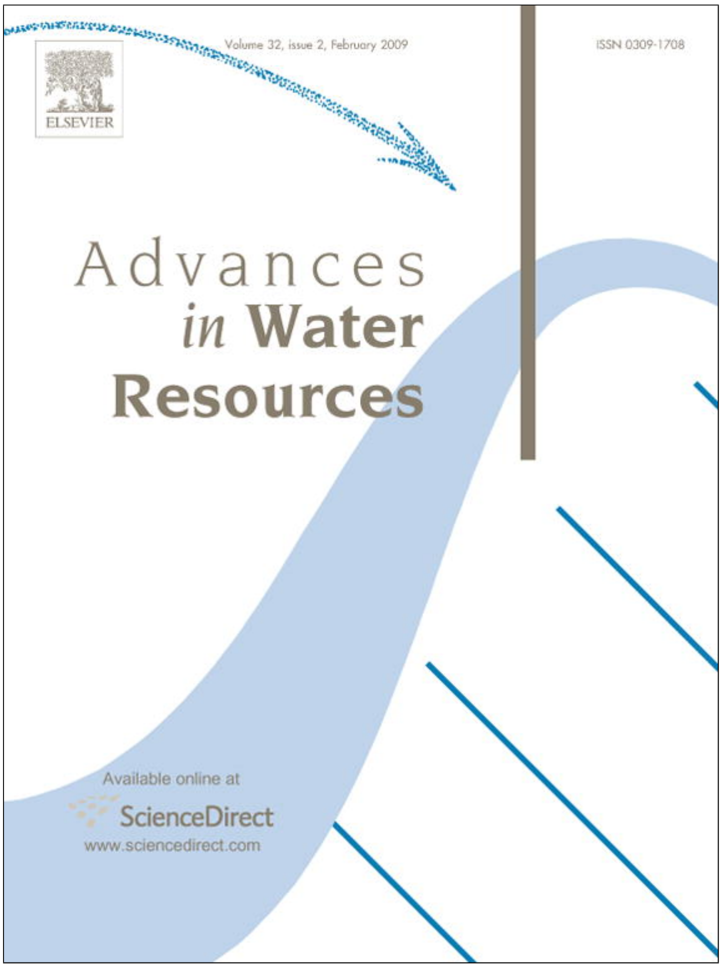
Volume 32, issue 2, February 2009
ISSN 0309-1708
ELSEVIER
Advances in Water Resources
Available online at
ScienceDirect
www.sciencedirect.com

Advances in Water Resources 32 (2009) 159–170

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) III: Scenario analysis

J.A. Huisman [a,*], L. Breuer [b], H. Bormann [c], A. Bronstert [d], B.F.W. Croke [e], H.-G. Frede [b], T. Gräff [d], L. Hubrechts [f], A.J. Jakeman [g], G. Kite [h], J. Lanini [h], G. Leavesley [i], D.P. Lettenmaier [h], G. Lindström [j], J. Seibert [k], M. Sivapalan [l,1], N.R. Viney [m], P. Willems [n]

[a] ICG-4 Agrosphere, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[b] Institute for Landscape Ecology and Resources Management, Justus-Liebig-University Giessen, Germany
[c] Institute for Biology and Environmental Sciences, Carl von Ossietzky University, Oldenburg, Germany
[d] Institute for Geoecology, University of Potsdam, Germany
[e] Integrated Catchment Assessment and Management Centre, The Australian National University, Canberra, Australia
[f] Afdeling Ecologie en Water, Lisec NV, Genk, Belgium
[g] Hydrologic Solutions, Pantymwyn, Flintshire, United Kingdom
[h] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA
[i] US Geological Survey, Denver, CO, USA
[j] SMHI, Norrköping, Sweden
[k] Department of Physical Geography and Quarternary Geology, Stockholm University, Sweden
[l] Center for Water Research, University of Western Australia, Crawley, Australia
[m] CSIRO Land and Water, Canberra, Australia
[n] Hydraulics Laboratory, Katholieke Universiteit Leuven, Belgium

ARTICLE INFO

Article history:
Received 13 July 2007
Received in revised form 27 June 2008
Accepted 29 June 2008
Available online 11 July 2008

Keywords:
Model intercomparison
Land use change
Reliability ensemble averaging (REA)
Ensemble modeling

ABSTRACT

An ensemble of 10 hydrological models was applied to the same set of land use change scenarios. There was general agreement about the direction of changes in the mean annual discharge and 90% discharge percentile predicted by the ensemble members, although a considerable range in the magnitude of predictions for the scenarios and catchments under consideration was obvious. Differences in the magnitude of the increase were attributed to the different mean annual actual evapotranspiration rates for each land use type. The ensemble of model runs was further analyzed with deterministic and probabilistic ensemble methods. The deterministic ensemble method based on a trimmed mean resulted in a single somewhat more reliable scenario prediction. The probabilistic reliability ensemble averaging (REA) method allowed a quantification of the model structure uncertainty in the scenario predictions. It was concluded that the use of a model ensemble has greatly increased our confidence in the reliability of the model predictions.

1. Introduction

Assessing the impact of land use change on water resources on a local, regional and global scale is a major challenge in hydrology. Typically, this is done by setting up a hydrological catchment model for the current land use, defining the expected changes in land use within a land use change scenario, re-running the model for the future land use and analyzing the differences between these two sets of simulations (e.g. [5,14,23,25] amongst others). Arguably, conventional practices for validation of hydrological models are not suitable for assessing the ability of a model to predict the impact of future environmental change [3]. Furthermore, the extensive calibration that is required to adapt most hydrological models to the current conditions makes one wonder whether these models are applicable for cases where the boundary conditions (e.g., climate, land use) have changed.

Ideally, predictions of the impact of land use change made with a specific model should be validated by comparison with data obtained after the land use change has occurred. However, such extensive validation is seldom performed (e.g. [33]). A main reason for this is the lack of suitable datasets for this purpose, despite the fact that there have been numerous experimental studies on the impacts of land use change in single and paired catchments (see [2,8,11,28]). As an alternative, Bathurst et al. [3] proposed to use a "blind" validation technique developed by Ewen and Parkin [13] in which the modeler is not allowed sight of the catchment output data so that the model cannot be calibrated for the

* Corresponding author. Tel.: +49 2461 618607; fax: +49 2461 612518.
E-mail address: s.huisman@fz-juelich.de (J.A. Huisman).

[1] Present address: Department of Geography, University of Illinois at Urbana-Champaign, IL, USA.

0309-1708/$ - see front matter
doi:10.1016/j.advwatres.2008.06.009

catchment under consideration. Such a blind validation test is harsh, and it is likely that many models would not pass such a test for a particular catchment.

Other environmental modeling communities also have to deal with predicting the impact of a change in boundary conditions. Most noticeable is the climate change community, which needs to predict the impact of rising CO_2 concentrations on the future climate. To deal with the uncertainty in these predictions of future climate, it has become common practice in this community to analyze scenarios with an ensemble of models instead of a single model [21]. There are two general approaches to the interpretation of ensemble modeling results. In the first so-called deterministic approach, an optimal combination of ensemble members is sought that results in better predictions than each single ensemble member (e.g. [12,16]). In the second so-called probabilistic approach, the single ensemble members are treated as possible (although not necessarily equally likely) realizations of the system response. In this probabilistic setting, quantitative methods to determine the uncertainty from an ensemble of scenario predictions have recently been proposed [17,26,29,30].

To increase the confidence in predictions of the impact of land use change on water resources, a set of hydrological models has been calibrated and validated on the same catchment and thereafter applied to the same set of land use change scenarios within the LUCHEM ("Assessing the impact of Land Use Change on Hydrology by Ensemble Modeling") project. This paper is the third in a series of four presenting the results of the LUCHEM project. The first paper ([10], this issue) describes the general set-up of the project, provides information on the relevant characteristics of the participating models and discusses the performance of these models for the current land use distribution. The second paper ([31], this issue) investigates the potential of deterministic model ensembles made up of some or all of the individual models to improve predictions of streamflow. The fourth paper ([7], this issue) investigates the effects of data resolution and spatial distribution of land use information on the simulated water balance for current catchment conditions and land use change scenarios for a subset of three models. In this third paper, the results of the application of these catchment models to the same set of land use change scenarios are analyzed. The aims of this paper are to (1) determine to what extent the models result in different simulations for the land use change scenarios and to understand the reasons behind these differences to the extent possible; (2) derive optimal (deterministic) scenario predictions of the impact of land use change from the ensemble of simulations and (3) quantify the uncertainty in the land use change predictions from the ensemble of simulations using probabilistic ensemble methods.

