainty. However, in this paper, it was assumed that uncertainty due to the homoschedastic model residuals cover the possible range of uncertainty due to parameterization. When uncertainties are studied in more details (e.g., input uncertainty, parameter uncertainty, and structure uncertainty), it is suggested that the uncertainty that results in the widest range be considered as the predictive uncertainty. To achieve the objective of this paper, it was considered sufficient to account for the predictive uncertainty using the probability distribution of the model residuals. Elshorbagy [2005] presented an alternative approach to estimate the predictive uncertainty of the same SDW model using the Bayesian approach. The results of both Monte Carlo simulation, used in this study, and the Bayesian approach are similar.

In the same way in which residuals were used in this paper to estimate MU_MRE, MPU_CRE, and MPU_MRE, other error measures such as root mean squared error (RMSE) can be used. The definition of error in this paper is expected to affect the way hydrologic models are calibrated. The error function to be minimized can be based only on the residuals that fall outside the boundary of uncertain measurements. This approach could provide the hydrologic models with more flexibility during calibration, which may result in better performance and generalization ability over validation data sets. Inductive models such as neural networks and genetic programming could also benefit from the re-definition of the error function. Considerable efforts by hydrologists are needed to quantify uncertainty about other hydrologic measurements, such as precipitation and streamflow.

Conclusions

The earlier efforts to address the parameter and predictive uncertainty of hydrologic models cannot be complete unless the issue of measurement uncertainty is also addressed. Uncertainties about the results of hydrologic models stemming from model inputs, parameters, or structure should be contrasted against uncertain measurements to obtain realistic statistics of model performance. The measurement uncertainty-based mean relative error (MU_MRE) as well as the measurement and predictive uncertainty-based credibility (MPU_CRE) and mean relative error (MPU_MRE) developed in this study could help assess model performance in cases where such uncertainties are quantifiable. The performance of the SDW model used in this study, which was assessed using the developed error measures, could be perceived differently when assessed using traditional error measures. The developed error measures can be extended to other types of error statistics and can also play a further role in calibration and validation of hydrologic models.

Acknowledgement

The author acknowledges the financial support of the Natural Sciences and Engineering Research Council (NSERC) of Canada through the Discovery Grant Program. The logistical support of Syncrude Canada Ltd. is also appreciated.

References

Bastidas, L. A., H. V. Gupta, K. Hsu, and S. Sorooshian (2003), Parameter, structure, and model performance evaluation for land-surface schemes, in *Calibration of Watershed Models, edited by*.Q. Duan, H. V. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte, American Geophysical Union, Washington, DC.

Bates, P. D., M. S. Horritt, G. Aronica, and K. Beven (2004), Bayesian updating of flood inundation likelihoods conditioned on flood extent data, *Hydrol. Proc.*, 18, 3347-3370.

Beven, K. (2004), Does an interagency meeting in Washington imply uncertainty? *Hydrol. Proc.*, 18, 1747-1750.

Boese, K. (2003), The design and installation of a field instrumentation program for the evaluation of soil-atmosphere water fluxes in a vegetated cover over saline/sodic shale overburden. *M.Sc. dissertation, University of Saskatchewan*, Saskatoon, SK, Canada.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow (2002), Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development, *Environ. Sci. & Tech.*, 36, 2109-2115.

Elshorbagy, A. (2005), Predicting the uncertainty of watershed models using a simple Bayesian approach, 17th Canadian Hydrotechnical Conference, Edmonton, AB, Canada, 1 :10.

Elshorbagy, A., A. Jutla, S. L. Barbour, and J. Kells (2005), System dynamics approach to assess the sustainability of reclamation of disturbed watersheds, *Can. J. Civ. Eng.*, 32, 1-15.

Franks, S. W. and K. Beven (1997), Estimation of evapotranspiration at the landscape scale: A fuzzy disaggregation approach, *Water Resour. Res.*, 33, 2929-2938.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Res. Res.*, 32, 2161-2173.

Jutla, A., A. Elshorbagy, and J. Kells (2005), Beyond rainfall-runoff modeling: Hydrologic simulation of reconstructed watersheds using system dynamics, 17th Canadian Hydrotechnical Conference, Edmonton, AB, Canada, 11 - 20.

