Hydrology and Earth System Sciences, 3(4), 477–489 (1999) © EGS

Multi-objective conditioning of a simple SVAT model

Stewart W. Franks[1], Keith J. Beven[2] and John H. C. Gash[3]

[1] Department of Civil, Surveying and Environmental Engineering, University of Newcastle, Callaghan 2308, New South Wales, Australia
[2] Institute of Environmental and Natural Sciences, Lancaster University, Lancaster, LA1 4YQ, U.K.
[3] Institute of Hydrology, Wallingford, OX10 8BB, U.K.
e-mail address for corresponding author: ceswf@civeng.newcastle.edu.au

Abstract

It has previously been argued that current Soil Vegetation Atmosphere Transfer (SVAT) models are over-parameterised given the calibration data typically available. Using the Generalised Likelihood Uncertainty Estimation (GLUE) methodology, multiple feasible model parameter sets are here conditioned on latent heat fluxes and then additionally on the sensible and ground heat fluxes at a single site in Amazonia. The model conditioning schemes were then evaluated with a further data set collected at the same site according to their ability to reproduce the latent, sensible and ground heat fluxes. The results indicate that conditioning the model on only the latent heat flux component of the energy balance does not constrain satisfactorily the predictions of the other components of the energy balance. When conditioning on all heat flux objectives, significant additional constraint of the feasible parameter space is achieved with a consequent reduction in the predictive uncertainty. There are still, however, many parameter sets that adequately reproduce the calibration/validation data, leading to significant predictive uncertainty. Surface temperature measurements, whilst also subject to uncertainty, may be employed usefully in a multi-objective calibration of SVAT models.

Introduction

It is becoming increasingly widely recognised that complex environmental models are generally over-parameterised with respect to typically available calibration data (see, for example, Beven, 1989; Duan *et al.*, 1992; Franks *et al.*, 1997a,b; Yapo *et al.*, 1998; Beven, 1999). Even for relatively well defined problems such as predicting the components of the surface energy balance, any attempt to represent the wide range of physical processes involved introduces too many degrees of freedom in the form of an excessive number of physical and physiological parameters to be specified. One consequence is that time series of calibration/validation variables are reproduced to acceptable accuracy by parameter sets from many parts of the feasible parameter space (e.g. Franks and Beven, 1997a; Franks *et al.*, 1997a,b) and it has been argued that the perceived necessity of incorporating ever increasing complexity into models may necessitate an explicit recognition that many different parameter sets will provide an acceptable fit to available observations (the *equifinality* problem of Beven, 1993).

One way of reducing the possibility of non-uniqueness or equifinality of parameter sets in the calibration of complex models with high dimensional parameter spaces is to increase the information of content of the calibration data. This may be achieved with longer periods of calibration data, but perhaps more importantly, by using observations of different, independently measured variables (see for example, Franks and Beven, 1997a; Klepper, 1997; Mroczkowski *et al.*, 1997; Gupta *et al.*, 1998; Kuczera and Mroczkowski, 1998; Yapo *et al.*, 1998; Lamb *et al.*, 1998). Such data may usefully provide further objective functions for model evaluation and hence constraint on the feasible parameterisations and resulting model predictive uncertainty.

In the context of rainfall-runoff modelling a number of studies have demonstrated the utility of conditioning on additional objectives (Chappell *et al.*, 1998; Franks *et al.*, 1997 a,b; Gupta *et al.*, 1998; Mroczkowski *et al.*, 1997). As such, multi-objective conditioning offers greater power in terms of discriminating between model structures and feasible parameter sets within any SVAT model structure.

In modelling the surface energy balance using a Soil Vegetation–Atmosphere Transfer (SVAT) model, Franks and Beven (1997a) demonstrated significant uncertainty associated with SVAT predictions following calibration of a simple SVAT model to latent heat fluxes alone. There is, however, the potential to use measurements of all heat flux components in the surface energy balance in a multi-objective conditioning scheme to constrain the feasible parameter space. During the daytime period, incoming net radiation may be partitioned into the two dominant fluxes of latent and sensible heat, and a further smaller component, the ground heat flux. Typically, SVAT models are calibrated on measured latent heat flux alone. Multi-objective calibration to all heat flux variables should, in principle, increase the calibration information content and hence reduce the parametric and predictive uncertainty. Other variables may also be included in multi-objective conditioning schemes such as sub-surface hydrological variables and appropriate surface temperature measurements.

