HYDROLOGICAL PROCESSES
Hydrol. Process. **17**, 887–900 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.1168

Erosion models: quality of spatial predictions

Victor Jetten,[1]* Gerard Govers[2] and Rudi Hessel[1]
[1] *Utrecht Centre for Landscape Evolution and Dynamics, Utrecht University, 3508 TC Utrecht, The Netherlands*
[2] *Laboratory of Experimental Geomorphology, Redingenstraat 16B, 3000 Leuven, Belgium*

Abstract:

An overview is given on the predictive quality of spatially distributed runoff and erosion models. A summary is given of the results of model comparison workshops organized by the Global Change and Terrestrial Ecosystems Focus 3 programme, as well as other results obtained by individual researchers. The results concur with the generally held viewpoint in the literature that the predictive quality of distributed models is moderately good for total discharge at the outlet, and not very good for net soil loss. This is only true if extensive calibration is done: uncalibrated results are generally bad. The more simple lumped models seem to perform equally well as the more complex distributed models, although the latter produce more detailed spatially distributed results that can aid the researcher. All these results are outlet based: models are tested on lumped discharge and soil loss or on hydrographs and sedigraphs. Surprisingly few tests have been done on the comparison of simulated and modelled erosion patterns, although this may arguably be just as important in the sense of designing anti-erosion measures and determining source and sink areas. Two studies are shown in which the spatial performance of the erosion model LISEM (Limburg soil erosion model) is analysed. It seems that: (i) the model is very sensitive to the resolution (grid cell size); (ii) the spatial pattern prediction is not very good; (iii) the performance becomes better when the results are resampled to a lower resolution and (iv) the results are improved when certain processes in the model (in this case gully incision) are restricted to so called 'critical areas', selected from the digital elevation model with simple rules.

The difficulties associated with calibrating and validating spatially distributed soil erosion models are, to a large extent, due to the large spatial and temporal variability of soil erosion phenomena and the uncertainty associated with the input parameter values used in models to predict these processes. They will, therefore, not be solved by constructing even more complete, and therefore more complex, models. However, the situation may be improved by using more spatial information for model calibration and validation rather than output data only and by using 'optimal' models, describing only the dominant processes operating in a given landscape. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS erosion models; prediction quality; spatial patterns

INTRODUCTION

Just as models of soil erosion by water have tended over time to place a greater emphasis on representing the physical processes that are responsible for erosion, so have they tended toward a more explicit representation of the area on which erosion occurs, albeit to a varying degree. Some models represent space in a simplified way. The USLE† (Wischmeier and Smith, 1978), RUSLE (Renard *et al*., 1991) and EPIC (Williams, 1985) are lumped models assuming in principle a spatially homogeneous uniform hillslope, although it is possible to apply them to more complex terrain (e.g. Foster and Wischmeier, 1974; Desmet and Govers, 1996). GLEAMS and CREAMS (Knisel, 1991), the model it is based on, are both field-scale models that assume that a linked system of interrill and channel elements adequately represents the area of erosion. The more recent WEPP (Flanagan *et al*., 2001), KINEROS2 (Smith *et al*., 1995) and EUROSEM (Morgan *et al*., 1998) models adopt

* Correspondence to: Victor Jetten, Department of Physical Geography, Faculty of Geographical Sciences, PO Box 80115, 3508 TC Utrecht, The Netherlands. E-mail: v.jetten@geog.uu.nl
† Model acronyms, names and references are listed in the Appendix

Received 10 February 2002
Accepted 29 April 2002

a basically similar element-based scheme. In terms of process descriptions, the models evolved from rainfall-based erosion prediction, via Soil Conservation Service Curve-Number-based runoff estimations, to a more physically based water balance approach.

With the rise in computing power and geographical information system (GIS) capabilities, spatially distributed catchment models have been developed that simulate the runoff and erosion dynamics of larger and more complex catchments. The potential advantage of these models is that they allow the identification of source and sink areas of water, sediment and associated chemicals within a catchment. Soil conservation measures can then be designed to prevent the problem from occurring or to minimize the runoff, sediment and chemicals that leave the catchment. The models mentioned above are adapted to the catchment scale by increasing the number of elements and catering for special elements such as channels and ponds, whereas models such LISEM (De Roo *et al*., 1996; Jetten and De Roo, 2001), EROSION3D (Schmidt *et al*., 1999), ANSWERS (Beasley *et al*., 1980), TOPMODEL (Beven and Freer, 2001) and MIKE-SHE (Refsgaard and Storm, 1995) are based on a regular grid of equal-sized raster cells.

All these models are based on a water and sediment balance that produces runoff and (suspended) sediment for each spatial element, which is then routed towards the outlet using a kinematic wave routine. Although, in principle, this approach allows one to provide the user with a distributed image of the runoff and erosion, the models are mostly used to calculate the discharge and soil loss from a catchment at only one point: the outlet. The majority of results reported in the literature are outlet based, where either simulated hydrographs and sedigraphs are compared with measured data, or the models are used to predict future events. Likewise, the majority of the model tests and sensitivity analyses deal with the outlet-based data only. There are surprisingly few studies that compare simulated erosion patterns with observed erosion patterns. This is not only true for soil erosion models: Beven (2002) states that there are also very few validations of distributed predictions against distributed measurements in runoff, subsurface flow and groundwater modelling. At the same time, many researchers report the phenomenon of 'predicting the correct result for the wrong reasons', i.e. the prediction of acceptable soil loss and discharge with an incorrect (sometimes completely wrong) pattern of the source and sink areas (e.g. see Jetten *et al*., 1996; Takken *et al*., 1999, Favis-Mortlock *et al*., 2001). Although field information on erosion sources and sinks is often relatively easy to obtain by mapping erosion and sedimentation phenomena, it is rarely used in calibration exercises to improve the models. The usefulness of such data is shown by Desmet *et al*. (1999) and Vandaele *et al*. (1997), who show by an analysis of field data, digital terrain models and aerial photographs that the locations of ephemeral gullies can be predicted with simple, relief-based criteria. Improvement of distributed model results has been realized by incorporating agricultural features such as tillage direction, wheel tracks and field boundaries (e.g. see Souchère *et al*., 1998; Takken *et al*., 2001; Moussa *et al*., 2002).

In this article we attempt to give an overview of the predictive quality of models for outlet predictions and spatial pattern prediction. Outlet-based predictions are simplified here to total and peak discharge, and to net soil loss from plots or catchments. It is impossible to include all available model tests where the fitness of a particular model for a particular dataset is reported; many such tests are available, with varying results. Rather, we will focus here on the tests where several models were compared. Spatial pattern predictions are seen here as the spatial patterns of erosion and deposition inside a catchment. Two studies are shown where the LISEM model has been used in catchments in China and Belgium, where field observations are available, to draw preliminary conclusions on model capacity to produce runoff and erosion patterns.

