10 A Role for Theoretical Models in Geomorphology?

Michael J. Kirkby
School of Geography, University of Leeds

ABSTRACT

Simulation models provide one of the crucial links between the study of process and the study of landforms, the two traditional activities of geomorphology. Only in exceptional cases can significant changes in landforms be observed directly, so that models provide a means of extrapolating from short-term process measurements to the long-term evolution of macroscopic landforms. The role of models is discussed, mainly in the context of hillslope profile models.

Models provide both improved understanding and forecasting capability. Preferred models are physically based, generally starting from the continuity of mass equations which also provide the formal link between space and time rates of change. Beyond this, many models for landform evolution are still based on gross simplifications of the detailed process mechanics. Effective models should generally be simple, subject to sufficient generality to allow transfer between areas, and of sufficient richness to link to cognate work. Furthermore, there is an important duty on geomorphologists to reconcile models at different spatial and temporal scales.

Although most simulation models may be used in a forecasting mode, perhaps their more important role is as a qualitative thought experiment, testing whether we have a sufficient and consistent theoretical explanation of landscape processes. The best model can only provide a possible explanation which is more consistent with known data than its current rivals. Every field observation, and especially the more qualitative or anecdotal ones, provides an opportunity to refute, or in some cases overturn, existing models and the theories which lie behind them.

The Scientific Nature of Geomorphology: Proceedings of the 27th Binghamton Symposium in Geomorphology held 27-29 September 1996. Edited by Bruce L. Rhoads and Colin E. Thorn.

INTRODUCTION

The science of geomorphology is severely constrained by the normally slow rates of landform change. Our data typically come from cross-sectional studies, either over time in stratigraphic sequences or over space as the distribution of current landforms. In addition we can make relatively short-term process studies in the field or the laboratory. Our preferred interpretations, however, are generally for time and space scales which increase together, from event-based erosion plot models to regional or global models with geological time scales (Figure 10.1). Other combinations can be found in the literature, but are generally less satisfactory. For example, there are many hillslope evolution models which are able to run for a million years or more, but they are deficient in providing the regional setting, particularly in terms of tectonics and basal boundary conditions. Similarly many remote sensing studies provide global snapshots of landscapes, but the geomorphology can only properly be interpreted in terms of regional geological history.

Models, seen as simplifying abstractions of reality, provide the basis for aggregating from the scales of the observations to the scales of interest. In principle a physically based understanding, obtained from process studies, can be applied to explain both current spatial distributions and at-a-point stratigraphic sections, and these cross-sectional studies can be used to calibrate and validate the physical model for more general application at all time and space scales.

We can implement this approach most readily if we are able to make some equilibrium assumptions: preferably equilibrium with respect to either time or space. Note, however, that conditions for equilibrium are themselves scale dependent. We may choose to assume that forms have reached equilibrium over time, because we believe that the duration of uniform conditions is long relative to the system response time. We may then interpret the spatial distribution of forms as a set of equilibrium responses to different climatic, lithological or tectonic environments. Still assuming equilibrium over time, we are able to interpret the stratigraphic record in terms of the palaeogeography of the site relative to

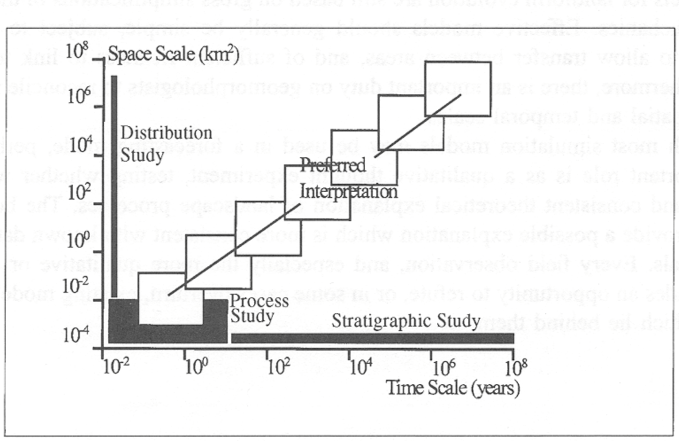

Figure 10.1 The relationship between field studies (shaded) and geomorphological interpretations

shorelines, plate margins, etc. Alternatively, we may choose to assume that equilibrium has been reached over space, because we believe that the lateral extent of uniform conditions lies within a uniform erosional or depositional environment. We then interpret the spatial distribution of landforms in terms of, say, Davisian stage; and the stratigraphic record as responses to environmental change.

In the simplest possible view, we adopt narrow uniformitarianism which allows us to argue directly from analogy. Using this method, we may, for example, infer the impact of global climate change in an area by looking at the landforms and processes acting in an analogue area, which matches the lithology and expected climate or land-use scenario for our area of interest. However, we quickly run into difficulties both because there are no exact analogues (the well-known failure of naïve uniformitarianism) and because of the wide range of relevant relaxation times which coexist in the landscape.