This present study represents an extension of that by Franks and Beven (1997a) to evaluate the utility of conditioning a SVAT model on multiple objectives. In this paper, a relatively simple SVAT model is tested within the Generalised Likelihood Uncertainty Estimation (GLUE) framework. Multiple parameterisations are selected from feasible ranges and their ability to reproduce time series of surface fluxes is evaluated. After calibrating the SVAT model against latent heat fluxes alone, feasible parameter distributions are assessed. Prediction bounds for a different period of data collected at the same site are then derived and evaluated. Multi-objective conditioning is then performed using all the observed components of the surface energy balance and the derived prediction bounds for the evaluation period are compared with those derived by single objective calibration.

This study therefore seeks to investigate the utility of conditioning a simple SVAT model on additional heat fluxes, in terms of increasing the constraints on the feasible parameter sets and the consequent predictive uncertainty. Additionally, model-output aerodynamic surface temperatures are then examined to indicate the potential utility of radiative temperature measurements in reducing uncertainty associated with SVAT model parameterisations.

Adaptation of the TOPUP-SVAT model

The TOPUP-SVAT model (Beven and Quinn, 1994; Franks *et al.*, 1997b) is an attempt to represent some of the key physical processes in a realistic but parametrically parsimonious way (Fig. 1) at the scale of a landscape patch. A novel feature of the model is that it includes a representation of the lateral redistribution of moisture provided by a dynamic water table following the basic concepts in TOPMODEL (Beven *et al.*, 1995). In previous applications, the TOPUP-SVAT model has utilised the Penman-Monteith equation (Monteith, 1981) to predict latent heat fluxes (see Franks *et al.*, 1997b for more details). As the aim of this study is to assess the utility of additional conditioning

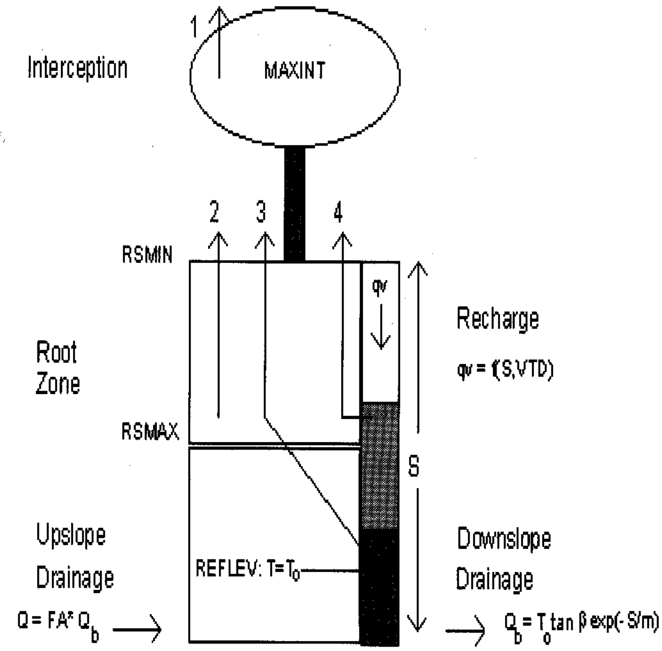

Fig. 1. *Schematic of the TOPUP SVAT. Note that in this application no water table is specified.*

afforded by all measured heat fluxes, the TOPUP-SVAT model was adapted so as to provide predictions of each component of the energy balance. Additionally, the Penman-Monteith equation eliminates the surface temperature from the calculation of latent heat fluxes through the use of the gradient of the saturation vapour pressure versus temperature function (the 'del' function, see for example Calder, 1990). To achieve closure of the energy balance, retaining the surface temperature as a predicted variable, the following equations were used;

$$R_n - G = \lambda E + H \qquad (1)$$

$$\lambda E = \frac{\rho C_P}{\gamma} \frac{(e_0 - e)}{r_a} \qquad (2)$$

$$\lambda E = \frac{\rho C_P}{\gamma} \frac{(e_s[T_0] - e_0)}{r_s} \qquad (3)$$

$$H = \rho C_P \frac{(T_0 - T_a)}{r_{ah}} \qquad (4)$$

where R_n is the net radiation, G is the ground heat flux, λE is the latent heat flux, H is the sensible heat flux, all in units of Wm^{-2}; T_0 is surface temperature, $e_s(T_0)$, e_o, e are the vapour pressures within the leaf stomata, at the leaf surface, and of the bulk atmosphere, respectively, ρ is the density of air, and C_P is the specific heat of air. Given a reliable measure of $R_n - G$, the partitioning of this incoming energy flux into latent and sensible heat fluxes may be achieved through the use of the above equations, describing aerodynamic evaporation, plant physiological evaporation, and sensible heat flux, respectively. The solution to the above equations is achieved through a robust iterative scheme.