OUTLET-BASED PREDICTIONS

The amount of water and sediment leaving the catchment (discharge and soil loss) are points of concern where damage downstream of the catchment may be expected (Boardman *et al*., 1994; Poesen *et al*., 1996). The design of erosion protection structures, such as retention basins, is based on the prediction of runoff from extreme events or design storms that may never have occurred in the area. Erosion models are technically capable of calculating the frequency and quantity of runoff and soil loss, but the question is whether the predictions are good enough.

In recent years, several exercises have been held whereby erosion models were compared and tested using common datasets for calibration and validation. The IGBP–GCTE (International Geosphere–Biosphere Programme–Global Change and Terrestrial Ecosystems) Soil Erosion Network (Ingram *et al.*, 1996) is testing the fitness of erosion models to predict the consequences of climate change for erosion. This is done in the following sequence: (i) field-scale water erosion models, (ii) catchment-scale water erosion models, (iii) wind erosion models and (iv) models with a landscape-scale and larger focus.

The first two exercises yielded results that are briefly reiterated here. For the model evaluations, common datasets were used which were split into a 'training set' for calibration and a 'testing set' for validation. The data used for the field-scale evaluation represented 73 site-years from seven sites in three countries: six field-scale erosion models took part in the evaluation. For the catchment-scale evaluation, data for ten events on a 40 ha catchment in the Netherlands were used to evaluate seven event-based catchment models. Details about the datasets, the models and the model specifics can be found in Boardman and Favis-Mortlock (1998) and De Roo (1999).

As an example, Figure 1 shows the results for total discharge for the catchment models (which was the output variable giving the best fit between predicted and observed values). From the graph, it can be seen that the overall performance of the models is moderate for the calibration dataset and moderate, at best, for the validation dataset. It should be said that some models performed better than others, and that some models were not intended to be used at this scale. The main conclusions drawn from the model comparisons were (Jetten *et al.*, 1999):

- Calibration is desirable for many models, and necessary for some. Calibration is most effective if the event(s) to be estimated are inside the range of calibration events.
- Calibration was almost always done on initial moisture content and saturated hydraulic conductivity K_{sat} (or the equivalents used in a particular model).
- Total discharge, is generally better predicted than peak discharge and both are better predicted than sediment discharge.
- For continuous-simulation models, long-term average results are better simulated than results for short time periods, and the prediction of when runoff actually occurs is not always correct (i.e. wrong days are predicted).
- Whereas for certain events the models may not perform well (absolute results), the correlation coefficients between observed and predicted values are more acceptable (relative results).
- From the discussions during the meetings it was clear that additional 'soft' information or 'meta data', in particular regarding the change in soil structure as a result of agricultural activities and/or climate, can greatly improve the quality of input data and model results.

Apart from these comparative studies, many individual model tests have been done by the model makers as well as by end-users. Their results support the conclusions drawn above. Zhang *et al.* (1996) tested the WEPP model and showed that, even for optimized values of K_{sat}, prediction was moderate, especially for smaller events. This over-prediction of small events also occurred in the other model comparisons, and it appears that, in general, erosion models have difficulties predicting small-scale events (e.g. see Nearing, 1998). Bathurst and Lukey (1998) tested the SHE model in the Draix area (France). Their comparison of measured and predicted soil loss showed that for some events the prediction was accurate, whereas for other events there was one to two orders of magnitude difference between observed and predicted values. These results contrast strongly with those obtained by Brochot and Meunier (1995). Using the same data as Bathurst *et al.* (1998) for the Draix area, they obtained good results using a simple regression model based on precipitation amount and intensity only: simple power equations resulted in an R^2 of 0·72–0·78 between observed and predicted sediment export for the validation dataset. De Roo (1993) compared the lumped USLE and MMF (Morgan, 2001) models

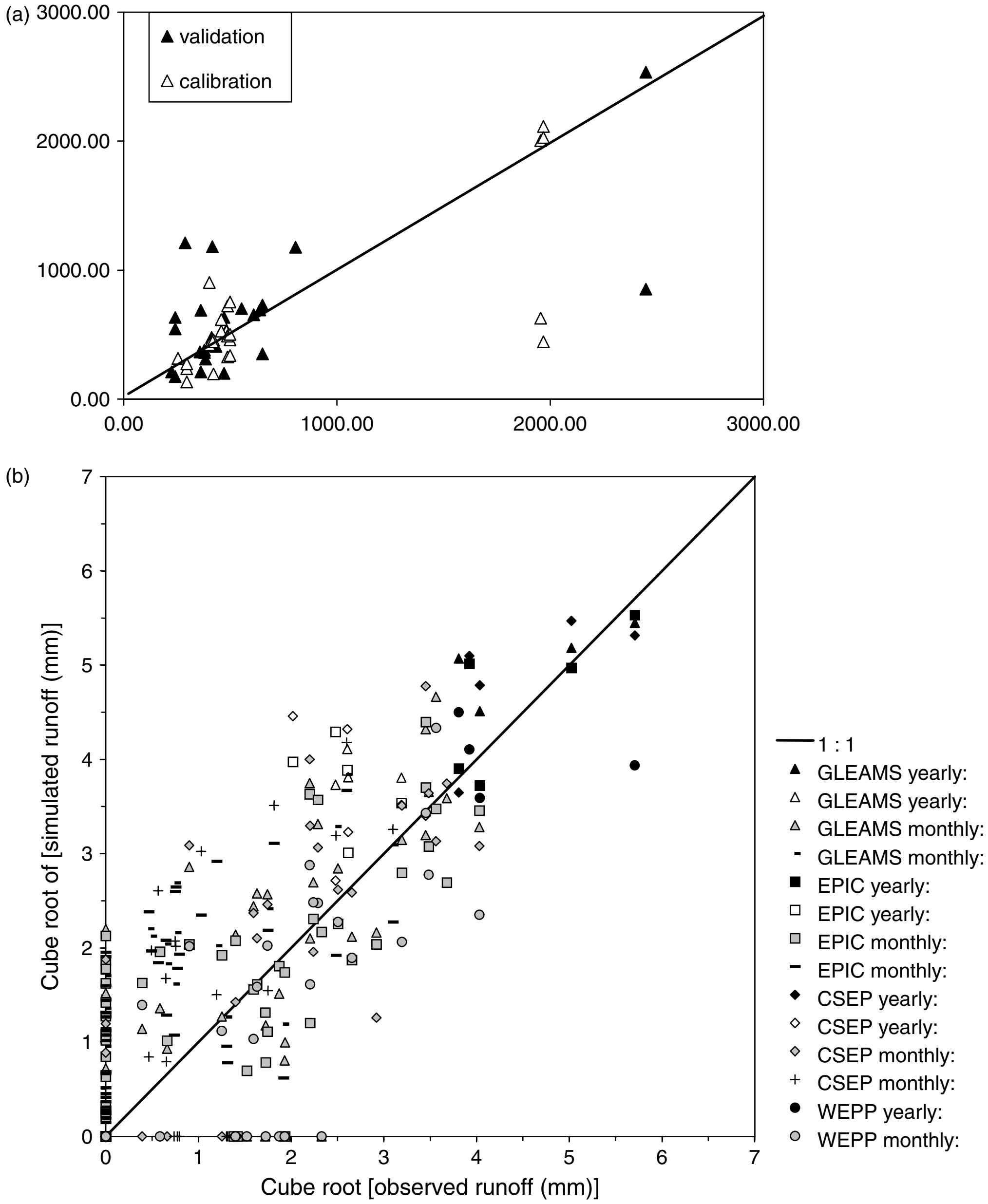

Figure 1. Observed and predicted total discharge for five calibration and five validation events using seven different catchment models (top), and four hillslope models (bottom) (Jetten *et al.*, 1999)