A CONTEXT: HILLSLOPE PROFILE MODELS

To provide a more concrete basis for discussing the properties of models in geomorphology or physical geography, we will focus on one-dimensional models for slope profiles, catenas or flow strips. These typically refer to spatial scales of 0.01-1 km^2, and have been used at time scales ranging from seconds to millions of years (Figure 10.2). As suggested above, some of these time scales are more appropriate and fruitful than others for the single profile, and the slope profile models may also be considered as components within catchments or other larger areal units. The models have been categorised for convenience into three ranges of time scales, which might be loosely linked to Schumm and Lichty's (1965) steady state, graded and cyclic spans, although it is recognised that these categories were originally formulated for rivers rather than hillslopes. For each time span, the lower limit might correspond to the fundamental computational iteration, and the upper limit to

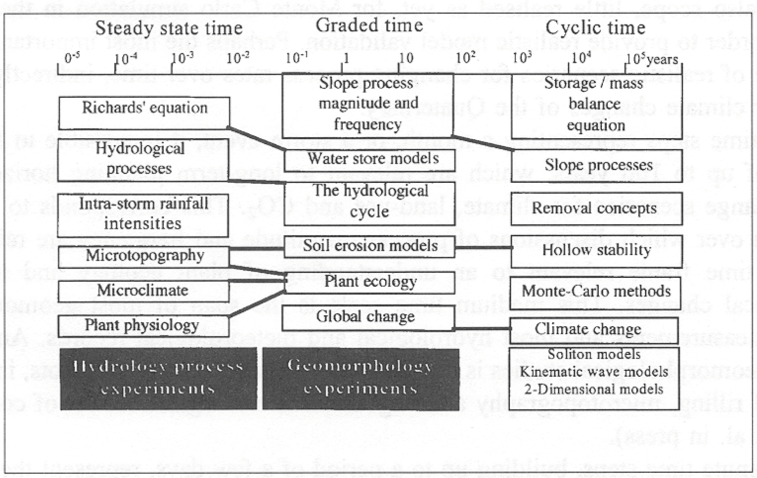

Figure 10.2 Relevant models for slope profiles at a range of time scales. Only the links between time spans are shown

the duration of 1000 such iterations, loosely related to the relevant time span for which a model may sensibly run.

As interpreted here, the time spans refer to hillslope sediment transport rather than to the hillslope form, and the periods involved then relate more closely to those for the states of the fluvial system. Steady-state time is interpreted as a period in which the slope form is essentially static, but with considerable dynamism in soil and surface hydrology. Graded time refers to a period over which hillslope sediment transport can be assigned a mean value and a distribution of magnitudes and frequencies. Cyclic time refers to a period over which sediment transport begins to have a cumulative impact on the form of the hillslope, and over which the parts of the hillslope profile interact with one another via sediment transport. In this sense cyclic time, as used here, includes periods in which the hillslope is both 'graded' and over which it shows net evolution.

Historically, slope models have most commonly combined a mass balance or continuity equation with a set of process 'laws' which express the variation of sediment transport rate in terms of topographic variables, typically as power laws in slope gradient and distance from the divide. Gradient is clearly a direct driver for sediment transport, and distance represents the collecting area for flow, which is important for wash transport. In this form, removal is generally transport (or flux) limited, the processes are effectively limited to diffusive and wash transport, and models may readily (if not realistically) run for iterations of 1000 years, and total periods which represent millions of years. Greater richness and realism may be introduced at the expense of some increased complexity.

The range of slope processes may be widened to include solution and mass movements, as well as taking on the distinctions between rainsplash, rainflow and rillwash. The addition of solution rates also allows changes in regolith depth to be simulated. By introducing the concept of travel distance, the mass balance framework can accommodate both transport (flux) and detachment (supply) limited removal, allowing better representation of mass movement processes especially, and allowing grain-size selective wash. Although one-dimensional models cannot deal with the nonlinearities associated with channel and valley extension, the criteria for stability may be evaluated in one dimension. There is also scope, little realised as yet, for Monte Carlo simulation in the parameter space in order to provide realistic model validation. Perhaps the most important extension is the use of realistic scenarios for changing process rates over time, indirectly reflecting the major climate changes of the Quaternary.

Using time steps representing a month, or a storm event, it is possible to build up to periods of up to 100 years, which are relevant to long-term planning horizons and to global change scenarios for climate, land-use and CO_2 This corresponds to the graded time span over which discussions of process magnitude and frequency are relevant. It is also the time frame relevant to an understanding of plant ecology and to seasonal hydrological changes. This medium time scale is the span of most geomorphological process measurements and most hydrological and meteorological records. An important class of geomorphological studies is related to soil erosion from runoff plots, in which the pattern of rilling, microtopography and vegetation is one significant set of controls (e.g. Kirkby et al. in press).

Five-minute time steps, building up to a period of a few days, represent the time scale appropriate for most detailed hydrological, meteorological and plant eco-physiological process studies. At this scale we may be concerned with representing microtopography, infiltration patterns and overland flow paths in an explicit way; and it is at this level that we generally feel most secure in claiming a physical understanding of the processes. Examples of this level of theory include the Richards equation for movement of water in soils, Reynolds and Darcy-Weissbach roughness equations built into a kinematic cascade for overland flow relationships and the Penman-Monteith equation for evapotranspiration.