2. Materials and methods

2.1. Catchment description and available data

Land use change scenarios were developed and investigated for the low mountainous Dill catchment (693 km^2) in Germany. The Dill catchment is characterized by shallow soils underlain by fissured bedrock aquifers. Cambisols are the dominant soil types, covering >60% of the area. As a consequence of solifluction on periglacial slope deposits, the hydraulic conductivity of the soils is anisotropic with larger conductivities in horizontal direction. Because of the shallow soils and the anisotropic hydraulic conductivity, discharge in the Dill catchment is dominated by lateral flow. Mean annual rainfall varies between 700 and 1100 mm within the catchment and is not only dependent on height, but also decreases from west to east. The annual mean temperature is 8 °C.

The landscape is characterized by a heterogeneous small structured land use pattern. The land use is comprised of deciduous forest (29.5%), coniferous forest (24.8%), pasture (20.6%), urban areas (9.2%), fallow (9.1%), cropland (6.5%), and water (0.3%). The typical crop rotation in the region is winter barley, winter rape, and oats. Besides shallow soils and unfavorable climatic conditions, the high proportion of fallow land is a consequence of the socio-economic structure of the area. High opportunity costs result in a disproportionate number of part-time farmers. This leads to high machinery costs, which are further reinforced by relatively small average field sizes (~0.7 ha).

A detailed description of the data provided to each of the LUCHEM participants is given in a companion paper [10]. In summary, digital data on land use, soils and elevation were provided on a 25 m grid. The land use distribution in 1994–1995 was obtained from multi-temporal Landsat TM 5 images [24]. Soil information was derived from digitized 1:50000 soil maps [19]. Climatic data for the period of 01.01.1980 to 31.12.1998 from the German weather service (DWD) were also provided on a daily basis. Available data included precipitation (mm), wind speed ($m\ s^{-1}$), global radiation ($MJ\ m^{-2}\ d^{-1}$), air temperature (°C) and relative humidity (%). Precipitation was measured at 12 stations inside and six stations outside the catchment, whereas the other climatic variables were only recorded at two stations inside the catchment.

2.2. ProLand model

The land use change scenarios used in this study were derived with the ProLand (prognosis of land use) model [22,32]. ProLand assumes that land use patterns are a function of natural, economic, and social conditions in a landscape. It postulates land rent maximizing behavior of the land user. Land rent is defined as the sum of monetary yields including all subsidies minus input costs, depreciation, taxes, and opportunity costs for employed capital and labor. Depending on the economic and ecological boundary conditions, the model calculates the land rent for a set of agricultural and forestry land use systems for each parcel of land. ProLand only simulates one type of forestry, namely mixed forests consisting of deciduous and coniferous trees (*Fagus sylvatica* beech, 40%; *Quercus* spp. oak, 20%; *Picea abies* spruce, 30%; *Pinus sylvestris* pine, 6%; and *Pseudotsuga menziesii* Douglas fir, 4%). This forest production system resembles the dominant forest species distribution in the landscape investigated in this work.

The ProLand version used in this scenario analysis is based on a pixel approach. Hence, every simulated parcel of land is equivalent to the area of a 25 m pixel. The land use system with the highest land rent is selected as the optimal land use for the pixel under consideration. Farmer sentiments and costs associated with land use change are not considered. In addition, ProLand does not consider neighborhood relationships. Thus, it can happen that a pixel with a particular land use is surrounded by different land uses. This restriction is model specific and may sometimes produce unrealistic land use patterns ("an island of cropland in the forest"). The output of the ProLand model consists of data describing the economic performance of the calculated set of land use systems and a spatially explicit map of the optimal land use distribution given the provided boundary conditions. Further details of the model set-up and performance are given in [22,32].

2.3. Field sizes scenarios

As in many other regions of Europe, the inheritance system has had a tremendous effect on average field sizes in the Dill catchment. Typically, fields were split equally amongst the inheritors. As a result, the average field size decreased more and more. In the Dill catchment, the average field size decreased to less than 0.1 ha in the middle of the 20th century in some municipalities. The average farm size was also very low with 15 ha.

From an economic point of view, the average field size is one of the major factors affecting the costs in agriculture. Large machinery cannot be used on small fields and purchasing machinery is only profitable if farms have an adequate size. The reallocation of small fields to create larger fields allows the use of more efficient machinery and could, therefore, reduce labor costs. In Germany, land reallocation has been conducted since the late 1960s to improve the structure of the agrarian sector.

The ProLand model was used to investigate the potential effects of reallocation on land use distribution. A sequence of land use change scenarios with average field sizes of 0.5, 1.5 and 5.0 ha were simulated. Open water and urban areas were held constant and were not affected by land use change. The resulting land use distribution for the three field size scenarios is shown in Fig. 1. Land use is dominated by forest and pasture in the 0.5 ha scenario. An increase of the average field size to 1.5 ha results in a strong increase of cropland at the expense of pasture and forest in the western and northern parts of the catchment. A further increase of the average field size to 5.0 ha leads to a more patchy land use distribution especially in the eastern part of the catchment where forested areas are now converted to cropland. Overall, the aggregation of fields from an average field size of 0.5 ha to a field size of 5.0 ha leads to an increase of cropland to more than 33% of the area accompanied by a reduction of pasture area by 11% and a reduction of forest by 22% of the area (Table 1).

2.4. Modeling approach within LUCHEM

In the LUCHEM project, 10 different models simulated hydrological fluxes of the Dill and three of its tributaries, the Aar, the Dietzhölze and the Obere Dill (see Table 2). Details on the model instructions for the LUCHEM participants are described in detail in [10]. In short, two sets of input data were used to evaluate the performance of the different models in terms of their ability to simulate hydrological fluxes under current conditions. For both data sets, the calibration period was 1983–1989, whereas the validation period was from 1990 to 1998. The first data set was provided prior to the LUCHEM workshop in order to set up the different models. However, during the workshop it became clear that a more homogenized dataset was necessary to ease the interpretation of the model results. Details on this homogenized dataset and the performance of the models for the current conditions are again provided by Breuer et al. [10].

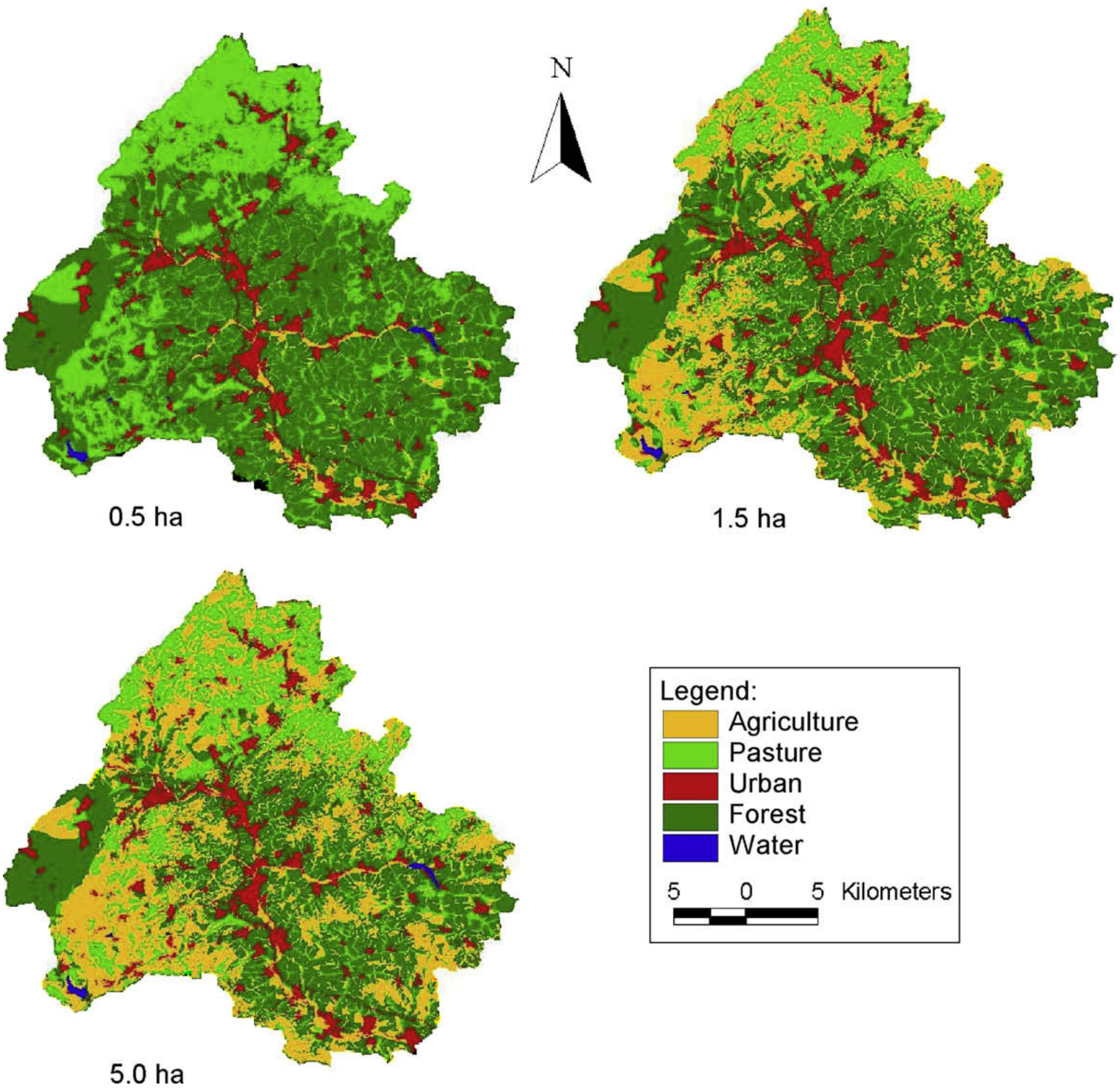

Fig. 1. Land use scenarios as predicted by the ProLand model assuming allocation of fields to average field sizes of 0.5, 1.5 and 5.0 ha.