with the spatially distributed ANSWERS and KINEROS models and concluded that they performed equally well when the models were tested for the same types of result (such as annual soil loss). Risse *et al.* (1993) tested the USLE for over 1700 years of erosion on 208 plots. They obtained an R^2 of 0·58 when comparing observed annual soil losses with predicted annual soil losses and an R^2 of 0·75 when comparing average annual values. These values are even somewhat higher then those obtained by Zhang *et al.* (1996) for the much more complex and physically based WEPP model (0·54 and 0·68 respectively).

In general, these results confirm those obtained in the GCTE model comparison exercises: accurate erosion prediction is still difficult. A fundamental reason for the poor to moderate predictive capabilities of erosion models is the spatial and temporal variability of the required input parameters and the uncertainty associated with them. Many studies have been done on the effect of spatial variability and error propagation, using sensitivity analysis based on Monte Carlo simulation, as well as other methods (e.g. see De Roo *et al*., 1992; Heuvelink, 1998; Brazier *et al*., 2000; Van Loon, 2001). They basically conclude that most of the model parameters are stochastic in nature, and measurement errors and uncertainty as a result of spatial interpolation add considerable uncertainty to the model results: using the EUROSEM, Quinton (1997) showed that the hydrograph and sediment graphs measured on erosion plots in Woburn (UK) were nearly always a member of the set of possible model outcomes if input parameter values were allowed to vary within reasonable bounds. However, the range in possible model outcomes was usually so large that an accurate prediction did not appear possible. Nearing *et al*., (1999) and Nearing (2000) tested general model fitness by comparing the soil loss of pairs of adjacent plots in the USA, selecting plots that could be considered identical in terms of received rainfall, soils and surface conditions. The coefficient of variation (CV) increased strongly with a decrease of event size, from *ca* 150% for very small events (0·01 tonne ha^{-1}) to *ca* 10% for very large events (about 400 tonne ha^{-1}). Keeping in mind that the plot size is very small (Wischmeier-type plots), extrapolation to a catchment scale could mean that it is fundamentally impossible for a model to predict small events. Nearing (2000) proposes that the 'physical model', represented by the replicated plot, can be seen as a standard for the erosion simulation model. They conclude that, given the state of our knowledge concerning erosion processes and environmental interactions and the inability to estimate model parameters properly, one would not expect the simulation model (of any type) to perform up to the standard of the physical model. Although the low CV for large events is encouraging, small events have received much interest in recent studies of eroded agricultural lands as a source of polluted sediment: only small quantities are sufficient to pollute surface water if the clay particles are loaded with agro-chemicals.

Several studies show that prediction quality is especially poor for small erosion events. The GCTE results, however, show that many models have problems with the prediction of extreme events as well. This could be due to a number of reasons. First, the system may not behave in the same way for medium and for large events. For example, cases have been observed where a wheat field was flattened during a rainstorm, thus providing a pass-through for the runoff that would otherwise have been delayed or dispersed. Also, barriers in the catchment that are normally not crossed by runoff may be overcome by the large amounts of overland flow produced during an extreme event, so that the actual runoff pattern is not the one used in the model. Second, many factors considered as constants in erosion models are, in fact, dynamic and may vary significantly within an erosion event: the rainfall energy destroys the surface structure during the event, effectively decreasing the infiltration capacity, the surface storage and the friction to runoff. Many studies have been conducted on the decrease of these soil parameters under the influence of rainfall (e.g. see Boifin and Monnier, 1986; Torri *et al*., 1999), but most of the relationships have only been tested under laboratory circumstances and are not incorporated in current models. The changes in topsoil structure will be much more rapid and more pronounced when the rainfall intensities are higher. Therefore, the discrepancy between model behaviour and real-world behaviour may increase for larger events.

The variability, and hence the uncertainty associated with input parameter values, is probably the most important reason why more complex physically based erosion models, in general, do not perform better than lumped regression-based models. More complex models with better process descriptions should, in principle, be capable of better output predictions; however, they also require more input data, with which there is an (often unknown) amount of uncertainty and error associated that will propagate through the model calculations and ultimately deteriorate the quality of the final results. The comparison of the results of Zhang *et al*. (1996) with those of Risse *et al*. (1993), and of those of Bathurst *et al*. (1998) with those of Brochot and Meunier (1995), suggests that the additional error resulting from introducing additional parameters often outweighs the potential improvement in prediction due to a better process description.

SPATIAL PATTERN PREDICTIONS

When designing erosion control measures, two aspects of the predictive quality of a model are important: (i) the quantity of water and sediment has to be correctly predicted, and (ii) the spatial patterns of erosion and sedimentation have to be correctly predicted. One can argue that the accurate prediction of the spatial pattern is even more important than the accurate prediction of the amount of runoff and sediment produced: it is much more cost effective to over-dimension an erosion control measure than to put it in the wrong spot. An example of an application of an erosion model to conservation design was given by Jetten and De Roo (2001), who showed that the efficiency of the grass strips was greatly improved by locating them in accordance with the tillage direction instead of the drainage pattern based on the steepest slope.

The focus in erosion research has almost entirely been on measuring hydrographs and sedigraphs at the outlet. Mapping erosion and sedimentation features (rills, gullies, colluvial deposits) for the purpose of modelling has only rarely been done (two examples are given below). Until recently, many models were not even capable of producing the predicted spatial patterns as an output. Here, the speed advantage of polygon-based models (such as WEPP, KINEROS2, EUROSEM), which normally use between 20 and 100 spatial elements, over raster-based models (LISEM, EROSION3D, TOPMODEL, MIKE-SHE), which use up to 10 000 pixels, turns into a disadvantage, because the erosion pattern inside a plane element is not shown. Since the plane and channel elements are subdivided into smaller elements to allow a finite difference kinematic wave approach, then giving some sort of erosion pattern would, in principle, be possible (the latest version of WEPP uses this technique). A further difficulty is that there are very few pattern comparison techniques, other than a visual comparison of maps. The fraction of pixels correctly classified, when the same classification intervals are used for predicted and observed maps, is a common value known as the 'kappa index' (Jetten *et al*., 1996). Another possibility is to use map-units (e.g. fields) and compare the predicted and measured erosion for each field (e.g. see Takken *et al*., 1999).