These examples will be used to explore our levels of understanding and to assess how far we are from the desired objective of creating a theory, or a family of models, which can be validated against field data, and used to link our knowledge of process and form. In the context of hillslope profiles, it is also worth noting that the stratigraphic dimension in Figure 10.1 is very poorly developed because hillslopes are essentially erosional forms, with minimal preservation potential, so that the need to rely on models is particularly strong.

WHAT SHOULD A MODEL PROVIDE?

Models ideally provide both an insight into the functioning of the natural environment, and a means of forecasting the range of likely outcomes, either in a forecasting sense or to assess the impact of alternative policies. The priority given to these objectives differs according to needs, but in the best practice there should be no conflict. A useful model will be able to make reliable forecasts for environments other than those for which they were originally constructed. As we attempt to combine our best field data to provide estimates for large areas and over long times (Figure 10.1), we will inevitably need to work well outside our original data set, so that we urgently need 'good' models!

Better Understanding

One essential features of any numerical model is that it must provide a logically sufficient and consistent explanation of the process or form it represents. Thus any model offers a view, usually simplified, of our understanding of the system of interest. The insights associated with a useful model, particularly with a conceptually simple model, generally have a much greater impact than any specific forecasts, because they provide components for work in related subfields, and allow new progress to be made, building on the understanding gained. In principle, the same benefits may accrue to non-numerical models which have a formal logical basis, but in practice the logical basis of many qualitative models is less exact, and therefore less effective at revealing any lack of consistency or incompleteness in the explanation.

Once a level of understanding has been achieved, there is generally some scope to assess the criticality and quantitative importance of each process and each state or storage to the overall explanation. Ideally a process of distillation can lead to an essential core of theory, eliminating secondary factors.

These processes of explanation and distillation have always been at the heart of theory development. Numerical modelling may be used to help this process, but may also obscure it with an overemphasis on the quality of the numerical forecasts, and on deriving parameter values rather than meaningful and consistent relationships. Intellectual insight into the working of natural processes is the crucial tool which allows us to go beyond the inductive content of even the best-designed experiment, and structure our world with a network of scientific theories. This search for deeper understanding must lie at the heart of the most significant modelling activity, in geomorphology as in all science, and underpins the critical dialogue between the development of theory and the design of critical experiments.

An important aspect of our search for understanding is to reconcile theories at different scales. Self-consistent theories at each scale typically make generalisations which are not immediately seen to be compatible with those on the next scale, and an important stage in mature theory development is to reconcile theories across scale differences.

Forecasting Potential

Forecasting is both a useful activity and a means of testing the validity and range of our understanding. While the best models may be those with an elegant simplicity, the best forecasts have to apply the understanding of principles to real examples, often in combination with other less complete theories and with strong empirical components.

A model or theory needs to be validated against field data, by comparing its forecasts with observations. In many cases, validation can only be achieved indirectly, since we cannot, for example, observe landscape change over geological time periods. In any case, there is no absolute criterion for acceptance of a theory, and Popper's (1972) view of theories as open to rejection but not acceptance, seems to provide a practicable programme for research. Validation also requires some statistical assessment of goodness of fit, which can be obtained from a distribution of acceptable model outcomes, from a distribution of acceptable real-world outcomes, or some combination of both. However, although forecasts strive towards empirical accuracy, it is clear that most geomorphological models are far from achieving it.

WHAT MAKES A GOOD MODEL?

If a model can be formally constructed from a body of existing theory with the addition of any necessary new development, if the theory takes into account the dominant processes operating, and if the theories are at the scale of interest, then the new theory should be as fundamental as that from which it is derived. Because many of our existing models have only an informal link to an accepted body of theory, development of new models has generally been slow. However, there are a number of factors which characterise a good model; an explicit physical basis, simplicity, generality, richness and the potential for scaling up.

Physical Basis

Models range from totally empirical black box models, often based on regression or neural net methods, to those where all parameters are independently determined physical constants, like $E = mc^2$. Even black box models usually have some physical basis, in the selection of variables and in choosing the form of the regression, arithmetic or logarithmic for example, which expresses some preconceptions about the form of the dependence. For example, the universal soil loss equation estimates soil loss as the product (rather than say the sum) of a series of factors.