Meier, D., and S. L. Barbour (2002), Monitoring of cover and watershed performance for soil covers placed over saline-sodic shale overburden from oilsands mining, National Meeting of the American Society of Mining and Reclamation, Lexington KY, ASMR, 3134 Montavesta Rd., Lexington, KY 40502.

Smith, R. M. S. and H. S. Wheater (2004), Multiple objective evaluation of a simple phosphorus transfer model, *Hydrol. Proc.*, 18, 1703-1720.

Strangeways, I. (2003), Measuring the natural environment, Cambridge University Press, Cambridge, UK.

Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian, (2003), A Shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, 1201-1216.

Walker, J. P., G. R. Willgoose, and J. D. Kalma (2004), In situ measurement of soil moisture: a comparison of techniques. *J. Hydrol.*, 293, 85-99.

Table 1. The performance of the SDW model with respect to predictive and measurement uncertainty.

	MRE (%)	MU_MRE	MPU_CRE	MPU_MRE
Peat layer (200 mm)	11.0	2.0	45.0	10.0
Till layer (800 mm)	2.0	0.0	0.0	0.0

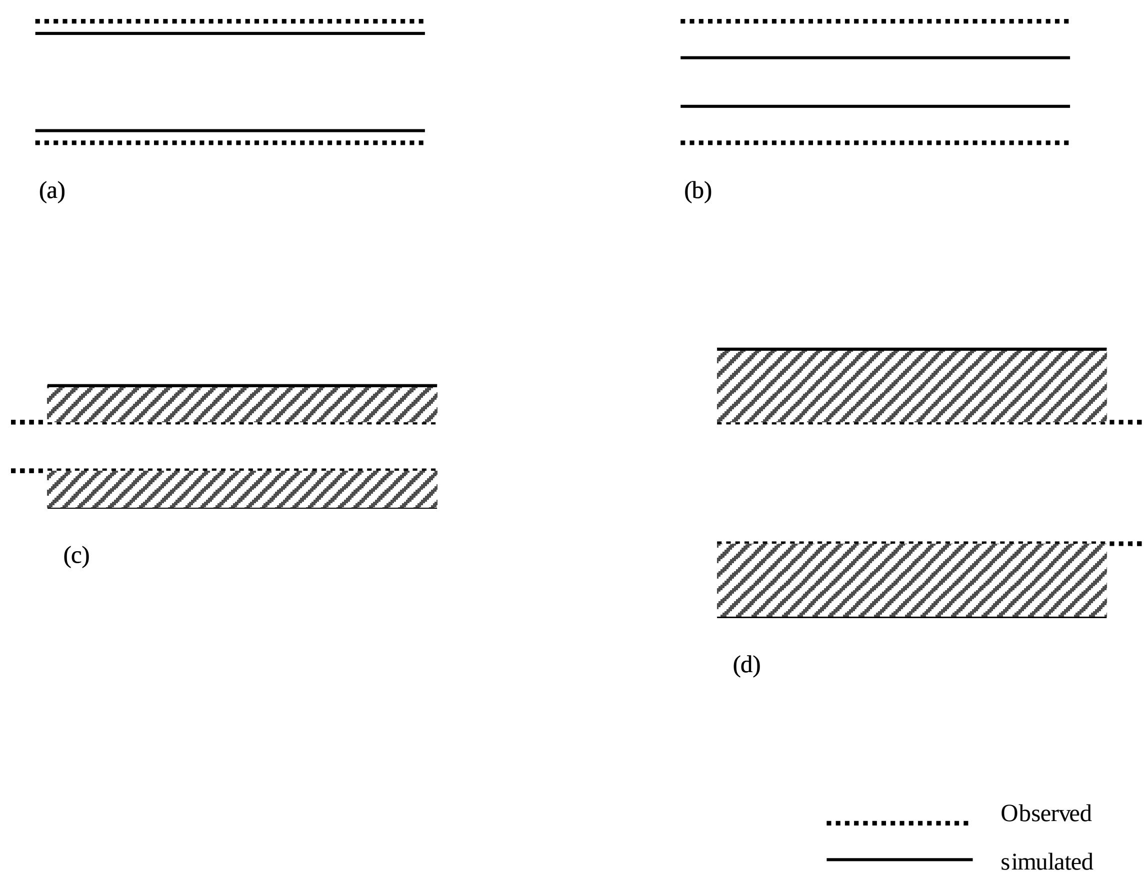


Figure 1. Possible cases of overlapping predictive and measurement uncertainty.