Intuitively, the finer the resolution, the more a pattern comparison should be feasible, even if it is only visually. However, one should be aware that model response is dependent on the spatial discretization. Vázquez *et al*. (2002) reported a dependency of the results of MIKE-SHE on the grid cell size, and WEPP appears to be sensitive to the dimensions of channel elements and the internal division of the plane elements (Baffaut *et al*., 1997).

Because of the lack of available studies, the remainder of this article will not attempt to be exhaustive on the pattern prediction quality of models. Rather, two questions are addressed: (i) What is the influence of spatial discretization of the landscape on the outcome of the LISEM model? (ii) How well do the predicted patterns compare with observed erosion in two examples taken from research in China and in Belgium?

THE EFFECT OF DISCRETIZATION OF SPACE

Simulated erosion patterns are directly related to the discretization of the catchment and the drainage network that is a result of this discretization: raster-based models often use a deterministic flow network where a cell can only flow towards one of the eight surrounding cells. Since the model (in this case LISEM) routes water and suspended sediment towards the outlet using a user-supplied network, the simulated erosion patterns are closely related to this network. In view of this, one question that arises concerns the influence of resolution (seen here as grid cell size) on the simulation results. Grid cell size is one of the most arbitrary choices in spatial modelling. It is often a balance between a subjective idea of accuracy acceptable to the user and practical reasons such as calculation time. In raster-based models the grid cell size determines everything: the discretization of space for the kinematic wave, the local slope derived from the digital elevation model (DEM), the land-use pattern, especially when the fields are small and have irregular shapes, and the actual catchment size. An inspection of finite difference solutions of the kinematic wave (e.g. see Chow *et al*., 1988) shows that they are as sensitive to the discretization of space as they are to the discretization of time. The same holds for polygon-based models; although the elements are delineated by hand and thought to represent the catchment, they are divided into smaller pieces to allow a finite difference solution of the kinematic wave equations, and a rill density has to be assumed by the user to simulate concentrated flow paths within the plane elements.

Runoff, erosion and deposition in relation to grid size are investigated for two catchments: the Da Nan Go catchment (210 ha) in northern China (Hessel *et al*., in press), which has very steep slopes and very small parcels, and the Catsop catchment (42 ha) in the south of the Netherlands, which has small slope angles and a few large fields (De Roo and Jetten, 1999). Table I shows the effect of changing grid cell size on the dataset: the results are represented relative to the 10 m resolution, because this grid size was used to calibrate both catchments on the outlet data. Table I shows that the average and maximum slope decrease and that the catchment size increases with increasing grid cell size. This has a large effect on the discharge: both peak discharge and runoff volume decrease with increasing grid cell size. The change in slope with grid cell size is probably the main factor explaining the change in response: rainfall falls on a horizontal projection of the surface and is recalculated in LISEM to the actual surface area by division with cosine of the slope angle. As a result, the water is spread out more, and lower water depths are predicted when the slope is steeper. Furthermore, the slope angle influences flow velocity directly (calculated with the Manning equation), as well as the unit stream power (product of velocity and slope), and, therefore, erosion and deposition. Although the Catsop catchment has a much lower slope angle, the effects are the same.

Erosion, deposition and net soil loss, however, behave very differently from the discharge. The variation of net soil loss with grid cell size is even fairly unpredictable. In fact, the effect of changes in grid cell size is of the same size as the sensitivity of LISEM to infiltration-related variables such as K_{sat} and initial moisture content (De Roo *et al*., 1996). Whereas runoff occurs practically everywhere in the catchment, important erosion is restricted to isolated areas on the steeper slopes; deposition is even more restricted to isolated pixels at the bottom of slopes and in the thalweg. The relatively large influence of isolated pixels makes the final result difficult to predict. For Da Nan Go the erosion and the deposition increase between 5 and 10 m resolution and then decrease again between 10 and 50 m resolution. For Catsop, erosion and deposition also decrease with increasing grid cell size, but the decrease is much stronger. The results also imply that a calibration done for a given grid cell size is not valid for any other size.

COMPARISON OF OBSERVED AND SIMULATED EROSION PATTERNS

In the Da Nan Go catchment, rills were mapped in two consecutive years on the agricultural fields. In 1999 only one major rainstorm occurred, and the resulting rill volumes were estimated and mapped. Total erosion was estimated by adding sheet erosion in a procedure explained by Hessel *et al*. (in press). The resulting observed erosion map was classified using four classes (0–8, 8–30, 30–80, >80 tonne ha^{-1}). LISEM was

Table I. Catchment properties and relative discharge and erosion for two catchments in China and The Netherlands. The catchments are calibrated for the 10 m resolution, which is taken as unity

	Da Nan Go (China)				Catsop (Netherlands)		
	5 m	10 m	20 m	50 m	10 m	20 m	50 m
Catchment size (ha)	210	210	214	228	42	43	49
Slope, avg % (max %)	55 (203)	54 (169)	48 (122)	36 (82)	6·1 (28·9)	5·8 (20·7)	4·9 (12·4)
Total discharge	1·12	1·00	0·92	0·65	1·00	0·92	0·69
Peak discharge	1·15	1·00	0·78	0·48	1·00	0·72	0·36
Flow detachment	0·98	1·00	1·09	0·97	1·00	0·91	0·73
Deposition	0·91	1·00	1·08	0·92	1·00	0·92	0·72
Soil loss	1·47	1·00	1·11	1·38	1·00	0·88	0·88

calibrated to give an acceptable hydrograph and sedigraph at the outlet. Figure 2 shows the observed and simulated patterns for the agricultural fields. LISEM predicts erosion for other types of area as well (orchards, rangeland, wasteland), but these were not mapped because no field information was available. The outlines of the agricultural fields are deliberately not shown in Figure 2 as they would be the same for both maps, which would visually suggest that it is a good simulation result of LISEM. The maps show that, in general, the simulated patterns resemble the mapped patterns, in fact reflecting the spatial distribution of the rainfall (the eastern part of the catchment received more rainfall). In detail, however, it is clear that there are many discrepancies: there are areas where severe erosion was mapped and none is simulated and *vice versa*. There may be a number of reasons why this is the case. On a grid-cell-to-grid-cell basis there may exist many differences between reality and the simulation results:

- The spatial variability of the cohesion, which determines the local flow detachment, is not correctly mapped. In fact, an average value was taken for each land-use type because there was no discernible spatial correlation between the cohesion measurements. Thus, the actual erosion for each grid cell can easily be wrong.
- The flow velocity and transport capacity in a grid cell may not be correct, e.g. because the local infiltration capacity is wrongly estimated, causing the runoff production to be too low or too high, or the local Manning's n may be wrong, causing the velocity to be too low or too high.
- Both detachment and deposition occur during the simulation and the net effect is an absence of soil loss from the pixel.

Moreover, the cause for discrepancies may also be found upstream of the pixel:

- The flow network itself is not correct, so that the actual amount of water flowing in a cell is not the real one. Figure 3 (left) shows erosion in a pattern that suggests a different network direction than the one used as an input map; severe erosion sometimes occurs at the top of the simulated network branches.
- The upstream hydrological conditions are not correctly simulated, causing too little or too much runoff inflow into a grid cell.