In geomorphology, few models rise far above empiricism, and most 'physically based' models are simply pushing the level of empiricism one level further down. For example, we may have a sound physical basis for relating wash sediment transport to water discharge, but we are still forced to assign one or more empirical soil erodibility parameters. As the physical basis improves, identification of model parameters becomes more consistent and allows closer links to be made to other theories, loosely along this scale:

Parameters
- calibrated by model optimisation
- can be consistently obtained from measured values
- directly identifiable from field measurements
- ‘universal' across a wide range of theory

Calibrated parameters can, in general, only be obtained by optimising forecast outcomes against real data. At the next level, the process of parameter estimation becomes the major part of model development, so that many of the factors in the Universal Soil Loss Equation can best be obtained from a set of look-up tables, based on crop types, soil series, conservation practices, etc. At the next higher physical level, we may hope to measure directly soil parameters related to, say, infiltration capacity which have a physical meaning in both our infiltration model and in the field. It is an advantage if these parameters relate to easily identified characteristics such as surface form or vegetation, as there is then much better knowledge of inherent variability and its spatial structure. Finally we might hope to develop models in which the parameters had a wider physical meaning, beyond the confines of the particular model. A few parameters, such as the gravitational acceleration, g, have this wider context, but they are generally in a minority.

Where models have a strong physical basis, this usually provides consistency with other theories. Such consistency helps to support both their validity, by providing additional theoretical support, and their acceptability, by providing more users within the scientific community. Consistency also allows theories to be woven together to explain other related phenomena, and may help to demonstrate consistency across scales, and/or relevant ways to aggregate or disaggregate

For the hillslope models outlined in Figure 10.2, an important physical basis is provided by the mass balance or storage equation, which already constrains the overall behaviour of any model. The equation guarantees continuity of mass in a sense which is central to the models, since the transfer and storage of water and/or soil masses are at their heart. For the long-term models, the remaining physical content is contained in the slope process rate 'laws', which are generally highly empirical, at least in the simplest versions of the models. Nevertheless, even loosely specified physical principles can have a powerful effect in constraining forecasts within reasonable bounds. Consider the following rather non-specific statements about rates of sediment transport on slopes:

> Net sediment transport is in a downslope direction.
> Overland flow is zero at the divide, and increases downslope.
> Wash transport increases more rapidly than overland flow discharge.
> Not all sediment transport depends on overland flow.

These can most easily be interpreted as suggesting that sediment transport takes the form

$$Q_s = K\Lambda^p + Bx^m\Lambda^n$$

where A is slope gradient, x is distance from the divide, and p, m, n, K and B are empirical constants with $m > 1$. The only additional (empirical) assumption implicit in this expression is that the relationships are power laws. The first term represents the non-flow-dependent or 'diffusive' processes, and the second the flow-dependent 'wash' processes. This form of expression was first proposed by Musgrave (1947). One commonly used simple form (Kirkby 1976) is a particular case in which the slope exponents p, n are equal, $m = 2$ and B has been replaced by K/u^2 :

$$Q_s = K\left[1 + \left(\frac{x}{u}\right)^2\right]\Lambda$$

Simplicity

Many models suffer from an excess of complexity, few from being too simple. Although some complex detail may be needed to apply a model in a specific context, the central concept of a good model must be simple. This may be argued pragmatically, both by analogy with existing successful models, and by experience of trying to create models. One part of the need for simplicity comes from the need for the model to be understood and communicated; the other part from the need for the modeller to understand how the model works. As a rule of thumb, it is difficult to construct the core of a model in which more than three (usually more than two) dominant processes interact at a time. For example, an important part of fluvial hydraulics is built on the Reynolds and Froude numbers, which are concerned with deciding which two sets of forces need to be considered in any situation. Similarly, it may be argued that landscape form is controlled primarily by the processes which are dominant in landscape denudation. We may consider the ratio of mass movement rates to wash rates; or of wash rates to solution rates as similar dimensionless ratios which determine the hillslope 'regime'. These ratios are controlled by a number of factors, including climate, topographic situation and lithology.

Figure 10.3 sketches a simple realisation of the process domains for wash, mass movement processes and solution, following the concepts of Langbein and Schumm (1958), and Langbein and Dawdy (1964). Comparing wash and mass movements, it shows rapid mass movements dominant on steep slopes and wash dominant below the landslide threshold gradient, both around the semi-arid peak, and in the humid tropics. For the temperate wash minimum and for very and areas, wash processes are slow enough for

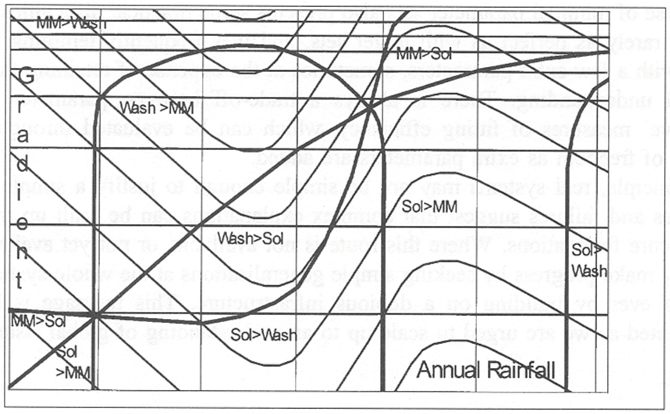

Figure 10.3 Schematic dominance zones for wash, mass movement and solution processes. Axes represent logarithmic scales of annual rainfall and slope gradient. Horizontal lines represent isolines of denudation by mass movement, which is represented as showing a low initial rate of increase for soil creep, and a more rapid increase beyond a threshold for landslides. Vertical lines represent isolines for solutional denudation, increasing linearly with rainfall, at a high rate where rainfall is less than potential evapotranspiration and at a lower rate thereafter. The parallel curves for wash indicate a semi-arid peak, a temperate minimum and a tropical increase, all at rates proportional to gradient. The heavy curves represent the thresholds of equal rates defined for each pair of processes, dividing the field into dominance fields. The captions indicate the two dominant processes in each field