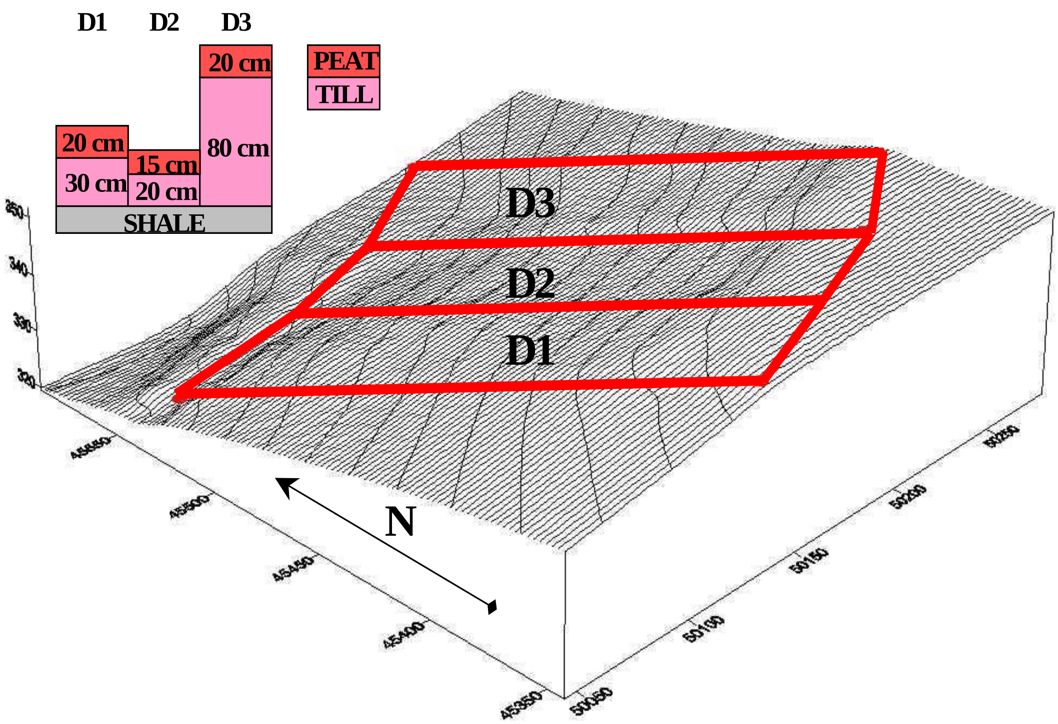
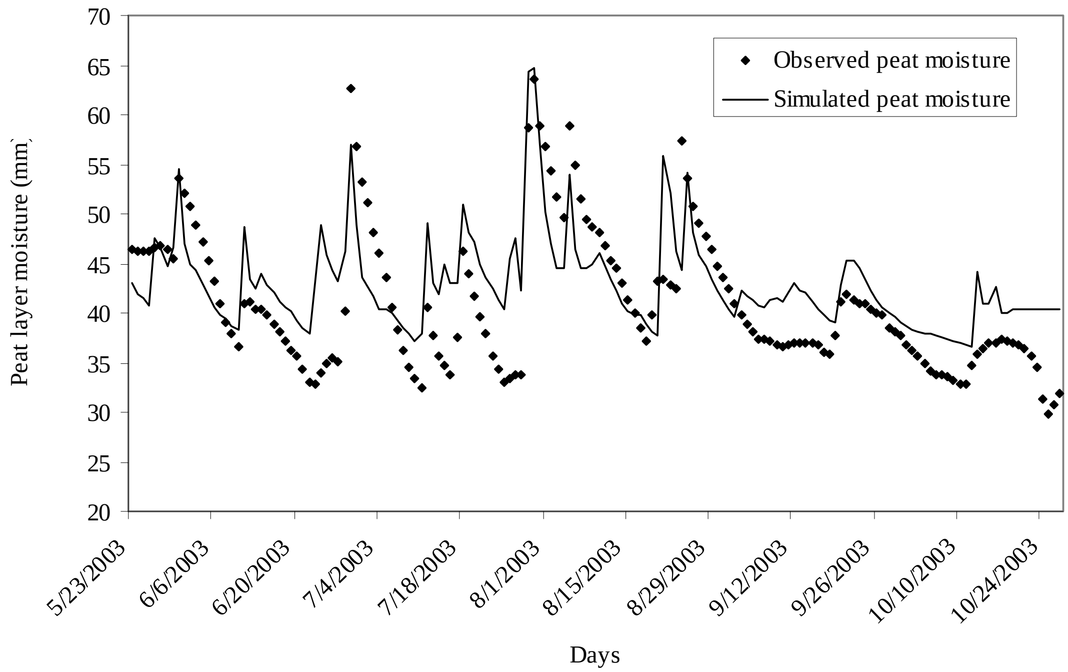