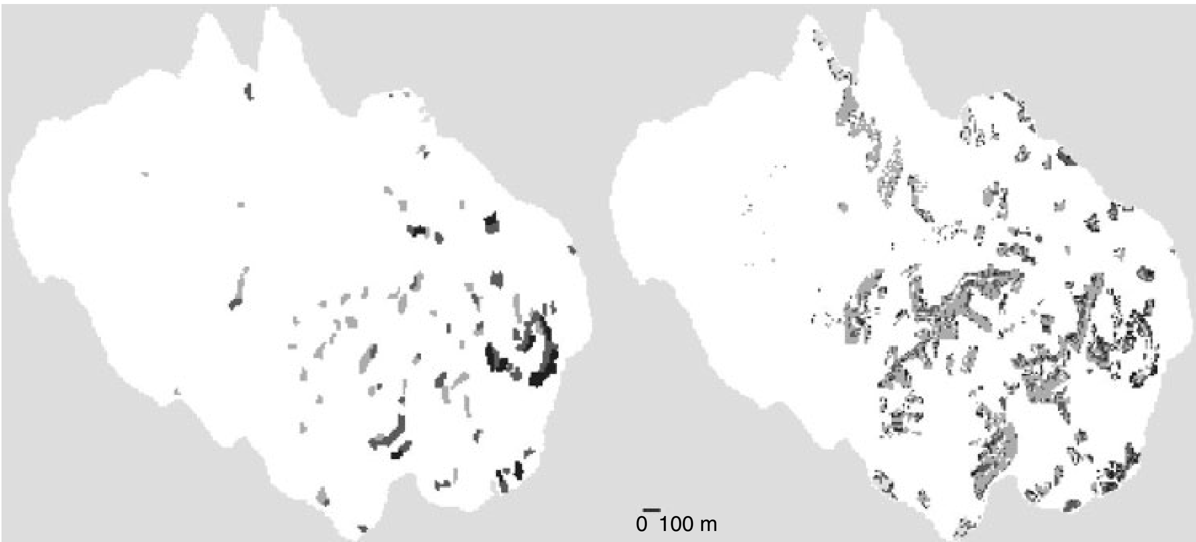

Figure 2. Observed (left) and predicted (right) erosion patterns of the 20 July 1999 event in the Da Nan Go catchment. Both maps are based on the same classification intervals

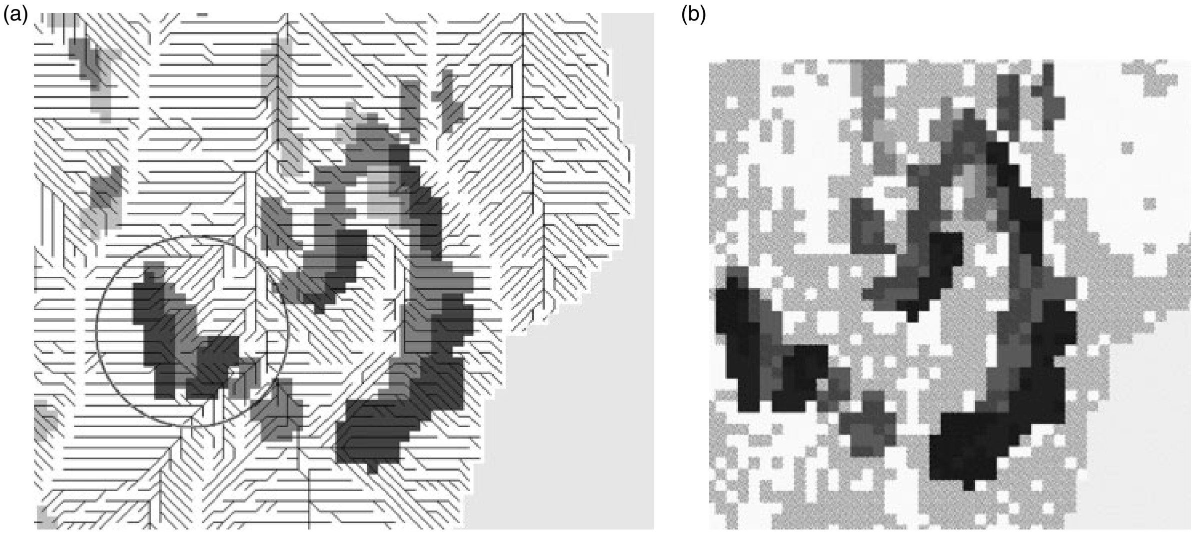

Figure 3. Left: suspicious drainage network compared with the eroded area (circle); heavy erosion occurs at the top of the branches. Right: erosion should only occur in the dotted areas, because outside these areas there is no transport capacity surplus throughout the simulation

- The simulated transport capacity in a pixel is already saturated, so that no net flow detachment is predicted. In the right-hand map of Figure 3, erosion should only occur in the dotted areas because there is no transport capacity surplus outside these areas throughout the simulation. Nevertheless, severe erosion is observed.

From the above it is clear that on a pixel-to-pixel level it is impossible to simulate the correct pattern. The input data will simply not be good enough to achieve this. A possible solution to this is spatial aggregation (Van Rompaey *et al.*, 1999). In order to investigate at what level of spatial aggregation the predicted map starts to resemble the observed map, an error index was calculated. This is simply the relative sum of the absolute difference between the classified maps (the lower the value, the more the maps show a resemblance):

$$\text{Error Index} = 1/n \sum_{i}^{n} |c_i^{\mathrm{o}} - c_i^{\mathrm{p}}|$$

where c^{o} and c^{p} are the class numbers of the observed and predicted grid cells and n is the number of grid cells (in this case of the agricultural areas only). Table II shows the results for different resolutions, obtained by resampling the original maps (which were at a 10 m resolution). It is clear that the maps only start to show a resemblance when the pixels are resampled to 50 m or larger, which is at five to ten times the original resolution. Moreover, it should be stressed that there are only four classes (the observed data did not permit a more detailed classification), which increases the likelihood of obtaining a correct class for a grid cell. Furthermore, spatial aggregation leads to error reduction by smoothing out local variations that are often due to errors in the input data: it will not allow one to reduce the error due to structural deficiencies in the model.