Previous studies have investigated the soil heat flux as a function of net radiation, R_n (e.g. Clothier *et al.*, 1986; Kustas and Daughtry, 1987; amongst others). The study by Clothier *et al.* (1986) revealed that the fraction of the net radiation contributing to the soil heat flux followed a diurnal cycle varying with the net radiation. Similar dynamics were also observed in the measured field data. In this study, a simple model was used to calculate the soil heat flux providing the dynamics observed by Clothier *et al.* (1986);

$$G = aR_n \tag{5}$$

$$a = b\frac{R_n}{R_n^{MAX}} \tag{6}$$

where a is the fraction of net radiation that is partitioned into soil heat flux at any instant, which is defined by a maximum fraction, b, and the ratio of instantaneous net radiation to the maximum net radiation, R_n^{MAX} (Eqn. 6). Note that this sub-model component has deliberately been kept as simple as possible (only one additional parameter) so as to restrict any increase in the dimensions of the parameter space. The number of parameters would not be a problem if additional effective parameter values required by the model could be uniquely measured in the field. However, for a patch scale SVAT model, this will not be the case even for physical parameters such as the soil thermal diffusivity.

The GLUE methodology

The GLUE methodology was developed as an extension of the Generalised Sensitivity Analysis of Spear and Hornberger (1980) to provide a means of assessing the predictive uncertainty of complex nonlinear models based on generalised likelihood measures. Its application so far has been predominantly in rainfall-runoff modelling (Beven and Binley, 1992; Beven, 1993; Romanowicz *et al.*, 1996; Freer *et al.*, 1996; Franks *et al.*, 1998; Lamb *et al.*, 1998; Cameron *et al.*, 1999) but has also been previously applied to the TOPUP-SVAT model (Franks and Beven, 1997a, b, 1999; Franks *et al.*, 1997b), in the prediction of critical loads (Zak *et al.*, 1997; Zak and Beven, 1999) and modelling soil nitrate fluxes (Schulz *et al.*, 1999). The sampling procedure used in this application of GLUE is based upon Monte Carlo simulation; a large number of model runs are made, each parameterised with independently chosen random values of the parameters selected from uniform distributions across the range of each parameter. The acceptability of each run is assessed by some chosen likelihood measure, calculated from a comparison of observed and simulated responses. The likelihood weights of the runs are rescaled so that their cumulative total is 1.0. At each time step, the predicted output from the retained runs are likelihood weighted and ranked to form a cumulative distribution of the output variable from which chosen quantiles can be selected to represent the model uncertainty.

Whilst GLUE contains a number of subjective elements (for example, the prior choice of parameter ranges and in the choice of the likelihood measure employed) it forces those choices to be made explicit. A large number of computer runs is also required, particularly for models with a large number of parameters. The uniform sampling generally used will in some cases be inefficient in comparison with the type of importance sampling of Monte Carlo Markov Chain methods but ensures the orthogonality of the chosen parameter sets in terms of sampling the parameter space. In practice, sampling efficiency has not proven to be a particular constraint, especially since Monte Carlo simulation is well suited to parallel computer systems. Against the disadvantage of computer run times is the considerable advantage that the approach is conceptually very easy to understand and easy to implement.

Meteorological Forcing Data

Meteorological forcing data and flux measurements were employed for an Amazonian, post-deforestation pasture site, collected as part of the ABRACOS UK-Brazilian collaboration (Shuttleworth *et al.*, 1991; Gash *et al.*, 1996; Gash and Nobre, 1997). This site is located at Fazenda Dimona, central Amazonia. The data set employed covers the period 29th June—10th September, 1991, (Wright *et al.*, 1995). This data set was divided to provide calibration and evaluation data sets, to test the ability of model calibration schemes (e.g. Franks and Beven, 1997a). The latent, sensible and ground heat flux records were measured independently. Details of the instrumentation employed at the site are given in Wright *et al.* (1992).

Selection of Parameter Ranges and Sampling Strategy

Parameter ranges may be selected for each parameter based on physical argument and experience. The *a priori* definition of parameter ranges to be considered already discounts parameter values outside these ranges, in effect, specifying values outside these ranges as having a likelihood of zero. The specification of the sampling ranges used in this application of the GLUE method follows the previous study of TOPUP SVAT model behaviour by Franks *et al.* (1997b). The parameter ranges used in this study are shown in Table 1. To sample the parameter space, 10,000 individual parameter sets were constructed with each constituent parameter value being randomly chosen from uniform distributions from each parameter residual series (even after allowing for heteroscedasticity and autocorrelation effects).