Table 1
Land use distribution of the ProLand land use scenarios

	Land cover (%)		
	Cropland	Pasture	Forest
Dill			
Baseline	6.5	20.6	54.3
0.5 ha	2.8	31.7	56.0
1.5 ha	27.1	17.6	45.8
5.0 ha	35.9	20.5	34.1
Aar			
Baseline	11.3	19.2	52.8
0.5 ha	3.3	21.7	68.3
1.5 ha	20.8	13.9	58.9
5.0 ha	34.5	16.2	42.8
Dietzhölze			
Baseline	3.2	17.0	60.9
0.5 ha	0.6	70.9	19.4
1.5 ha	31.7	47.3	11.8
5.0 ha	35.2	49.7	6.0
Obere Dill			
Baseline	2.9	17.0	64.1
0.5 ha	0.5	58.0	36.2
1.5 ha	38.3	32.0	24.2
5.0 ha	41.4	42.4	10.8

Urban and water land covers are not reported because they are identical in the baseline and the scenarios.

The results of the scenario analysis presented in this paper are based on the homogenized data set. Evaluation of the land use change scenarios was performed for the entire simulation period (1983–1998). All model groups provided results for the three land use change scenarios for the Dill catchment. Apart from the MIKE-SHE model group, all other groups also provided results for the three subcatchments.

After model set-up, model calibration and model validation for the current conditions, the next step in a typical scenario analysis of the impact of land use change on hydrology is the implementation of the scenarios. For distributed models, this step most often involves running the model with the parameterization obtained for the current conditions, but with different fractions and a different spatial arrangement of the land use classes. For lumped models, it involves deriving a new parameterization based on the calibration results and the new land use fractions. Obviously, a key factor for the simulation of land use change effects is to account for differences in evapotranspiration (ET) of different land use classes.

Within the LUCHEM project, the models differ with respect to the spatial and temporal representation of key factors determining actual ET. Key differences between the 10 models are summarized in Table 2. First of all, the spatial scale (i.e., the representative area for which calculations are performed) ranges from 1 ha to areas much larger than 1 km^2. In addition, some of the models are fully distributed (e.g., DHSVM) whereas other models are lumped (e.g., IHACRES). The impact of spatial scale on the simulation results is explored in a companion paper [7] for a selection of distributed models. Second, the models used different levels of aggregation for land use. For example, HBV and IHACRES did not distinguish between crops and pasture but used a land use type "open vegetation" instead. Finally, different methods were used to calculate potential ET. The methods ranged from the Penman–Monteith method, which requires detailed meteorological information and plant-specific parameters (e.g., canopy resistance), to much less complex temperature-driven methods. For all models except IHACRES, the temporal change of LAI was a key parameter in the temporal representation of ET. The IHACRES model used empirically derived relationships between actual ET and land cover that are based on an investigation by Zhang et al. [34]. To avoid differences in model predictions due to different model parameterizations of LAI, a simplification of the SWAT crop growth model was used to derive mean monthly maximum LAI for the land use classes cropland, pasture and forest. This average monthly LAI was used by all models considering temporal changes in LAI as the driving factor for potential ET, except for the SWAT model where the LAI was based on the actual growing conditions within a year. However, differences obviously remain between the models. For example, there are different implementations for the Penman–Monteith method. DHSVM calculates the canopy resistance based on temperature, vapour pressure deficit and soil water content, whereas SWAT only considers vapour pressure deficit. There is no consensus on how to implement the Penman–Monteith method, which inevitably leads to differences.

The next critical issue causing differences between the models is the calculation of actual ET. Most models calculate a so-called reference (potential) ET, which describes the amount of ET under well-watered conditions. In a next step, it is determined how much the soil can deliver given this demand. Typically, this is done for each layer in the model by assigning a potential root water uptake to each layer, for example based on the root length density within each layer. If the available water is less than the demand, the actual amount of ET is reduced. Depending on the number of soil layers, the method to calculate vertical flow (e.g., fill and spill, Darcy equation), the assumed (or modeled) root length density, and whether one soil layer can compensate water deficits occurring in other layers, the actual ET calculated by different models can vary strongly. Although the amount of water available for ET depends on the rooting depth and density, we did not homogenize this information between the models. It should also be noted that not all models use this demand-and-supply concept. For example, DHSVM directly calculates actual ET by modifying the crop resistance in response to water shortage. Finally, interception is an important contribution to actual ET, especially for forests. The parameterization of this process was also based on vegetation-type dependent default values within each model.

Table 2
Overview of spatial scale, land use type and potential evaporation method used in the scenario runs for each model

Model	Spatial scale	Land use types	ET_{pot}
DHSVM	100 m	Forest, crops, pasture	Penman–Monteith
MIKE-SHE	200 m	Forest, crops, pasture	Penman–Monteith
TOPLATS	100 m	Forest, crops, pasture	Penman–Monteith
WASIM	200 m	Forest, crops, pasture	Penman–Monteith
SWAT	HRU in 52 subcatchments	Forest, crops, pasture	Penman–Monteith
PRMS	HRU in 25 subcatchments	Forest, crops, pasture	Jensen–Haise
SLURP	Fraction of land use in 39 subcatchments	Forest, crops, pasture	Penman–Monteith
HBV	Fraction of land use in 10 subcatchments	Forest, open vegetation (crops + pasture)	Penman–Monteith, temperature-driven monthly factors
LASCAM	Fraction of land use in 29 subcatchments	Forest, crops, pasture	Based on solar radiation assuming linear trend with latitude
IHACRES	Fraction of land use	Forest, open vegetation (crops + pasture)	Empirical, T- and P-driven factors [34]

Of course, ET is not the only flux that changes with a change in land use. For example, the partitioning of precipitation at the land surface and the partitioning between interflow and groundwater recharge might also be affected. However, we consider this to be of secondary importance for the Dill catchment [6,20], and therefore we refrain from providing a detailed description of these processes for each model. For this, we refer to the original publications, which are referenced in a companion paper [10].

2.5. Deterministic and probabilistic ensemble methods

Deterministic ensemble predictions may be constructed in a number of ways. The simplest approach is to take the raw mean of the set of model predictions for each day. Other simple approaches are to take the daily median or to use a trimmed mean. More complex approaches are to determine the weights for each model based on model performance within a specified calibration period and using these "optimal" weights to make predictions (e.g. [16]). Alternatively, it is also possible to determine the model weights from multiple linear regression or use conditional ensembles where the model weights depend on the state of the system (e.g. [1]).

A large number of deterministic ensemble approaches is compared in a companion paper [31]. Their results showed that of the simple ensemble methods, the trimmed mean (i.e., the mean of four to six central model predictions for each day) produced a higher model efficiency than the raw mean, the median or any single member of the ensemble. In addition, they showed that the more complex combination methods often performed better during the calibration period, but that the performance during the validation period was similar or worse than the trimmed mean. Therefore, we decided to use a six-member trimmed mean of daily streamflow as the best deterministic combination for the Dill catchment. For the gauged subcatchments of the Dill, a five-member trimmed mean was used because MIKE-SHE did not provide scenario results for these gauges.

Deterministic ensemble methods result in a single "optimal" prediction based on some combination of single ensemble members. However, the model ensemble can also be interpreted in a probabilistic sense. Such a probabilistic interpretation was presented by Giorgi and Mearns [18] based on the work of Giorgi and Mearns [17]. They called their method reliability ensemble averaging (REA) and it was used to quantify the average and uncertainty range of simulated climate change from an ensemble of different atmosphere–ocean global circulation models. For this study, the REA method was adapted to hydrological scenario analysis.