Table II. Relative cumulative error of classified pixels comparing observed and predicted erosion maps of the Da Nan Go catchment, at different levels of resampling. A smaller value indicates a higher resemblance

Resolution (m)	10	20	50	100	200
Error index	0·692	0·377	0·186	0·054	0·039

A POSSIBLE MEANS OF IMPROVEMENT

The current discussion in distributed runoff and erosion modelling moves towards a greater interaction between the landscape and the model, at least for runoff and erosion models. Beven (2002) suggests that for each particular catchment represented by a set of distributed variables one may be able to find one or more optimal models. Thus, all superfluous process descriptions that do not improve the result but only add uncertainty should be pruned. Evidently, the optimal model will then be dependent on the landscape characteristics and the dominant process operating. Van Loon (2001) has tested this approach for two catchments in West Africa and Costa Rica on different scales and was able to identify a set of models that have the greatest 'fitness' to simulate the runoff in these catchments. An interesting aspect about this exercise is that they integrate the landscape and the model into a single matrix that is used to determine the best set of models automatically. Moreover, spatial observations of runoff are included in the matrix. Van Rompaey and Govers (2002) propose a technique that allows the determination of an optimum model structure provided that the error distribution on the input data is known. As an example they applied the RUSLE to a small catchment in the Loess Belt of central Belgium. They concluded that, in this case, it was better to assume a constant soil erodibility value K instead of using spatially distributed soil information derived from the Belgian soil map. Indeed, the error due to uncertainty on the soil information more than compensated for the improvement in the model's predictive ability due to a better process description.

On a much more practical level, several studies have shown that, when the model is allowed to include observed patterns *a priori*, the results may be greatly improved. For instance, Desmet and Govers (1997) and Takken *et al*. (2001) showed a marked improvement when the imposed drainage network followed the tillage direction instead of the steepest slope. These *a priori* imposed patterns need not be precise and deterministic, but rather they can zones or critical areas in which a process is allowed to take place. An example is given from the Kinderveld catchment in Belgium. Based on simple landscape characteristics derived by Desmet *et al*. (1999) and Nachtergaele *et al*. (2001), thresholds for ephemeral gully incision can be derived from a DEM. These thresholds are based on power functions of contributing area and slope (much like the wetness index; e.g. see Quinn *et al*., 1991). When incorporating them into LISEM to limit the modelling of gully incision to those 'critical areas' only (the dotted areas in Figure 4), and assuming rill erosion for the remaining parts of the catchment, fairly good results were obtained for the predicted gully dimensions (Figure 5). Without this spatial restriction, LISEM would predict gully incision at many more locations.

DISCUSSION AND CONCLUSIONS

The conclusions from this paper are fairly negative: the overall status of erosion models is that they perform moderately at best with respect to the outlet data. A fundamental reason for this observation is the high spatial and temporal variation of erosion and sediment transport and our inability to assess and/or describe this variability in terms of the input parameters normally used in erosion models. This poses fundamental limits to the degree of accuracy that erosion models may achieve; this problem is probably more important for minor to moderate events than for large erosion events. However, the simulation of large events poses specific problems as well. More complex, physically based models do not necessarily perform better than lumped, regression-based models, mainly because input errors increase with increasing model complexity.

This touches on the discussion of what constitutes a good model. Quinton (1994) suggests an iterative, stepwise approach in calibrating a model, whereby the 'fitness' of the model for a specific purpose may increase when more data and effort are added. In that sense, the GCTE exercises mentioned in the first part of this article may be seen as a fair test, but perhaps not as 'good modelling', and the results might have been better if the participants had more knowledge and data of the areas involved. Strategies to determine an optimal model structure, depending on the characteristics of the landscape and the quality of the input data, are beginning to emerge and may help to avoid additional prediction error due to model over-parameterization.

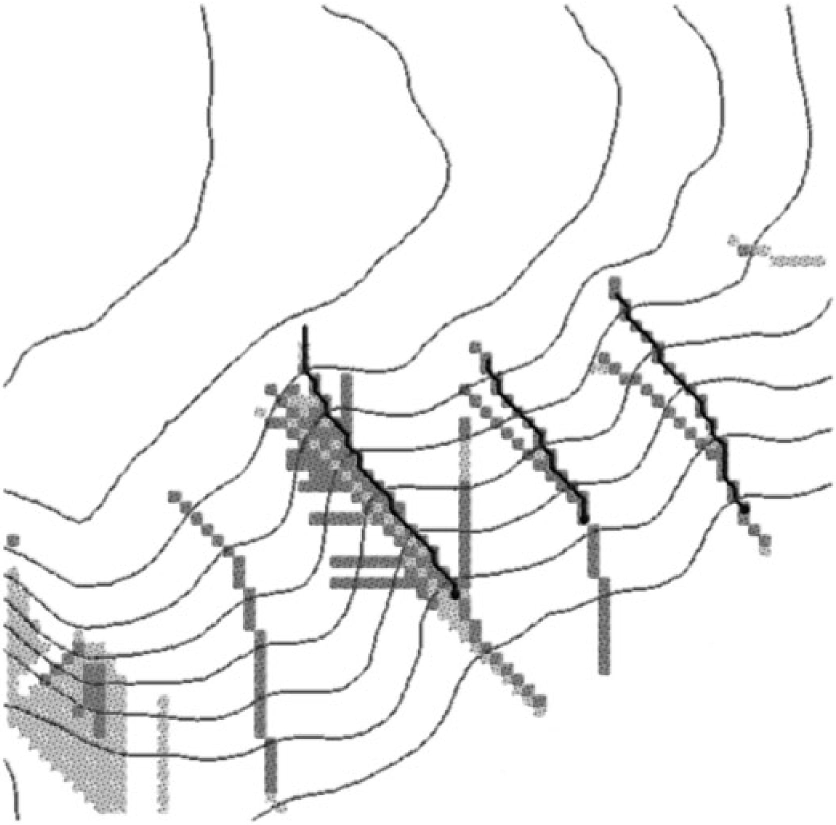

Figure 4. Prediction of gully incision by LISEM (dark-grey areas) on a slope of the Kinderveld catchment, when the incision is restricted to 'critical zones' (dotted areas). The three observed gullies are shown as black lines

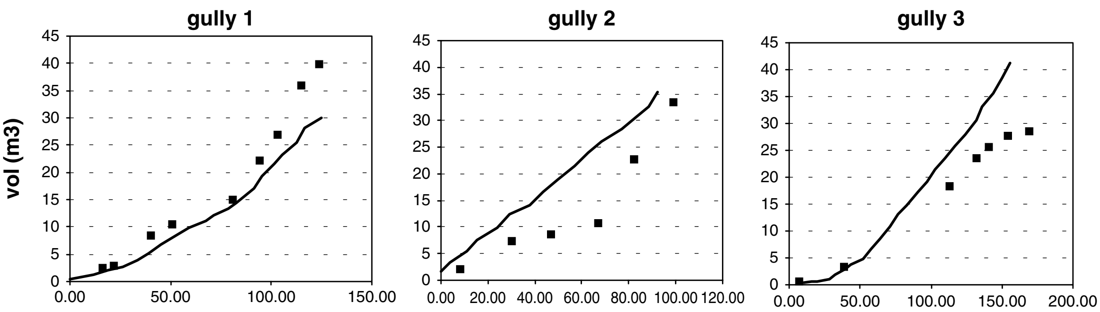

Figure 5. Predicted (lines) and observed gully volumes (dots) along the mapped gullies (shown as lines in Figure 4)

The evident disadvantage of adapting the model structure to local circumstances is that no longer can a 'universal' model be proposed.