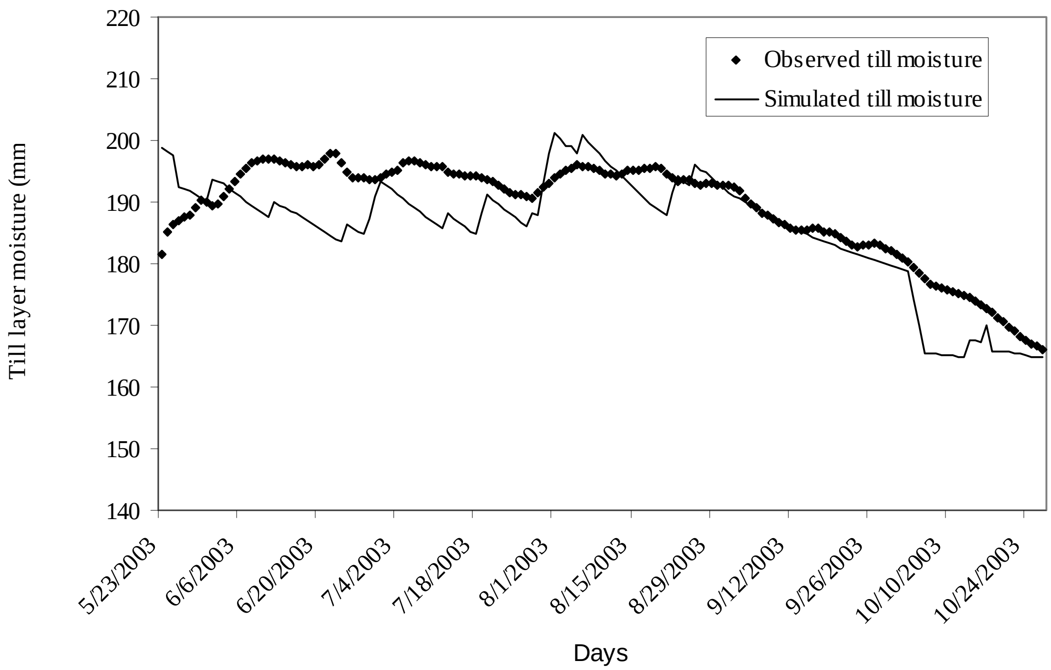