slow mass movements to dominate. Taking account of solution, we may schematically define fields dominated by a pair of dominant processes. Thus the mass movement (MM) > solution (Sol) field may be characterised by periodic stripping of the soil by mass movement, at a rate determined by bedrock weathering; whereas the Sol > MM field is characterised by a creeping saprolite regolith. MM > Wash gives boulder veneered rock slopes, while Wash > MM gives size-sorted pediment and fan surfaces, the two types often linked at a semi-arid break in slope.

Since the majority of geomorphic processes are significantly nonlinear, there is great scope for the development of nonlinearity and chaotic unpredictability in model behaviour. In practice this is often constrained by the strongly diffusive nature of many landscape processes, which allows landscapes, and models representing them, to run forward over time in a stable manner. Diffusive models, or models with a strong diffusive component, tend to converge as a negative exponential towards stable forms. This convergence creates strong equifinality in the evolution, ultimately towards peneplains for a stable tectonic regime, which creates difficulties in assigning a unique model to an observed landform. It also means that models cannot be ran backwards in time, since arbitrarily small initial perturbations then grow exponentially over the reversed time scale.

Large numbers of parameters also tend to provide many opportunities for equifinality in the output, usually including cases where it is qualitatively plain that the 'right' answer is being produced by the 'wrong' set of processes. Furthermore, the process of parameter estimation, either by optimisation or by measurement, becomes increasingly laborious, expensive and indeterminate as the number of relevant parameters increases.

The use of minimal parameter sets also presents some practical difficulties. The quality of fit is rarely as perfect as with larger sets, and it is frequently tempting to tweak the model with a few extra parameters, sometimes at the expense of retaining the underlying physical understanding. There is always a trade-off between parameter number and 'objective' measures of fitting efficiency, which can be evaluated through the loss in degrees of freedom as extra parameters are added.

In principle, real systems may not be simple enough to justify a simple model. Past successes and failures suggest that complex explanations can be built up in small steps from secure foundations. Where this route is not available, or not yet available, then we can only make progress by seeking simple generalisations at the whole-system scale, and rarely if ever by building on a dubious infrastructure. This message is being amply documented as we are urged to scale up to an understanding of global issues.

Generality and Richness

There is always a danger that models for particular areas have no validity outside the field area. A model rises above pure numerical description only when it has some transferability to other areas, or may be applied in other contexts. Clearly any improvement in the physical basis helps to enhance transferability, as it improves the consistency of model parameters. These are important components of the concept of model generality. Models also differ in richness: in how much they help to explain. A rich slope profile model may be able to give some information about soil and vegetation conditions, in addition to the bare form of the profile. Such a model provides greater opportunities for cognate understanding and reduces the risk of equifinal outcomes and model misidentification.

Even a well-specified physical model may, however, face difficulties in being transferred to a new area. One of the main problems is that different processes may be dominant, so that transferability may only be possible over a limited range. For example, a hillslope model may only be valid provided that the pair of dominant processes remains the same. Models may adopt different strategies towards greater transferability, or greater relevance to the conditions of a specific site. If transferability is a high priority, then we may prefer to seek generic rather than specific physical understanding. For example, if a geomorphological model includes explicit response to vegetation cover, the optimal model for a local scale may reflect the particular dominant species, and their response to small differences in environment. For transferring to other areas, or other conditions, it is necessary to include individual curves for each species or functional group, or directly represent the envelope relationship across the whole range of species which might replace the current dominant species, based perhaps on considerations of energy conversion. This is shown schematically in Figure 10.4. In the same way, a broadly relevant hillslope model might either include all relevant processes, or use an envelope relationship. In this case, a series of individual process models may be preferable, because the variety of hillslope processes responds to rather different variables and because of the relatively small number of ftmctional process groups. Even the simplistic view illustrated in Figure 10.3 may be too complex to lend itself to the envelope curve approach, so that geomorphologists may prefer to consider each group of processes separately, and strive to include all processes which might be dominant within the planned range of model transferability.