Models should always be carefully calibrated for a given area before being used for predictions. This calibration should not only focus on outlet data. The model's capability to represent the processes occurring within the catchment can be much better assessed when the spatial pattern of erosion and deposition as it is observed within the catchment are also used.

The application of the LISEM tested here shows that accurate predictions at the grid-cell resolution at which the model is run are impossible. The level of detail of the input data and knowledge needed is prohibitive. In fact, it seems we need to know everything before we can simulate an event. This is clearly not possible, and a redefinition of the modelling goals is necessary. Since the 'spatial error' (expressed here as the relative difference between classified predicted and observed pixels) decreases rapidly when the maps are resampled to lower resolutions, it may be possible to predict the location of sources and sinks of runoff and sediment adequately on a given scale and resolution, which is an important goal in erosion modelling. The proper simulation at a catchment scale of the effect of small-scale detailed anti-erosion measures, such as grass strips or terraces, especially when the optimal location has to be determined by the model, seems to be impossible for the time being.

Model results are likely improved by using *a priori* terrain knowledge of an area, used to delineate 'critical areas' where certain processes, such as the incision and formation of gullies in the example, are allowed. These critical areas may be derived from field observations of runoff and erosion patterns, an analysis of the DEM, aerial photographs or remote sensing. Although, on the one hand, this seems very deterministic, as we are telling the model what to do and where to do it beforehand, on the other hand it is good practice to use all knowledge available.

APPENDIX

Acronym	Name	Reference
ANSWERS	Areal non-point source watershed environmental response simulation model	Beasley *et al.* (1980)
CREAMS	Chemicals, runoff and erosion from agricultural management systems	Knisel (1991)
EPIC	Erosion–productivity impact calculator	Williams (1985)
EROSION3D	3D erosion model	Schmidt *et al.* (1999)
EUROSEM	European soil erosion model	Morgan *et al.* (1998)
GLEAMS	Groundwater loading effects of agricultural management systems	Knisel (1991)
KINEROS2	Kinematic runoff and erosion model	Smith *et al.* (1995)
LISEM	Limburg soil erosion model	Jetten and De Roo (2001)
MIKE-SHE	—	Refsgaard and Storm (1995)
RUSLE	Revised universal soil loss equation	Renard *et al.* (1991)
TOPMODEL	—	Beven and Freer (2001)
USLE	Universal soil loss equation	Wischmeier and Smith (1978)
WEPP	Water erosion prediction project	Flanagan *et al.* (2001)

REFERENCES

Baffaut C, Nearing MA, Ascough JC, Liu BY. 1997. The WEPP watershed model: II. Sensitivity analysis and discretization on small watersheds. *Transactions of ASAE* **40**(4): 935–943.

Bathurst JC, Lukey B. 1998. Modelling badlands erosion with SHETRAN at Draix, southeast France. *Modelling Soil Erosion, Sediment Transport and Closely Related Hydrological Processes*, Summer W, Klaghafer E, Zhang W (eds). *IAHS Publication* no. 249. IAHS: Wallingford; 129–136.

Beasley DB, Huggins LF, Monke EJ. 1980. ANSWERS: a model for watershed planning. *Transactions of ASAE* **23**(4): 938–944.

Beven K. 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrological Processes* **16**: 189–206.

Beven K, Freer J. 2001. A dynamic TOPMODEL. *Hydrological Processes* **15**: 1993–2011.

Boardman J, Favis-Mortlock D. 1998. *Modelling Soil Erosion by Water*. *NATO ASI Series 1: Global Environmental Change*, vol. 55. Springer-Verlag: Berlin.

Boardman J, Ligneau L, De Roo APJ, Vandaele K. 1994. Flooding of property by runoff from agricultural land in northwestern Europe. *Geomorphology* **10**: 183-196.

Boifin J, Monnier G. 1986. Infiltration rate as affected by soil surface crusting caused by rainfall. In *Assessment of Soil Surface Sealing and Crusting*, Callebaut F, Gabriels D, De Boodt M (eds), Proceedings of the Ghent Symposium 1985; 210–217.

Brazier RE, Beven KJ, Freer J, Rowan JS. 2000. Equifinality and uncertainty in physically-based soil erosion models: application of the GLUE methodology to WEPP—the Water Erosion Prediction Project for sites in the UK and US. *Earth Surface Processes and Landforms* **25**: 825–845.

Brochot S, Meunier M. 1995. Erosion de badlands dans les Alpes du sud: synthèse. In Meunier M (ed.). *Compte-Rendu de recherches n°3: BVRE de Draix*, *Equipements pour l'eau et l'environnement* no. 21. Cemagref-Editions: 141–174.

Chow VT, Maidment DR, Mays LW. 1988. *Applied Hydrology*. McGraw-Hill.

De Roo APJ. 1993. *Modelling Surface Runoff and Soil Erosion in Catchments Using Geographical Information Systems*. *Netherlands Geographical Studies* 157. Faculty of Geographical Sciences, Utrecht University.

De Roo APJ. 1999. Modelling of soil erosion by water on a catchment scale. *Catena* **37**: 275–541.

De Roo A, Jetten V. 1999. Calibrating and validating the LISEM model for two datasets from the Netherlands and South Africa. In *Modelling of Soil Erosion by Water on a Catchment Scale*, De Roo APJ (ed.). *Catena* **37**: 521–541.

De Roo APJ, Hazelhoff L, Heuvelink GBM. 1992. Estimating the effects of spatial variability of infiltration on the output of a distributed runoff and soil erosion model using Monte Carlo methods. *Hydrological Processes* **6**: 127–143.

De Roo APJ, Wesseling CG, Ritsema CJ. 1996. LISEM: a single event physically-based hydrologic and soil erosion model for drainage basins. I: theory, input and output. *Hydrological Processes* **15**: 1107–1117.

Desmet PJJ, Govers G. 1996. A GIS procedure for automatically calculating the USLE LS factor on topographically complex landscape units. *Journal of Soil and Water Conservation* **51**: 427–433.

Desmet PJJ, Govers G. 1997. Two-dimensional modeling of the within-field variation in rill and gully geometry and location related to topography. *Catena* **29**: 283–306.

Desmet PJJ, Poesen J, Govers G, Vandaele K. 1999. Importance of slope gradient and contributing area for optimal prediction of the initiation and trajectory of ephemeral gullies. In *Modelling of Soil Erosion by Water on a Catchment Scale*, De Roo APJ (ed.). *Catena* **37**: 377–392.

Favis-Mortlock DT, Boardman J, MacMillan VJ. 2001. The limits of erosion modeling: why we should proceed with care. In *Landscape Erosion and Evolution Modelling*, Harmon RS, Doe WW (eds). Kluwer Academic: New York; 477–516.

Flanagan DC, Ascough JC, Nearing MA, Laflen JM. 2001. The Water Erosion Prediction Project (WEPP) model. In *Landscape Erosion and Evolution Modelling*, Harmon RS, Doe WW (eds). Kluwer Academic: New York; 145–199.