The REA method relies on two general "reliability criteria" to assess the reliability of simulated changes in mean annual discharge. In hydrological terms, the first is based on the ability of the models to predict current river discharge. Models that represent the current discharge well can be expected to produce more reliable scenario predictions. The second criterion is based on the convergence of simulations by different models for a given scenario. Greater convergence of the scenario predictions implies a robust change prediction, despite differences in model structure. In mathematical terms, these two reliability criteria can be stated as a model reliability factor, R_i for a particular model i as [17]

$$R_i = R_{B,i} \times R_{D,i} = \left(\frac{\varepsilon}{\mathrm{abs}(B_i)}\right)\left(\frac{\varepsilon}{\mathrm{abs}(D_i)}\right) \tag{1}$$

In this equation, R_B is a factor that measures the model reliability as a function of the model bias (B) in simulating present-day discharge. Model bias is defined in [10]. The higher the bias, the lower is the model reliability. Here, the bias is calculated for the entire 16-yr simulation period (1983–1998) of the LUCHEM project. R_D is a factor that measures the model reliability in terms of the distance (D) of the change for a specific model from the REA average change ($\Delta\widetilde{Q}$) and is defined as

$$D_i = \Delta Q_i - \Delta\widetilde{Q} = \Delta Q_i - \frac{\sum_{i=1}^{N} R_i \Delta Q_i}{\sum_{i=1}^{N} R_i} \tag{2}$$

where ΔQ_i is the simulated change for ensemble member i and N is the total number of ensemble members. Again, the higher the distance, D, the lower is the reliability. Since the REA average change is not known a priori, it is obtained in an iterative way as described in detail by Giorgi and Mearns [17]. To calculate the uncertainty range around the REA average change, the weighted root mean square difference (RMSD) is calculated:

$$\mathrm{RMSD} = \left(\frac{\sum_{i=1}^{N} R_i (\Delta Q_i - \Delta\widetilde{Q})^2}{\sum_{i=1}^{N} R_i}\right)^{1/2} \tag{3}$$

Assuming that the probability density function is somewhere between uniform and Gaussian, the REA average change plus and minus the RMSD can be interpreted as a 60–70% confidence interval [17].

The parameter ε in Eq. (1) is a measure of natural variability in 16-yr average discharge. To calculate ε, we computed 16-yr moving averages for the discharge time series of the catchments under consideration. Typically, 35–45 yr of discharge measurements were available. As suggested by Giorgi and Mearns [17] for short time series, we then used the difference between the minimum and maximum values of the moving average time series after linear detrending as a measure of the natural variability. An important aspect of the REA method is that R_B and R_D are set to 1 when B and D are smaller than ε, respectively. Essentially, this implies that a model prediction is reliable when both its bias and its deviation from the weighted ensemble average are lower than the natural variability.

The concept of the reliability factor can also be used to estimate the probability of the predicted changes from a model ensemble. Giorgi and Mearns [18] suggested that the likelihood of a model-simulated change (P_i) is proportional to the reliability factor defined above. The normalization of the likelihood yields the definition

$$P_i = \frac{R_i}{\sum_{i=1}^{N} R_i} \tag{4}$$

In other words, it is assumed that the change simulated by a more reliable model (in the sense of Eq. (1)) is more likely to occur. From Eq. (4) it follows that, for a given land use change scenario, the probability of a discharge change exceeding a threshold ΔQ_{th} is given by

$$P_{\Delta Q_{\mathrm{th}}} = \sum_i P_i = \sum_i \frac{R_i}{\sum_{i=1}^{N} R_i}, \quad \Delta Q_i \leqslant \Delta Q_{\mathrm{th}} \tag{5}$$

Thus, the probability that a certain threshold is not exceeded equals the sum of the probability of all models that do not exceed the threshold. If all models are equally reliable, this equation corresponds to the probability analysis method proposed by Räisänen and Palmer [26].

3. Results and discussion

3.1. Scenario analysis

Fig. 2 presents the results of the scenario predictions as compared to the baseline run for the mean annual discharge of the Dill catchment. As discussed in a companion paper [10], the baseline runs show a considerable variation in bias in cumulative volume

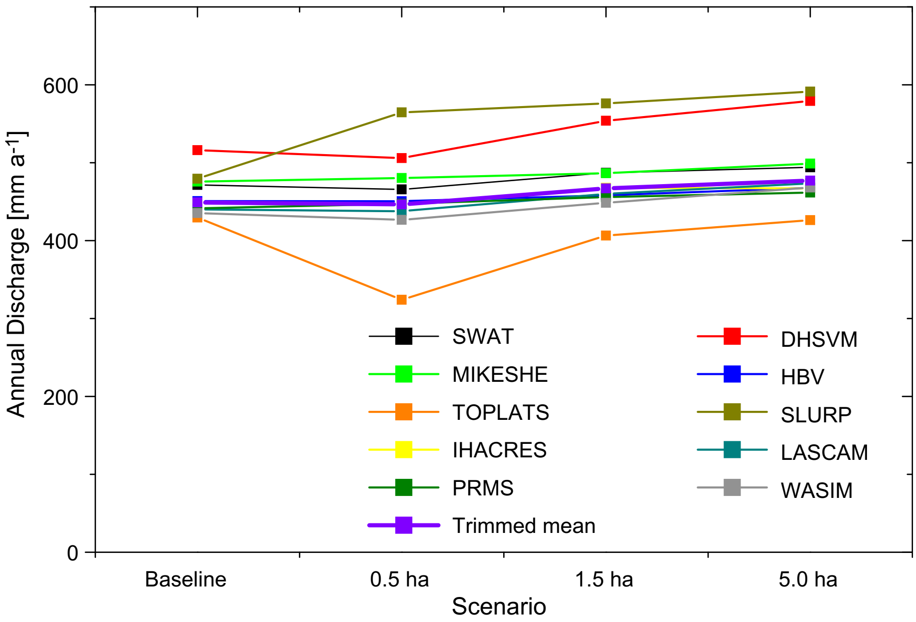

Fig. 2. Annual discharge in the Dill catchment for the baseline and three field-size scenarios. Results are presented for all participating models and a deterministic ensemble method (trimmed mean).

with respect to the observed discharge (437 mm yr^{-1}). However, based on an analysis of several performance metrics it was concluded that no single model outperformed the others and it was, therefore, suggested that all models contain useful information that could be exploited in an ensemble analysis [10,31]. It can also be seen in Fig. 2 that all models predict that mean annual discharge will increase with an increase in field size. As with the IPCC climate predictions [21], this agreement with respect to the direction of change increases our confidence in the reliability of these scenario predictions. However, it is obvious that the bias in the baseline runs is propagated within the scenario analysis and that the variation in bias between the models is responsible for a significant fraction of the variation in the scenario predictions. This suggests that in order to reduce the uncertainty in the scenario predictions, it might be just as important to focus on reducing the model structure uncertainty (i.e., reducing the variability in the bias). In the following, we interpret the results of the scenario analysis by looking at the change in mean annual discharge relative to baseline runs (Fig. 3). It is quite common to analyze scenarios in this way. For example, IPCC climate change predictions are also reported relative to baseline runs and not relative to climatic observations [21]. However, there is no doubt that confidence in change predictions would further increase when model structural uncertainty is reduced.

Although all models predict the same direction of change, the rate of increase differs between models (Figs. 2 and 3). For example, the lowest difference between the 0.5 and the 5.0 ha scenario was 14.5 mm yr^{-1} and the highest difference was 102.1 mm yr^{-1} for the Dill catchment. Based on Figs. 2 and 3, it appears that 7 of the 10 models provide rather similar predictions for the Dill catchment, whereas the three remaining models perform quite differently (DHSVM, TOPLATS and SLURP). Although these three models simulated the lowest (TOPLATS) and highest (DHSVM, SLURP) mean annual discharge in the baseline run, it only becomes apparent from the current scenario analysis that these models behave differently. Obviously, the possibility to compare scenario predictions has turned out to be a big advantage of using an ensemble of models, highlighting differences in model performance that were not fully apparent previously.

Fig. 3 also presents the results for the three subcatchments of the Dill. The results for the subcatchments again confirm that all models predict an increase in discharge with increasing field sizes. The difference between the 0.5 and 5.0 ha field size scenario for the subcatchments is also similar to the scenario predictions for the Dill. For example, the median increase in discharge going from the 0.5 ha to the 5.0 ha scenario is 31 mm yr^{-1} for the Dietzhölze and 27 mm yr^{-1} for the Dill catchment. However, the models predict a much wider range of differences between the scenario runs and the baseline run for both the Dietzhölze and the Obere Dill. Finally, the change in mean annual discharge for a dry year, a wet year and for the summer and winter periods for the three scenarios relative to the baseline is shown in Fig. 4 for the Dill catchment. Again, the direction and magnitude of change are similar for seven models, and DHSVM, SLURP and TOPLATS perform differently.

From water balance considerations, it is obvious that differences in mean annual discharge must be attributed to differences in actual ET between the models because variability in mean annual precipitation was relatively small [10]. To understand differences in simulated actual ET between the models, it is important to note that (1) the main land use change within the scenarios is an increase in cropland at the expense of pasture and/or forest; (2) the current land use used in the baseline run is most similar to the 0.5 ha scenario for the Dill catchment; (3) the current land use for the Dietzhölze and Obere Dill catchment deviates strongly from the land use in all three scenarios; and (4) the baseline run considers both coniferous and deciduous forest, whereas the scenario analysis considers mixed forest only.