Figure 2. The prototype cover site [Elshorbagy et al., 2005].

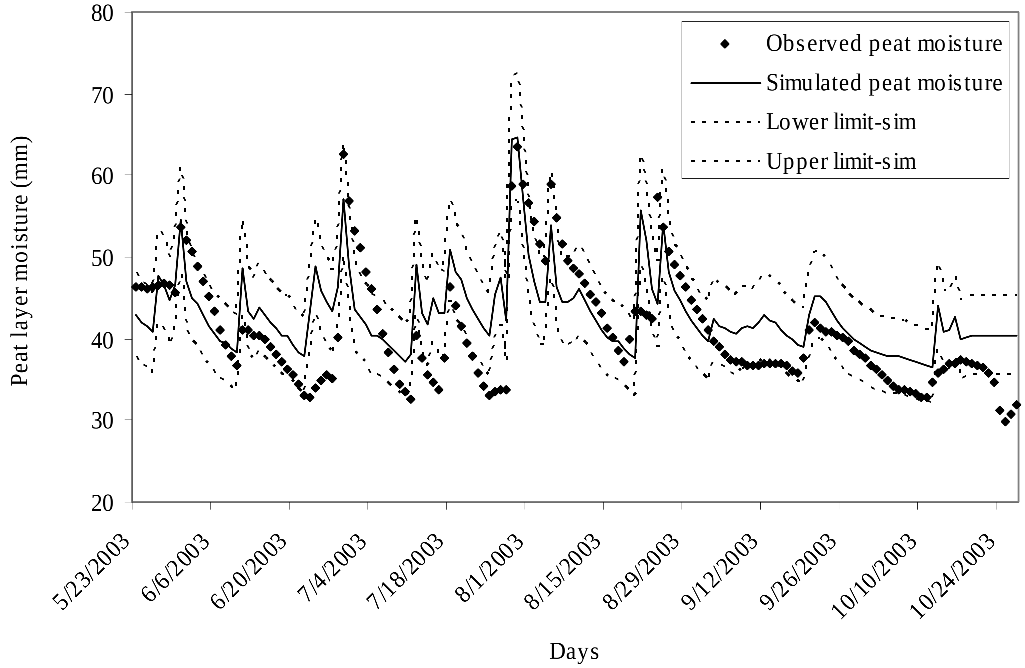


(a)

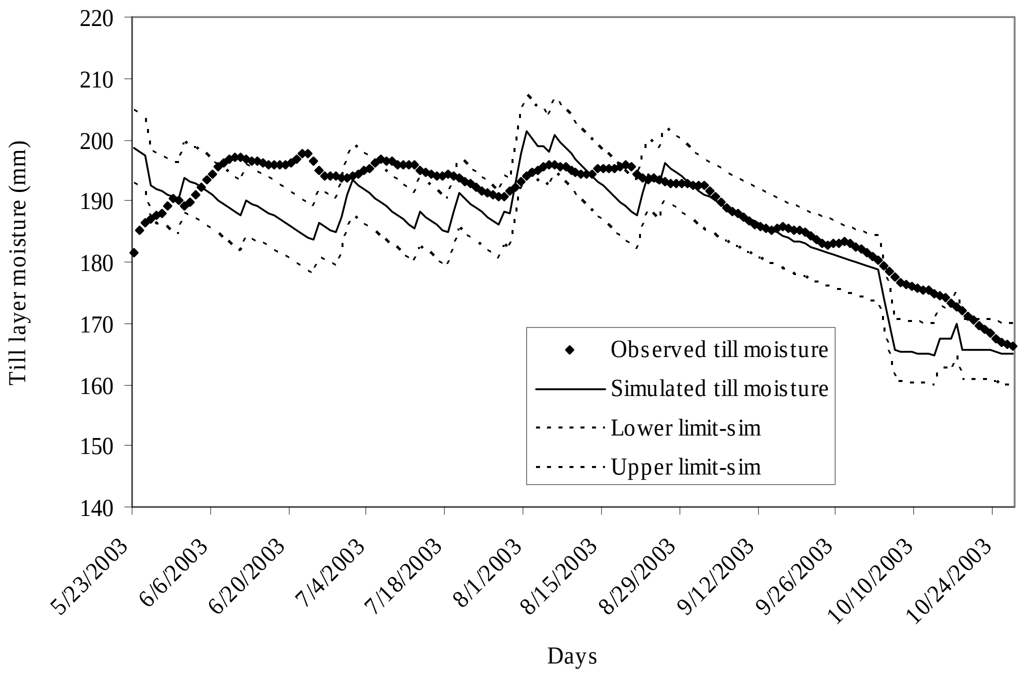


(b)

Figure 3. Observed and simulated soil moisture content; (a) upper peat layer (200mm), (b) lower till layer (800 mm).