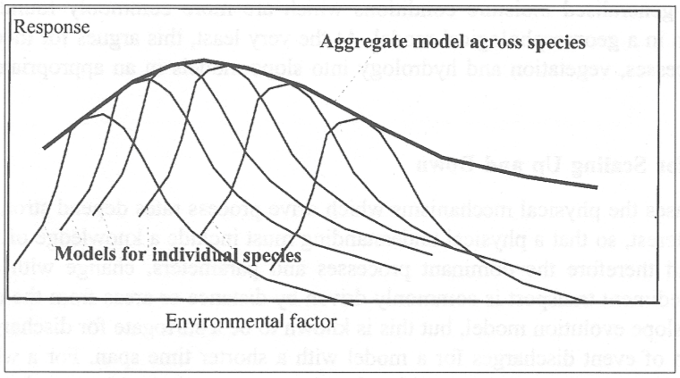

Figure 10.4 Schematic relationship between local and global optima in a model, for the example of vegetation species responses

An important component of richness lies in the notion of the net information gain of the model, defined conceptually as the net change in information content or entropy, comparing output to input values. Highly distributed hydrological models have very large input data requirements, and in many cases are used only to forecast the output hydrograph, so that their net information gain is strongly negative. Lumped models, such as TOPMODEL (Beven and Kirkby 1979), have a more favourable net gain, using a small number of parameters to forecast both outflow hydrographs and, in this example, distributions of saturated area or soil moisture deficits for many points in the catchment.

The information gain is significantly affected by uncertainties attached both to input and output data. Uncertainties in inputs are generally transmitted to the outputs, with gain or attenuation according to the sensitivity of the model to each input. The quality of the output, however, generally also responds to the uncertainties and simplifications built into the model as a representation of reality. Some of this uncertainty is linked to our qualitative categorisation of the model, loosely along this scale:

Principle	e.g. continuity of mass
Law	e.g. Newtonian gravitation
Theory	e.g. plate tectonics
Hypothesis	e.g. the geomorphological unit hydrograph (Rodriguez-Iturbe and Valdez 1979)
Conjecture	e.g. landscape entropy (Leopold and Langbein 1962)

It is clear that few models or theories in geomorphology are far from the bottom of this scale, and that few would agree about which models belong in which categories.

Hillslope models can clearly benefit from attempts to increase their richness. The slope profile is a simple, but a relatively spare description of the landscape. The description, and the associated models, can be greatly enriched by adding the information on soils, vegetation and generalised moisture conditions which are more commonly found in a soil survey than in a geomorphological model. At the very least, this argues for incorporating solute processes, vegetation and hydrology into slope models in an appropriate way.

Potential for Scaling Up and Down

In many cases the physical mechanisms which drive process rates depend strongly on the scale of interest, so that a physical understanding must include a knowledge of how those drivers, and therefore the dominant processes and parameters, change with scale. For example, sediment transport is commonly driven by distance or areas from the divide in a long-term slope evolution model, but this is known to be a surrogate for discharge and the distribution of event discharges for a model with a shorter time span. For a whole-slope model, wash processes may be considered to be transport limited but, at the scale of a runoff plot, travel distances become important and removal is more controlled by detachment factors.

If, and generally only if, a model has an explicit and well-understood physical basis, there is, in principle, the scope to apply it at a range of different scales. Fine-scale models should be capable of aggregation up to coarser scales, although the reverse process of disaggregation is not generally possible without additional insights. In principle, aggregation may be achieved by integrating over relevant frequency distributions, provided that these have a well-behaved structure. In practice this condition is usually met where the distributions have well-behaved means and variances, and that they remain well behaved when combined with the nonlinearities of the system. These conditions are met if, for example, the underlying distributions are normal, exponential or gamma in form, at least for extreme values, and the nonlinearities in, say, sediment transport are in the form of power laws. The extremes of the distribution will then take forms like $x^n\exp(-x)$ or $x^n\exp(-x^2)$, which are themselves of gamma or normal form, and so still have finite means and variances. Two examples of this kind of aggregation are the conversion of event-based wash erosion to an integrated average in long-term slope evolution models (Kirkby and Cox 1995) and the integration of flow depths over microtopography (Kirkby et al. 1995).

In the first of these simplified examples, the overland flow runoff production, j, from a single storm of rainfall r, assuming a fixed soil-water storage threshold, h, is

$$j = (r - h)$$

The frequency density of days with rainfall r is, at the simplest, approximated by the exponential distribution (or as a sum of exponential and gamma terms):

$$N(r) = N_0 \exp(-r/r_0)$$

where N_0, r_0 are empirical parameters fitted to the distribution of daily rainfalls. Summing over this distribution, we obtain the total overland flow production:

$$J = \int_h^\infty (r-h)N(r)\mathrm{d}r = N_0 r_0 \exp(-h/r_0)$$

The sediment yield, assumed proportional to discharge squared, for a single rainfall event of r is

$$t \propto (r - h)^2$$

Again summing over the frequency distribution, the total sediment yield is

$$T \propto \int_h^\infty (r-h)^2 N_0 \exp(-r/r_0)\mathrm{d}r = 2N_0 r_0^2 \exp(-h/r_0)$$

In this case the aggregation is a simple summation process, but it should be observed that the main parameter of the distribution, r_0 appears explicitly in the aggregated form for the climatic erodibility as a strong control on the long-term value.