The bulk of the models show little sensitivity to changes in land use. In fact, it seems likely that some of the simulated differences are insignificant given the typical uncertainty in model parameterization. However, a quantification of the uncertainty in the predicted changes for individual ensemble members as was done in [9] is beyond the scope of this study. The lack of sensitivity to land use change indicates that actual ET estimates for the different land use classes are quite similar for these models. For example, the SWAT model predicts mean annual actual ET values of 477 mm for forest, 487 mm for pasture and 396 mm for crop land for the baseline run. These mean actual ET values are similar for the land

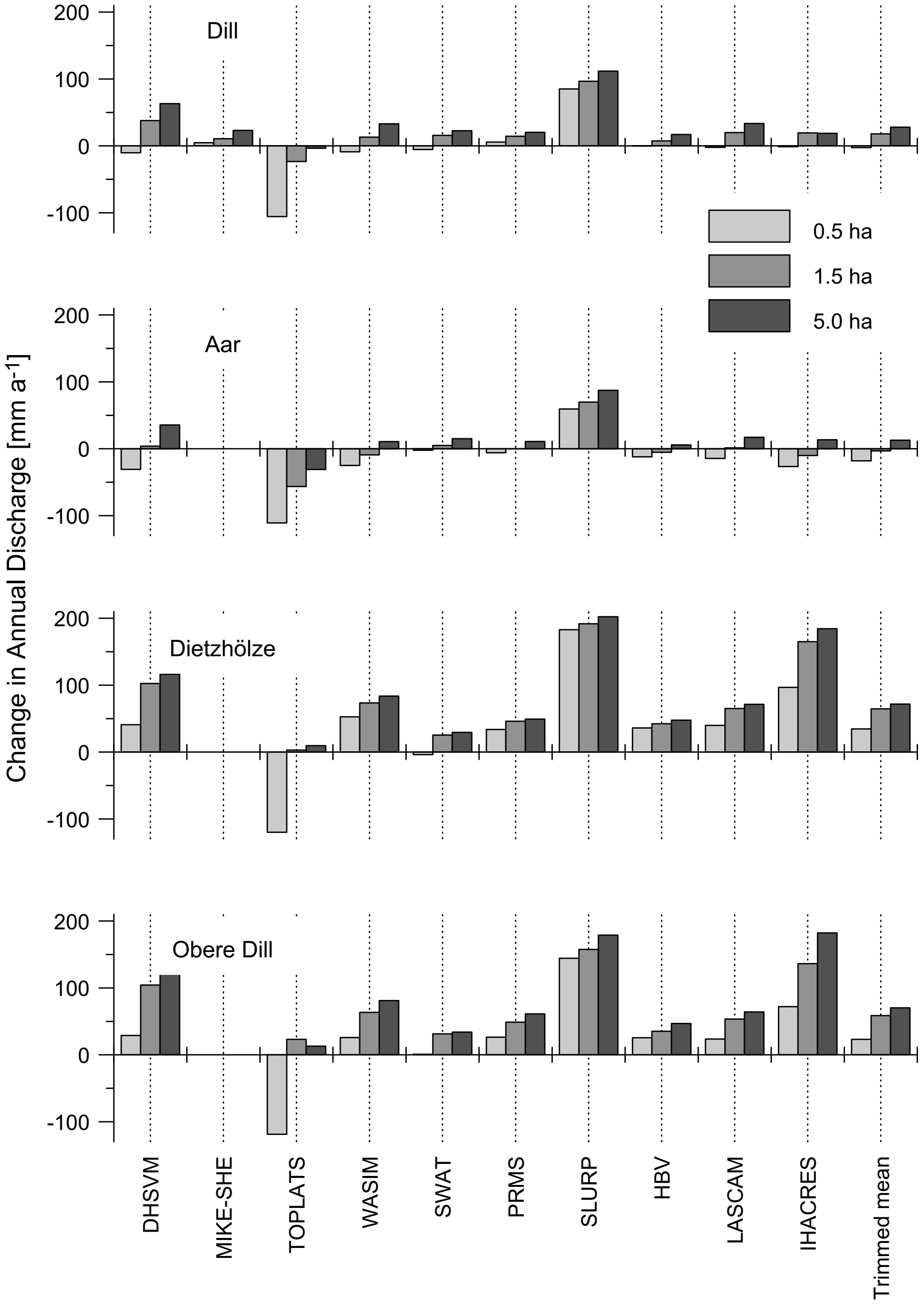

Fig. 3. Predicted change of annual discharge for all subcatchments and land use change scenarios as predicted by 10 hydrological models and a deterministic ensemble method (trimmed mean). Change in annual discharge is expressed as the difference from the baseline scenario. Models are ranked according to their complexity from left (complex) to right (simple). Ranking is discussed in more detail in [8].

use scenarios. The difference of 80–90 mm between cropland and the other land uses nicely agrees with the predicted ~30 mm increase in discharge with a 30% increase in cropland between the 0.5 and 5.0 ha scenarios for the Dill catchment. The fact that this simple calculation works out so well foreshadows the finding in [7] that the spatial arrangement of land use does not strongly affect the scenario results for a subset of the semi-distributed and distributed models.

As noted above, three models have deviating scenario predictions. Fig. 2 shows that there is a strong difference between the baseline and scenario run for two of these models (SLURP and TOPLATS). For SLURP, we completely attribute this to the presence of mixed forests in the scenario runs and deciduous and coniferous forest in the baseline run. Apparently, the mixed forest parameterization in SLURP underestimates the actual ET for this land use type. This conclusion is based on the following three observations: First, the difference between baseline and scenario runs is larger than the changes between scenarios for all catchments despite large changes in land use between scenarios. Second, there is a large increase in discharge between the baseline run and the scenario runs for the Dill catchment despite similar land use fractions. Finally, the increase between baseline and scenario runs is strongest for the Obere Dill and Dietzhöze catchment where the change in forest is largest between the baseline and scenario runs.

The deviating results for TOPLATS are also partly related to an inadequate parameterization of the mixed forest. In the baseline run, mean annual actual ET is 330 mm for deciduous forest and 609 mm for coniferous forest. The mean annual actual ET of mixed

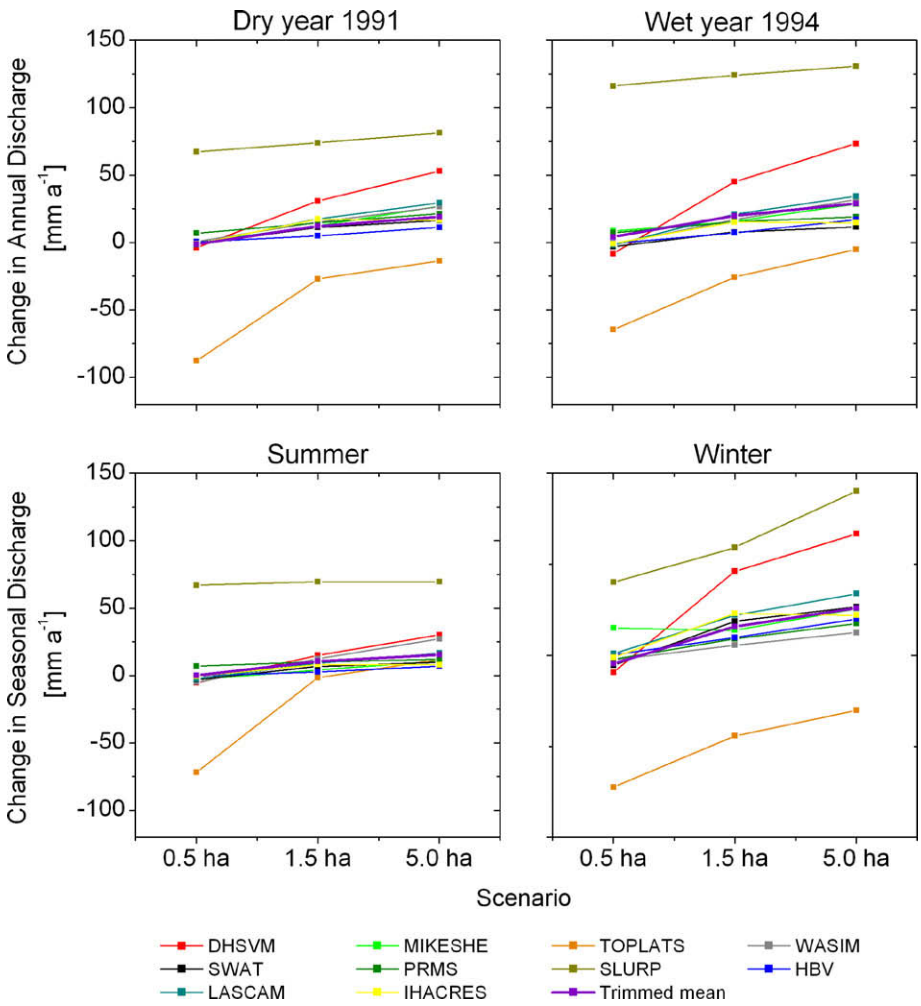

Fig. 4. Effect of dry and wet years as well as season on predicted discharge for the Dill catchment. Winter season is defined from 01.11. to 30.04. Dry year and wet year were selected based on the years with lowest and highest annual precipitation, respectively. Results are presented for all participating models and a deterministic ensemble method (trimmed mean).

forest is 554 mm in the scenarios, which does not adequately reflect the proportions of deciduous and coniferous forest in the baseline run. This explains the difference between the baseline run and the 0.5 ha scenario. The relatively strong increase in discharge going from the 0.5 ha to the 5.0 ha scenario is related to differing mean actual ET rates for each land use. TOPLATS calculated a high mean annual ET of 760 mm for pasture and a low mean annual rate of 323 mm for cropland. Again, these mean annual rates were relatively independent of the scenario under consideration [7]. Combining these rates with the land use fractions for the 0.5 and 5.0 ha scenarios (Table 1) results in a predicted decrease in mean annual ET of 100 mm. This corresponds very well with the increase in discharge of 102 mm yr^{-1} presented in Fig. 3. The high rate for pasture as compared to cropland was attributed to the high LAI (1.5–5.3) and the low stomatal resistance (72 s m^{-1}) used for pasture.