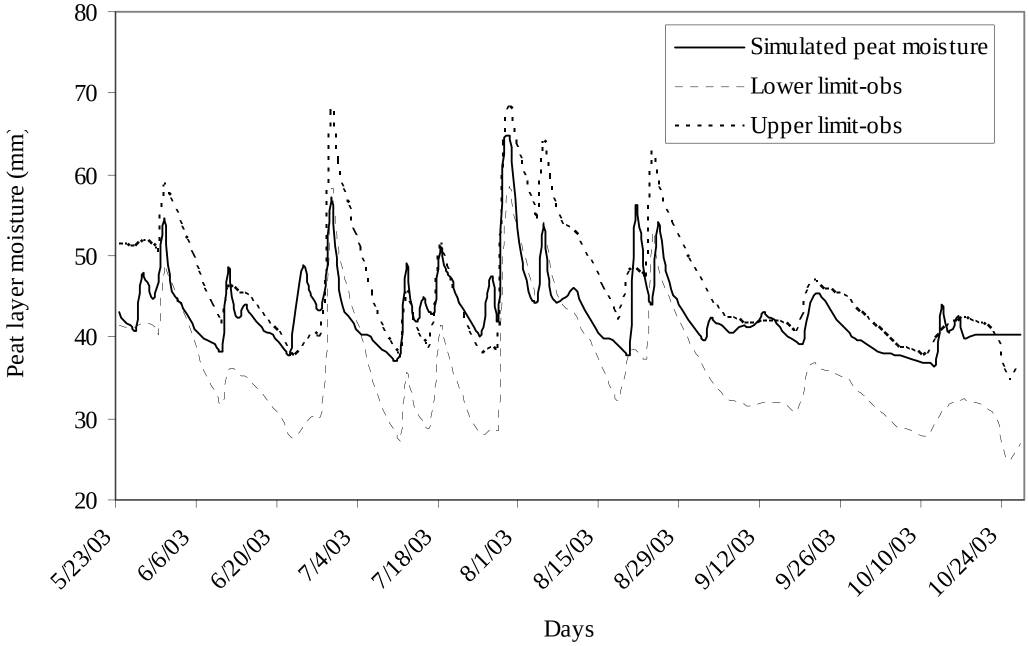


(a)

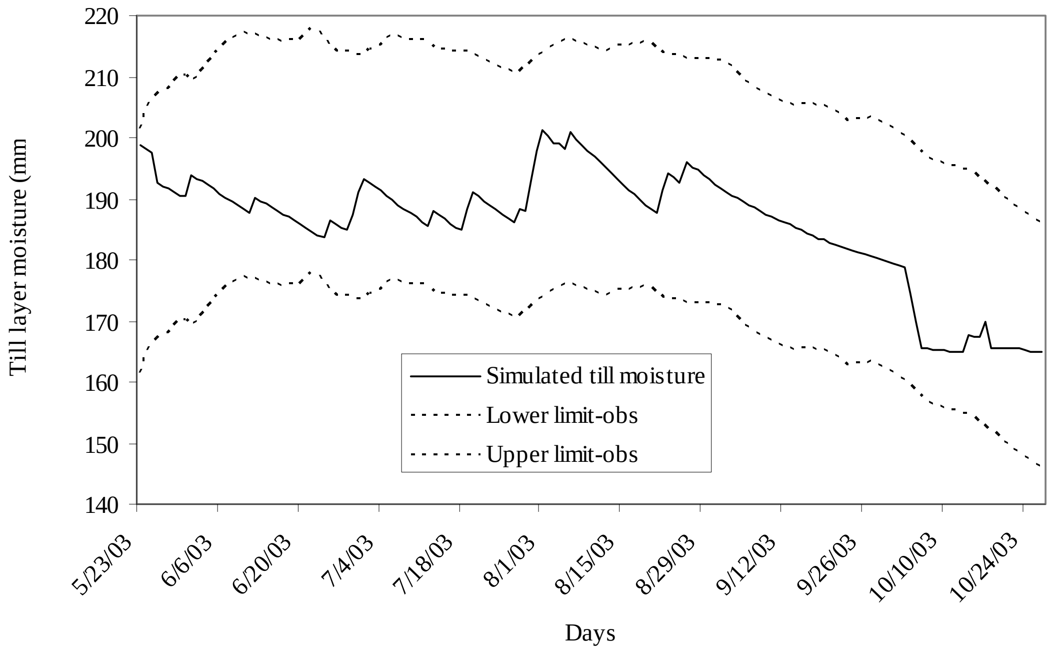


(b)

Figure 4. Predictive uncertainty about the SDW model results; (a) upper peat layer (200mm), (b) lower till layer (800 mm).