The second example is for aggregation of overland flows and sediment transport across a rough surface, which is here envisaged as a series of grooves running up- and downslope rather than as terraces or furrows along the contour. For a flow over such a microtopography, stage h may be defined relative to the mean elevation of the surface. The distribution of elevations on the surface can be well approximated empirically by a normal distribution, also referred to the mean elevation, and characterised by a standard deviation, h_0. Points on this surface at elevation z occur with probability density:

$$p(z) = \frac{1}{h_0\sqrt{2\pi}} \exp\left(\frac{-h^2}{2h_0^2}\right)$$

Using this probability density as a weighting, we have, for the mean flow depth, z_0, total water *(q)* and sediment (S) discharge:

$$z_0 = \frac{1}{h_0\sqrt{2\pi}} \int_{-\infty}^{h} (h-z) \exp\left(\frac{-z^2}{2h_0^2}\right)\mathrm{d}z$$

$$q = \frac{c}{h_0\sqrt{2\pi}} \int_{-\infty}^{h} (h-z) \exp\left(\frac{-z^2}{2h_0^2}\right)\mathrm{d}z$$

$$S = \frac{Kc^2}{h_0\sqrt{2\pi}} \int_{-\infty}^{h} (h-z)^2 \exp\left(\frac{-z^2}{2h_0^2}\right)\mathrm{d}z$$

Again it may be seen that the integrals over these distributions remain well behaved, and that the roughness, h_0, appears in the integrated forms as an important determinant of the aggregate rate. These expressions, which can be integrated numerically, show that, for a given overland flow discharge, rougher surfaces create greater concentration of the flow into depressions, and consequently greater sediment transport. The influence of the roughness term is greater at low flows, whereas at flows high enough to inundate the entire surface, the effect is weaker.

It may been seen that, in both of these examples, the effect of processes at the finer scales modifies the rates at the coarser scales. Thus both storm distribution and microtopography may influence long-term rates of slope evolution. In the latter case, it is probable that microtopography not only influences the mean rate of wash transport, but that its variation downslope may also affect the rate of change of sediment transport downslope. Furthermore, the erosional history of the slope is, in turn, likely to control the evolution of the microtopography, in a feedback loop which is largely ignored in current slope models.

An important unifying concept is the span of relevance for each process, in transferring between scales. For example, discussions of landform change in semi-arid environments are concerned primarily with the impacts of climate and imposed land-use. Within discussions of climate impact, there has long (e.g. Leopold and Miller 1954; Cooke and Reeves 1977) been a discussion of the relative importance of changes in total precipitation and in its frequency distribution. The analysis above clearly shows explicit dependence on the frequency distribution of daily rainfalls (through r_0), as well as on total rainfall (roughly equal to $N_0 r_0$). It is clearly legitimate to ask also about what scales are most relevant, even within the frequency distribution.

There are many scaling issues to be addressed. Within the topic of wash erosion, even, we are still not clear exactly how to model the influences of variations in rainfall intensity within storms, although we know it to be important (Yair et al. 1978). The influence of surface stoniness is also clearly important. On soil-covered humid slopes, wash shows little size selectivity, whereas on some stony semi-arid slopes there is strong size sorting which can largely counteract the effect of gradient between about 10^o and 30^o, leading to sharp breaks in slope.

In many cases, a theoretical understanding of how aggregation is achieved can also provide important insights into the magnitude and frequency distribution of the process. For example, the aggregation of event sediment yields into long-term averages, set out in simplified form above, also gives estimates of return periods for dominant events, with clear implications for, among other things, the design of field experiments. Where dominant return periods are long (> 100 years, say) for example, there is little point in carrying out monitoring experiments for a few years, and measurements should be based primarily on extensive surveys.

MODEL VALIDATION?

For most hillslope models, serious attempts at formal model validation are at a very early stage. Validation is more advanced for hydrological models and substantive work has also been done on assessing uncertainties (e.g. Beven and Binley 1992) for some simpler hydrological models. Only a relatively crude approach to uncertainty in forecasts can, however, be applied to most distributed models.

Validation must, in many cases, be preceded by extensive calibration for the less physically based parameters, usually based on optimisation methods, although there are practical difficulties in exhaustive optimisation where there are many parameters. Further problems arise from uncertainties in the calibration data sets, with which forecasts are compared. For many geomorphological models, including those for hillslope evolution, there are major uncertainties about initial landscape forms and dates, about boundary conditions, particularly for basal removal, and about the variations in long-term process rates, taking account of climatic and anthropogenic changes. Thus the best that can generally be achieved is to strive for consistency with both what is known about plausible process rates and their variations, and what is known about the external conditions and constraints. This conclusion may be taken in two ways, which may both be fruitful. On one hand, it argues for treating most models as thought experiments at least as much as for practical forecasting. Alternatively, it requires us to improve our measures of goodness of fit, so that we may evaluate our progress in providing better parameter values or more satisfactory models.

If we construct a landscape evolution model for a slope profile or a catchment, we need to begin from one or a set of logically selected initial forms at a given date; with the changing rates of slope processes in relation to regional knowledge of climate, sea level, land-use and other relevant conditions; and with the surrounding area as reflected in the behaviour of the model boundary conditions, at the outlet(s) and/or divide(s). Goodness of fit to an observed topographic form can be assessed directly as a least-squares departure or similar measure.