The third model that provides deviating results is DHSVM. This model has the largest bias for the baseline run. It should be noted that this was introduced by the process of data homogenization [10] when the yearly course of LAI was prescribed. With the prescribed LAI, the actual ET of forest was considerably reduced and a positive bias in mean annual discharge was introduced. In addition, DHSVM reacts more sensitive to changes in land use than the other models (Figs. 2 and 3), which is related to differences in actual ET in the winter period (Fig. 4). The high sensitivity in winter indicates strongly differing mean actual ET rates between the perennial plants and cropland in this period. Since ET rates are low in winter for cropland because of bare soils and low energy, it is concluded that the simulated mean actual ET of perennial plants in the winter period must be higher in DHSVM than in the other models in the ensemble.

From the above, it is clear that we can interpret the differences in scenario predictions between the models in terms of different ET estimates for each land use type. The question then arises whether the "minor changes" prediction of the bulk of the models is realistic. We believe that this is the case. For large parts of the year, the Dill catchment is energy-limited. Given the conditions within the catchment, it is reasonable to conclude that actual ET rates for cropland, pasture and forest are relatively similar and in the range of 400–600 mm yr^{-1}. These small contrasts in actual ET will cause only small to moderate changes in discharge, even in the case of substantial changes in land use. As outlined above, the models differ considerably in their approaches to calculate actual ET. However, in a dominantly energy-limited environment these model structural differences are not propagated into the scenario predictions. Therefore, it might be interesting to repeat this kind of ensemble scenario analysis for a water-limited environment, where stronger differences between models would be expected.

One might argue that in energy limited environments runoff is an important component of the water balance and how the models capture the runoff generation processes will also be important. Our perceptual model of runoff generation in the Dill catchment consists of a dominant interflow component in addition to minor fractions of quick (e.g., surface runoff) and slow (e.g., base flow) flow components. Because these flowpaths are determined by the geomorphology and geology and not so much by the land use, we believe that it is reasonable that we consider the impact of land use change on the flowpaths and the subsequent feedback to actual ET of secondary importance. As argued in [6,20], this is a peculiarity of the Dill catchment and should not be extended to other environments.

Finally, it is interesting to note that all three models providing deviating results use the highly parameterized Penman–Monteith model, and that the development of two of these three models (DHSVM and TOPLATS) has focused on a physically-based representation of soil–vegetation–atmosphere relationships. Arguably, the results for these models partly reflects the shortcomings of the current calibration practice, where catchment models are mainly calibrated to discharge without considering internal catchment processes, such as ET of different land use types [4]. It is now well recognized that calibration to a single gauge only does not guarantee that internal processes are simulated correctly [15,27]. In this study, only discharge data were available for the main catchment outlet and the outlets of three subcatchments. Therefore, the parameterization of the ET modules and the validation of the ET results were largely based on experience of the modelers. In future ensemble studies of this kind, more attention should be paid to the validation of ET predictions for single land use types. This can be achieved by direct comparison with measurements (e.g., derived from remote sensing or micrometeorological methods) or by selecting a catchment with gauged subcatchments with strongly varying land uses.

3.2. Deterministic ensemble results

In a companion paper [31], it was argued that an optimal combination of model simulations could average out model structural error and provide an estimate of the most probable state of the system. This is explored in this section. Figs. 2–4 show the results for a six-member trimmed mean of daily streamflow for the Dill catchment and a five-member trimmed mean for the gauged subcatchments. The second paper in this series [31] has shown that this simple ensemble method results in superior predictions of streamflow as compared to the more complicated combination methods for the Dill catchment (model efficiency of 0.937 and a bias of 3.2% for the validation period). Although no validation is possible for the scenario results, we assume that this trimmed mean also provides the best estimate of the expected changes in the scenario analysis.

Figs. 2–4 show that the ensemble method results in predictions close to the median of scenario predictions. Although this seems obvious because a trimmed mean discards extreme predictions that might distort the raw mean, it is not self-evident because the trimmed mean is calculated on a daily basis. All models contributed to the trimmed mean for a considerable amount of time. For the Dill catchment, the model with the lowest average contribution for all scenarios (SLURP) is still selected 38% of the time. The highest contributing model (PRMS) was selected 79% of the time. The contributions of each model to the trimmed mean ensembles for the gauged subcatchments were similar. For the DHSVM model applied to the Dill catchment, there was a strong decrease in contribution going from the trimmed mean ensemble for the actual conditions (68% of the time) to the average contribution to the trimmed mean ensembles for the land use change scenarios (52% of the time). This reflects that the scenario predictions of this model deviate more from the other models in the scenario analysis than in the application to the current conditions. The contributions of the other models were approximately equal for the actual conditions and the land use change scenarios.

The deterministic ensemble predictions presented here are one option to summarize the results of the model ensemble. It has been shown that multi-model ensemble methods typically outperform all individual models [31], and references therein and, therefore, it can be expected that scenario predictions obtained here are also more reliable. However, the deterministic approaches do not consider the opportunity to use the ensemble results to not only derive an optimal scenario prediction but to also derive an estimate of the uncertainty in these scenario predictions. Clearly, this would be a significant contribution, and, therefore, a probabilistic method that allows such uncertainty quantification is discussed next.

3.3. Reliability ensemble averaging (REA)

To quantify the uncertainty in the scenario predictions, the model ensemble was used to derive approximate probability density functions (PDF). Fig. 5 presents the PDF describing the probability that the change in discharge will exceed a certain threshold for the Dill and the Dietzhölze for the 5.0 ha scenario. Two methods were used to derive these PDF. In the first method, all models were weighted equally and in the second method the REA method was

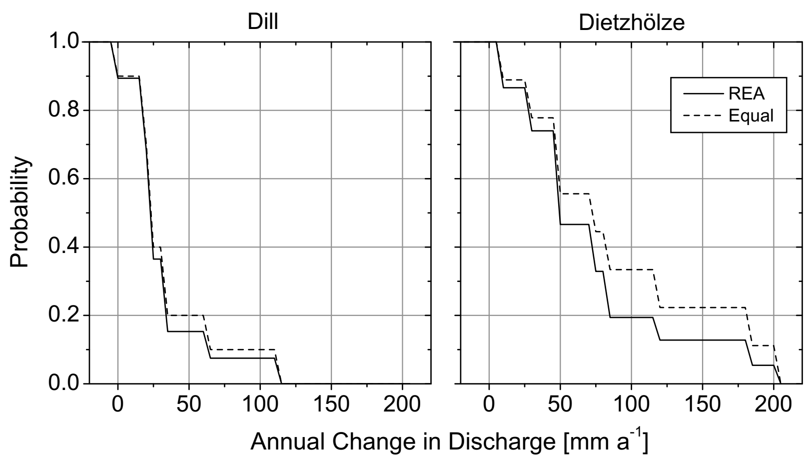

Fig. 5. Comparison of the effects of equal weighing and REA for the probability distribution of changes in annual discharge. Predictions shown are for the 5.0 ha land use scenario for the Dill and the Dietzhölze catchment.

used to weight the different models. The REA method assumes that models that perform poorly for the present-day discharge (see [10]) and models that deviate strongly from the average change predicted by the ensemble are less reliable. As discussed in Section 3.1, there are three models that provide deviating scenario predictions for reasons discussed earlier. Since this deviating behavior was not evident from the present-day conditions, it seems appropriate to consider both the present-day performance and the similarity to the scenario predictions of the other models in the ensemble analysis. Fig. 5 shows that the difference between the two methods is small for the Dill catchment. For example, both methods predict that there is a 50% probability that the change in annual discharge will exceed 23 mm yr^{-1} for this land use change scenario. The small difference can be explained by the coherency of the scenario predictions and the relatively low bias in simulating present-day discharge. One model was deemed slightly less reliable because the model bias for the actual conditions exceeded the natural variability (DHVSM, $R = 0.74$), and the reliability of another model was decreased because its scenario prediction deviated more from the ensemble mean than expected from the natural variability (SLURP, $R = 0.71$). The difference between the two methods is much larger for the Dietzhölze. For this catchment, four models were less reliable because of a high bias and three models were less reliable because of strongly deviating scenario predictions (R values ranged from 0.39 to 1.00). It should be noted that there is a weak correlation between bias and deviation from the ensemble mean of the scenario predictions, which means that the reliability of some models is reduced for both reasons. The observed differences in Fig. 5 are in line with the results presented in [10], who showed that the individual models performed better for the Dill catchment than for the subcatchments.