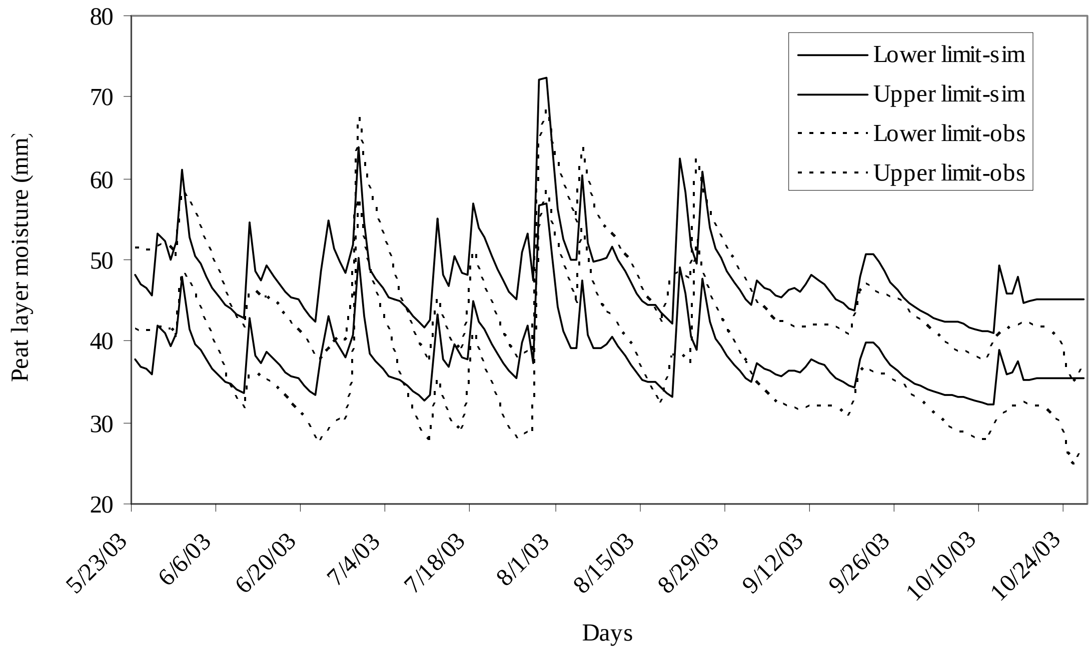


(a)

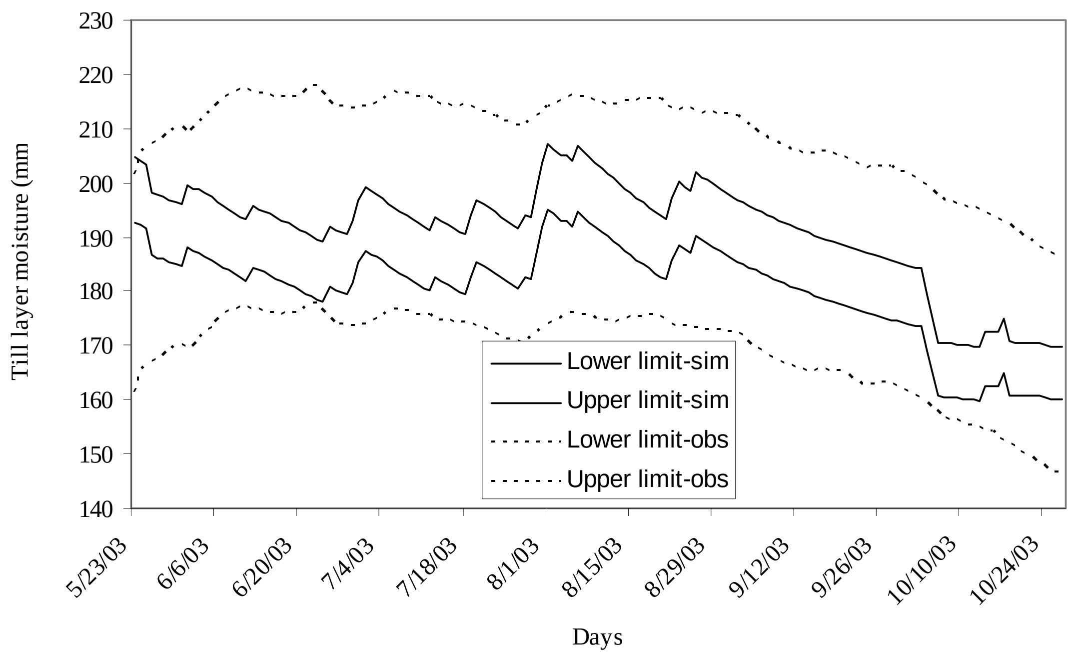


(b)

Figure 5. The SDW model results vs. uncertain measurements; (a) upper peat layer (200mm), (b) lower till layer (800 mm).



(a)



(b)

Figure 6. The uncertain results of the SDW model vs. uncertain measurements; (a) upper peat layer (200mm), (b) lower till layer (800 mm).