Foster GR, Wischmeier WH. 1974. Evaluating irregular slopes for soil loss prediction. *Transactions of ASAE* **17**: 305–309.

Hessel R, Jetten V, Liu B, Zhang Y, Stolte J. In press. Calibration of the LISEM model for a small loess plateau catchment. *Catena*.

Heuvelink GBM. 1998. *Error Propagation in Environmental Modelling with GIS*. Taylor & Francis: London.

Ingram JSI, Lee JJ, Valentin C. 1996. The GCTE Soil Erosion Network: a multi-participatory research program. *Journal of Soil and Water Conservation* **51**: 377–380.

Jetten V, De Roo APJ. 2001. Spatial analysis of erosion conservation measures with LISEM. In *Landscape Erosion and Evolution Modelling*, Harmon RS, Doe WW (eds). Kluwer Academic: New York; 429–445.

Jetten VG, Boiffin J, De Roo APJ. 1996. Defining monitoring strategies for runoff and erosion studies in agricultural catchments: a simulation approach. *European Journal of Soil Science* **47**: 579–592.

Jetten V, De Roo APJ, Favis-Mortlock DT. 1999. Evaluation of field-scale and catchment-scale soil erosion models. In *Modelling of Soil Erosion by Water on a Catchment Scale*, De Roo APJ (ed.). *Catena* **37**: 521–541.

Knisel WG. 1991. CREAMS/GLEAMS: a development overview. In *Proceedings of the CREAMS/GLEAMS Symposium*, Beasley DB, Knisel WG, Rice AP (eds). Publication 4, Agricultural Engineering Department, University of Georgia, Athens, GA, USA; 9–17.

Morgan RPC. 2001. A simple approach to soil loss prediction: a revised Morgan–Morgan–Finney model. *Catena* **44**: 305–322.

Morgan RPC, Quinton JN, Smith RE, Govers G, Poesen JWA, Auerswald K, Chisci G, Torri D, Styczen ME. 1998. The European soil erosion model (EUROSEM): a process-based approach for predicting soil loss from fields and small catchments. *Earth Surface Processes and Landforms* **23**: 527–544.

Moussa R, Voltz M, Andrieux P. 2002. Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events. *Hydrological Processes* **16**: 393–412.

Nachtergaele J, Poesen J, Vandekerckhove L, Oostwoud Wijdenes DJ, Roxo M. 2001. Testing the ephemeral gully erosion model (EGEM) for two Mediterranean environments. *Earth Surface Processes and Landforms* **26**: 17–30.

Nearing MA. 1998. Why soil erosion models over-predict small soil losses and under-predict large soil losses. *Catena* **32**: 15–22.

Nearing MA. 2000. Evaluating soil erosion models using measured plot data: accounting for variability in the data. *Earth Surface Properties and Landforms* **25**: 1035–1043.

Nearing MA, Govers G, Norton LD. 1999. Variability in soil erosion data from replicated plots. *Soil Science Society of America Journal* **63**: 1829–1835.

Poesen J, Vandaele K, Van Wesemael B. 1996. Contribution of gully erosion to sediment production in cultivated lands and rangelands. In *Erosion and Sediment Yield: Global and Regional Perspectives*, Walling DE, Webb BW (eds). *IAHS Publications* no. 236: IAHS: Wallingford; 251–266.

Quinn PF, Beven KJ, Chevallier P, Planchon O. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes* **5**: 59–79.

Quinton JN. 1994. *The validation of physically-based erosion models—with particular reference to EUROSEM*. PhD thesis, Cranfield University, UK.

Quinton JN. 1997. Reducing predictive uncertainty in model simulations: a comparison of two methods using the European soil erosion model (EUROSEM). *Catena* **30**: 101–117.

Refsgaard JC, Storm B. 1995. MIKE SHE. In *Computer Models of Watershed Hydrology*, Singh VP (ed.). Water Resources Publications: Colorado, USA.

Renard KG, Foster GR, Weesies GA, Porter JP. 1991. RUSLE: revised universal soil loss equation. *Journal Soil and Water Conservation* **46**(1): 30–33.

Risse LM, Nearing MA, Nicks AD, Laflen JM. 1993. Assessment of error in the universal soil loss equation. *Soil Science Society of America Journal* **57**: 825–833.

Schmidt J, Werner V, Michael A. 1999. Application of the EROSION 3D model to the Catsop watershed, The Netherlands. In *Modelling of Soil Erosion by Water on a Catchment Scale*, De Roo APJ (ed.). *Catena* **37**(3–4): 449–456.

Smith RE, Goodrich DC, Quinton JN. 1995. Dynamic, distributed simulation of watershed erosion: the KINEROS2 and EUROSEM models. *Journal of Soil and Water Conservation* **50**(5): 517–520.

Souchère V, King D, Daroussin J, Papy F, Capillon A. 1998. Effects of tillage on runoff directions: consequences on runoff contributing area within agricultural catchments. *Journal of Hydrology* **206**: 256–267.

Takken I, Beuselinck L, Nachtergaele J, Govers G, Poesen J, Degraer G. 1999. Spatial evaluation of a physically based distributed erosion model (LISEM). *Catena* **37**: 431–447.

Takken I, Govers G, Jetten V, Nachtergaele J, Steegen A, Poesen J. 2001. The effect of tillage on runoff and erosion patterns. *Soil and Tillage Research* **61**: 55–60.

Torri D, Regués D, Pellegrini S, Bazzoffi P. 1999. Within-storm soil surface dynamics and erosive effects of rainstorms. *Catena* **38**: 131–150.

Van Loon EE. 2001. *Overland flow: interfacing models with measurements*. PhD thesis, Wageningen University, The Netherlands.

Vandaele K, Poesen J, Marques de Silva JR, Govers G, Desmet P. 1997. Assessment of factors controlling ephemeral gully erosion in southern Portugal and central Belgium using aerial photographs. *Zeitschrift für Geomorphologie* **41**(3): 273–287.

Van Rompaey A, Govers G. 2002. Data quality and model complexity for regional scale soil erosion modeling. *International Journal of GIS*: in press.

Van Rompaey A, Govers G, Baudet M. 1999. A strategy for controlling error in distributed environmental models by aggregation. *International Journal of GIS* **13**(6): 577–590.

Vázquez RF, Feyen L, Feyen J, Refsgaard JC. 2002. Effect of grid size on effective parameters and model performance of the MIKE-SHE code. *Hydrological Processes* **16**: 355–372.

Williams JR. 1985. The physical components of the EPIC model. In *Soil Erosion and Conservation*, El-Swaify Sa, Moldenhauer WC, Lo A (eds). Soil Conservation Society of America: Ankeny, IA; 272–284.

Wischmeier WH, Smith DD. 1978. *Predicting Rainfall Erosion Losses*. Agricultural Research Service Handbook 537. US Department of Agriculture.

Zhang XC, Nearing MA, Risse LM, Mcgregor KC. 1996. Evaluation of WEPP runoff and soil loss predictions using natural runoff plot data. *Transactions of the ASAE* **39**(3): 855–863.