The impact of the REA method on the PDF is evident from Fig. 5b. The reduced reliability for scenario prediction of seven models has led to an assumed smaller likelihood for the scenario predictions of these models. In case of the Dietzhölze, this caused a tightening of the PDF as compared to equal weighting of each scenario prediction. Because models with large deviations from the average scenario change are also deemed less reliable, the REA method will often cause such a tightening. However, if models that predict close to the average scenario change are less reliable because of a high bias, the PDF can also be wider for the REA weighting as compared to equal weighting of all scenario predictions. In the following, we apply the REA method to the remaining scenario predictions because it seems a convenient way to deal with varying model quality and deviating scenario predictions.

Fig. 6 presents the PDF for the four catchments and three field size scenarios. Table 3 presents the mean, the trimmed mean, the REA average and the REA RMSD of the ensemble of scenario predictions. Fig. 6 and Table 3 show that the ensemble predicts that the mean annual discharge will increase with increasing field size, as was already observed in Figs. 2 and 3. For example, the REA mean of the difference between scenario and baseline run in the Dill catchment increases from −1.9 mm yr^{-1} for the 0.5 ha scenario to 30.8 mm yr^{-1} for the 5.0 ha scenario (Table 3). Table 3 also shows that the REA average is closer to the trimmed mean prediction than to the raw mean prediction. It should be noted that this raw mean ensemble prediction was shown to have a rather large bias and less than optimal model efficiency in a companion paper

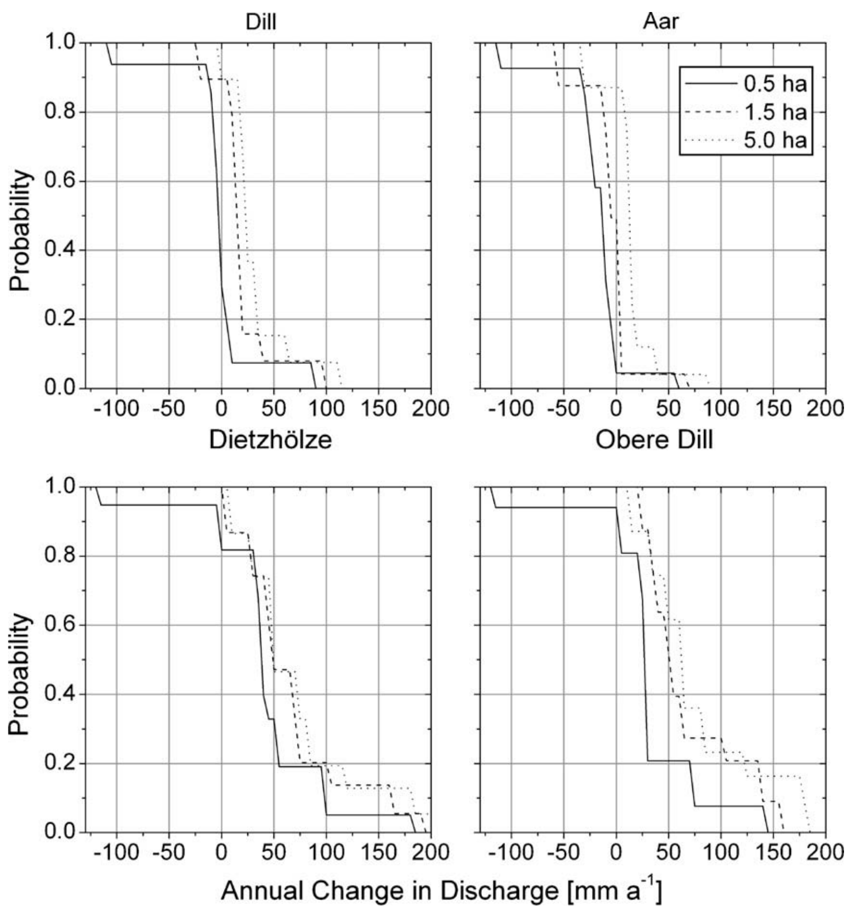

Fig. 6. Probability of change in annual discharge for the Dill catchments and three of its subcatchments. Probability was estimated using the reliability ensemble averaging (REA) method.

Table 3
Trimmed mean, mean, REA mean and REA root mean square difference (RMSD) for the Dill catchment and three of its subcatchments for three field size scenarios

	Trimmed mean (mm yr^{-1})	Mean (mm yr^{-1})	REA mean (mm yr^{-1})	REA RMSD (mm yr^{-1})
Dill				
0.5 ha	−2.5	−3.8	−1.9	35.3
1.5 ha	18.0	21.2	18.6	26.7
5.0 ha	28.0	33.9	30.7	41.3
Aar				
0.5 ha	−18.0	−18.7	−19.3	36.1
1.5 ha	−3.2	−0.1	−3.3	27.6
5.0 ha	12.8	18.3	14.9	26.5
Dietzhölze				
0.5 ha	34.6	39.9	41.6	56.0
1.5 ha	64.9	79.4	65.2	51.5
5.0 ha	71.7	88.2	71.8	54.0
Obere Dill				
0.5 ha	23.1	25.4	29.0	52.8
1.5 ha	58.6	72.6	69.1	44.8
5.0 ha	70.2	87.1	78.0	54.2

[31] and this might explain this tendency. In addition, both the trimmed mean and the REA method punish simulations that deviate strongly from the mean prediction, which might also have caused the similarity.

The PDF in Fig. 6 also illustrates the uncertainty in the predictions. For example, for the 5.0 ha scenario, the ensemble predicts that there is a 74% probability that the increase in discharge is between 15 and 35 mm yr^{-1} in the Dill catchment. It also predicts that the probability of a change in discharge exceeding >35 mm yr^{-1} is 15%. Besides the direct interpretation of the PDF, the uncertainty in the scenario predictions can also be quantified with the REA RMSD provided in Table 3. However, due to the asymmetric density functions for our scenario predictions (Fig. 6), this uncertainty measure is less suited to convey the uncertainties within the model scenario ensemble.

Fig. 6 also nicely illustrates that there are substantial differences between the scenario predictions for the four catchments. For example, the uncertainty in the scenario predictions for the Dietzhölze and the Obere Dill is much larger than for the Dill and the Aar catchment. Interestingly, Fig. 6 also shows that the 1.5 and 5.0 ha scenarios result in similar annual changes in discharge for the Dietzhölze and the Obere Dill catchment, as would have been expected from the land use distribution in Table 1. This observation was not apparent from the more qualitative Fig. 3.

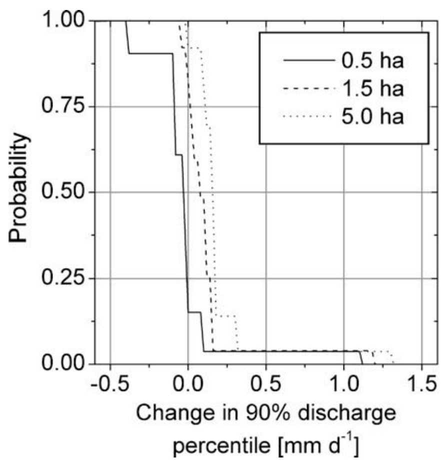

Fig. 7. Probability of change in 90% discharge percentile for the Dill catchment. Probability was estimated using the reliability ensemble averaging (REA) method.

The analysis so far has focused on quantifying the uncertainty in the change in mean annual discharge for the three field-size scenarios. However, it is also of interest to consider high flows. Fig. 7 presents the PDF for the change in the 90% discharge percentile for all three scenarios in the Dill catchment. To obtain these PDF, the REA method was slightly adapted. The model quality for current discharge (R_B in Eq. (1)) was defined as the difference between observed and simulated 90% percentiles. The parameter ε in Eq. (1) was computed from the difference between the minimum and maximum 90% discharge percentile of a 16-yr long moving window after linear detrending on the 35-yr long Dill discharge time series. As for the mean annual discharge, the model ensemble predicts an increase in high flows with an increase in field size. The REA mean changes were −0.019 mm d^{-1}, 0.112 mm d^{-1} and 0.190 mm d^{-1}, respectively for the 0.5, 1.5 and 5.0 ha scenarios. A comparison with the results in Table 3 shows that the high flows reacted more sensitive to land use change than the mean flows. Most likely, this is related to the increase of quickflow components with an increase in cropland.