How do we combine this measure with other numerical measures, such as that for the water and sediment yield from the profile/catchment outlet? Immediately we have to make a subjective judgement about relative weightings. We may also have forecasts of vegetation cover, soil depth and distributions of soil moisture, among others. It is likely that the quality of these data will be highly variable, particularly in a large area. Some data will be in the form of quantitative surveys, but much will be qualitative, for example in maps of soil capability or erosion sensitivity. Our assessment of changing process rates over time relies on other knowledge (or models) which link rates with climate and land-use, so that this aspect of the landscape model performance is also based on our confidence in other theoretical relationships.

A probabilistic of fuzzy logic scheme may be one way to assess what constitutes acceptable goodness of fit for each distinct criterion. Formally we seek to maximise the probability that a model, defined by a parameter set which includes an error distribution, will provide an acceptable forecast, taking all criteria into consideration. It is important to proceed along some such path of formal optimisation and validation in a framework which allows criteria of different types to be evaluated together, and thus to formalise and confirm our progress in making better forecasts. It is, however, clear that there are many subjective judgements within this process. In other words, there is now, and is likely to remain for the foreseeable future, a key role, and I would argue the key role, for models as essentially qualitative thought experiments. The quantitative nature of most computational and mathematical models obscures this role, by appearing to offer a precision which is generally spurious. Nevertheless, we are still in the position where only a few of our variables are known within an accuracy of $\pm$ 50%, and where one good field observation can overturn, or at least radically modify, what had previously seemed to be a well-founded model.

CONCLUSION

Models in geomorphology form an essential bridge between the scales at which we can observe and the generally larger scales at which we seek explanations and forecasts. At present our models are most effective as thought experiments which help to refine our understanding of the dominant processes acting. This theoretical understanding is best achieved though models which have a strong physical basis, a high degree of generality, and a richness and power which link them to other areas of environmental and earth sciences. One major challenge lies in the search for improved methods for testing the quality of model forecasts, taking full account of the richness in the forecasts, and combining qualitative and quantitative criteria for success. The other major challenge for modelling is to address problems of up-scaling, for application in global change and geophysical contexts. This not only requires new model concepts, but an explicit understanding of compatibility between scales of interest.

REFERENCES

Beven, K. and Binley, A.M. 1992. The future of distributed models: calibration and predictive uncertainty, *Hydrological Processes,* **6**, 279-298.

Beven, K.J. and Kirkby, M.J. 1979. A physically based, contributing area model of basin hydrology, *Hydrological Sciences Bulletin,* **24**, 43-69.

Cooke, R.U. and Reeves, R.W. 1977. *Arroyos and Environmental Change in the American Southwest,* Clarendon Press, Oxford, 213 pp.

Kirkby, M.J. 1976. Tests of the random network model and its application to basin hydrology, *Earth Surface Processes,* **1**, 197-212.

Kirkby, M.J. Abrahart, R., McMahon, M.D., Shao, J. and Thornes, J.B. in press. MEDALUS soil erosion models for global change, paper presented at the 3rd International Geomorphology Conference, Hamilton, Ontario (for publication in an edited volume).

Kirkby, M.J. and Cox, N.J. 1995. A climatic index for soil erosion potential (CSEP), including seasonal factors, *Catena,* **25**, 333-352.

Kirkby, M.J. McMahon, M.L. and Abrahart, R.J. 1995. *The MEDRUSH model, MEDAL US II Final Report,* 1/1/93-30/9/1995.

Langbein, W.B. and Dawdy, D.R. 1964. Occurrence of dissolved solids in surface waters of the United States, *USGS Professional Paper 501-D,* 115-117.

Langbein, W.B.. and Schumm, S.A. 1958. Yield of sediment in relation to mean annual precipitation, *Transactions American Geophysical Union,* **39**, 1076-1084.

Leopold, L.B. and Langbein, W.B. 1962. The concept of entropy in landscape evolution, *USGS Professional Paper 500-A,* 20 pp.

Leopold, L.B. and Miller, J.P. 1954. A post-glacial chronology for some alluvial valleys in Wyoming, *USGS Water Supply Paper 1261,* 90 pp.

Musgrave, G.W 1947. The quantitative evaluation of factors in water erosion, a first approximation, *Journal of Soil and Water Conservation,* **2**, 133-138.

Popper, K.R. 1972. *The Logic of Scientific Discovery,* 6th revised impression, Hutchinson, London.

Rodriguez-Iturbe, I. and Valdez, J.B. 1979. The geomorphological structure of hydrological response, *Water Resources Research,* **15**, 1490-1520.

Schumm, S.A. and Lichty, R.W. 1965. Time, space and causality in geomorphology, *American Journal of Science,* **263**, 110-119.

Yak, A., Sharon, D. and Lavee, H. 1978. An instrumented watershed study of partial area contribution runoff in the arid zone, *Zeitschrift für Geomorphologie,* Supplementband, **29**, 71-82.