Although the probabilistic interpretation of the land use change scenarios is considered to be an important step, it should be realized that the sample size used to derive these PDF was still rather limited. Clearly, care should be taken not to overextend the quantitative interpretation of Fig. 6. In climate change scenario analysis, the number of simulations has typically been increased by additionally considering model runs made with a single model, but with different equally probable parameterizations. In this study, it has not been attempted to increase the sample size with additional Monte Carlo simulations, but it is worthwhile to consider it for future studies of this kind. Finally, it is worth mentioning that the REA method can easily be extended with additional reliability measures. One could think of alternative quantitative error measures, such as the fit to interior basins not included in the calibration and the deterioration from calibration to validation, or perhaps one could even consider less conventional measures, such as the number of calibration parameters and the number of successful applications.

4. Conclusions

In the LUCHEM project, an ensemble of hydrological models was applied to the same set of land use change scenarios. Although the model ensemble produced a considerable range of predictions for the scenarios and catchments under consideration, there was general agreement amongst the models with respect to the direction of change. Compared to previous analyses of these land use change scenarios with a single model, this coherence within the model ensemble increases our confidence in the scenario predictions. Nevertheless, there remains the possibility that all models did not capture the essential hydrological processes determining the impact of land use change. For this reason, it would be worthwhile to repeat this type of model experiment in a well-instrumented catchment that has experienced (or still is experiencing) land use change.

The predicted amount of change varied between models. Seven out of ten models showed a small increase in mean annual discharge, which was explained by similar mean annual ET for the different land use classes. Three models provided deviating scenario predictions because of deviating mean annual ET for single (or multiple) land uses, despite an acceptable performance for the present-day conditions. This is a consequence of current calibration and validation practice, where the focus is on predicting discharge correctly. Unfortunately, internal catchment data for improved calibration and validation were not available in this study. Although model ensembles cannot entirely overcome problems inherent to hydrological model calibration, it can be used to either average out model structural error by optimally combining ensemble members or to obtain an estimate of uncertainty introduced into the scenario predictions by variations in model structure.

A trimmed mean was used to combine the ensemble member predictions into a single deterministic scenario prediction. Although it cannot be verified, it is assumed that this prediction is more accurate than the prediction of every single model because this is typically observed when deterministic ensemble methods are applied to actual conditions where validation data are available. Although deterministic ensemble methods are promising, it seems a missed opportunity to not use the information present in the ensemble to derive uncertainty estimates. To this end, we applied the probabilistic reliability ensemble averaging (REA) method. This method considers both the performance of the ensemble member for the current conditions and the deviation of the scenario prediction from the ensemble average to derive weights for each model within the ensemble. It was concluded that the REA method is a convenient way to summarize the information in the multi-model ensemble in a quantitative way. To the best of our knowledge, this is the first attempt to quantify the uncertainty in scenario predictions due to structural differences in hydrological models, which is often argued to be an important source of uncertainty. As such, we believe that the uncertainty estimates of the scenario predictions derived in this study are more realistic than previous estimates based on a single model. Clearly, realistic estimates of predictive uncertainty are paramount for risk assessment and communication with decision makers.

Acknowledgements

This study has been supported by the Deutsche Forschungsgemeinschaft within the scope of the Collaborative Research Centre (SFB) 299. A special thank goes to Bernd Weinmann for his efforts with the ProLand land use change scenarios.

References

[1] Ajami NK, Duan Q, Gao X, Sorooshian S. Multimodel combination techniques for analysis of hydrological simulations: application to Distributed Model Intercomparison Project results. J Hydrometeorol 2006;7:755–68.

[2] Andréassian V. Waters and forests: from historical controversy to scientific debate. J Hydrol 2004;291:1–27.

[3] Bathurst JC, Ewen J, Parkin G, O'Connell PE, Cooper JD. Validation of catchment models for predicting land-use and climate change impacts. 3. Blind validation for internal and outlet responses. J Hydrol 2004;287:74–94.

[4] Beven KJ. Rainfall–runoff modelling. The primer. Chichester: Wiley; 2001.

[5] Bhaduri B, Harbor J, Engel B, Grove M. Assessing watershed-scale, long-term hydrologic impacts of land-use change using a GIS-NPS model. Environ Manage 2000;266:643–58.

[6] Bormann H, Breuer L, Gräff T, Huisman JA. Analysing the effects of soil properties changes associated with land use changes on the simulated water balance: a comparison of three hydrological catchment models for scenario analysis. Ecol Mod 2007;209:29–40.

[7] Bormann H, Breuer L, Gräff T, Huisman JA, Croke BFW. Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) IV: model sensitivity on data aggregation and spatial (re-)distribution. Adv Water Resour; this issue.

[8] Bosch JM, Hewlett JD. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. J Hydrol 1982;55:3–23.

[9] Breuer L, Huisman JA, Frede H-G. Monte Carlo assessment of uncertainty in the simulated hydrological response to land use change. Environ Model Assess 2006;11(3):209–18.

[10] Breuer L, Huisman JA, Willems P, Bormann H, Bronstert A, Croke BFW, et al. Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) I: model intercomparison of current land use. Adv Water Resour; this issue.

[11] Brown AE, Zhang L, McMahon TA, Western AW, Vertessy RA. A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. J Hydrol 2005;310:28–61.

[12] Doblas-Reyes FJ, Hagedorn R, Palmer TN. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. Tellus A 2005;57:234–52.

[13] Ewen J, Parkin G. Validation of catchment models for predicting land use and climate change impacts. 1: Method. J Hydrol 1996;175:583–94.

[14] Fohrer N, Haverkamp S, Frede H-G. Assessment of long-term effects of land use patterns on hydrologic landscape functions – sustainable land use concepts for low mountain range areas. Hydrol Proc 2005;193:659–72.

[15] Franks S, Gineste P, Beven KJ, Merot P. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process. Water Resour Res 1998;34:787–97.

[16] Georgakakos KP, Seo DJ, Gupta H, Schaake J, Butts MB. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. J Hydrol 2004;298:222–41.

[17] Giorgi F, Mearns LO. Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging" (REA) method. J Clim 2002;1510:1141–58.

[18] Giorgi F, Mearns LO. Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. Geophys Res Lett 2003;3012:1629. doi:10.1029/2003GL017130.

[19] HLUG. Hessisches Landesamt für Umwelt und Geologie. Digital soil map 1:50000; 1998.

[20] Huisman JA, Breuer L, Frede HG. The sensitivity of simulated hydrological fluxes towards changes in soil properties in response to land use change. Phys Chem Earth 2004;29(11–12):749–58.

[21] IPCC. Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2001.

[22] Kuhlmann F, Möller D, Weinmann B. Modellierung der Landnutzung – Regionshöfe oder Raster-Landschaft? Ber Landwirtsch 2002;803: 351–92.

[23] Niehoff D, Fritsch U, Bronstert A. Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany. J Hydrol 2002;267: 80–93.

[24] Nöhles I. Landnutzungsklassifikation mit multitemporalen Landsat TM-Szenen in einer kleinstrukturierten Agrarregion. [Bd. 32], 1–160. 2000. Justus-Liebig-Universität Gießen. Schriftenreihe zur Bodenkunde, Landeskultur und Landschaftsökologie.

[25] Ott B, Uhlenbrook S. Quantifying the impact of land-use changes at the event and seasonal time scale using a process-oriented catchment model. Hydrol Earth Sys Sci 2004;81:62–78.

[26] Räisänen J, Palmer TN. A probability and decision-model analysis of a multimodel ensemble of climate change simulations. J Clim 2001;14:3212–26.

[27] Seibert J, Bishop KH, Nyberg L. A test of TOPMODEL's ability to predict spatially distributed groundwater levels. Hydrol Proc 1997;11:1131–44.

[28] Stednick JD. Monitoring the effects of timber harvest on annual water yield. J Hydrol 1996;176:79–95.

[29] Tebaldi C, Mearns LO, Nychka D, Smith RL. Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. Geophys Res Lett 2004;31:L24213. doi:10.1029/2004GL021276.

[30] Tebaldi C, Smith RW, Nychka D, Mearns LO. Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multi-model ensembles. J Clim 2005(10):1524–40. doi:10.1175/JCLI3363.1.

[31] Viney NR, Bormann H, Breuer L, Bronstert A, Croke BFW, Frede H-G, et al. Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) II: ensemble combinations and predictions. Adv Water Resour; this issue.

[32] Weinmann B, Schroers JO, Sheridan P. Simulating the effects of decoupled transfer payments using the land use model ProLand. Agrarwirtschaft 2006;555(6):248–56.

[33] Whitehead PG, Johnes PJ, Butterfield D. Steady state and dynamic modelling of nitrogen in the River Kennet: impacts of land use change since the 1930s. Sci Total Environ 2002;282–283:417–34.

[34] Zhang L, Dawes WR, Walker GR. Response of mean annual evapotranspiration to vegetation changes at catchment scale. Water Resour Res 2001;373:701–8.