In this study, an appropriate measure must be selected that can treat the multi-variate problem at hand. A number of earlier studies have employed least-squares based likelihood measures under the assumption of an error model based on zero-bias, normally-distributed errors (e.g. Kuczera, 1983).

The multi-variate problem can be expressed as below;

$$\begin{bmatrix} LE_i \\ H_i \\ G_i \end{bmatrix} = \begin{bmatrix} f(\underline{\theta}_i, \underline{Y}) \\ g(\underline{\theta}_i, \underline{Y}) \\ h(\underline{\theta}_i, \underline{Y}) \end{bmatrix} + \begin{bmatrix} \alpha_i \\ \beta_i \\ \chi_i \end{bmatrix} \quad (9)$$

where, LE_i, H_i and G_i are the latent, sensible and ground heat fluxes observed at time, i, $\underline{\Theta}_i$ is a given parameter set, $\underline{Y}$ is the set of forcing observations, $f(\ )$, $g(\ )$, $h(\ )$, α, β, and χ represent the model structure and prediction errors for the prediction of latent, sensible and ground heat fluxes, respectively.

Box and Tiao (1973) define a simple likelihood measure as;

$$L(\underline{\Theta}_i|\underline{Y}) \propto (S_\alpha^2)^{-n/2} \quad (10)$$

where $S_\alpha{}^2$ represents the sum of squared errors α_i for a given parameter set, θ, and n is the number of independent measures of a variable. Franks and Beven (1997a) used a similar likelihood measure in conjunction with the TOPUP-SVAT model with the addition that the variance of the errors was raised to a power of $-N$, where N is a subjectively chosen shaping parameter to accentuate the form of the response surface

In this study, as multiple objectives are to be used, then there is the additional requirement that each variable is treated *a priori* equally. To achieve this, the variance of the errors may be normalised by the minimum estimate of the variance of the errors for each of the different objectives (heat fluxes). A simple single objective likelihood measure may therefore be defined as;

$$L(\underline{\Theta}_i \mid \underline{Y}, \sigma_\alpha^2) \propto \left(\frac{\sigma_\alpha^2}{\hat{\sigma}_\alpha^2}\right)^{-n/2} \quad (11)$$

where θ is the parameter set, $\underline{Y}$ is the set of forcing data, $\sigma_\alpha{}^2$, is the variance of the errors for the variable of interest, and $\hat{\sigma}_\alpha{}^2$ represents the maximum likelihood estimate of the variance of the errors of the variable. In terms of the multi-objective problem, the normalised measures for each individual variable may be combined through Bayes Equation, thus;

$$L(\underline{\Theta}_i|\underline{Y}, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\chi^2) \propto L(\underline{\Theta}_i|\underline{Y}, \sigma_\alpha^2) L(\underline{\Theta}_i|\underline{Y}, \sigma_\beta^2) L(\underline{\Theta}_i|\underline{Y}, \sigma_\chi^2) \quad (12)$$

The normalisation of the error variance for the three variables ensures that the derived likelihood measure treats each of the variables *a priori* equally. The developed likelihood measure therefore explicitly accounts for the inherently different ability of the model to simulate the different objectives. This is particularly important in the calibration of a patch (or plot) scale SVAT model as numerous sources of error may exist in the different data sets. Both latent and sensible heat flux measurements are made at a point in space at some measurement height. However, the measured fluxes are representative of areally-integrated flux variability, where the area sampled is a function of fetch and wind direction (and hence, the variable micro-meteorological data). Such measurements cannot, therefore, be treated as precise point measures. In contrast, the measures of ground heat fluxes are made with heat plates that are set in fixed point locations. Because of the mis-match of scales at which the various fluxes are made, an inherent residual in the total energy balance must be expected at any point in time. In contrast to the inherent residual in the observed energy balance, the model must always balance these energy components. By normalising the error variances of each of the fluxes by the minimum error variance in the likelihood measure, we have explicitly acknowledged the variable degree to which the different heat fluxes can be reproduced by the model.

At this point it is worthwhile noting that a traditional statistical likelihood function in terms of Eqn. 10 requires raising the error variance term to the power of $-n/2$, where n is the equivalent number of independent observations. Equation 12 also implicitly assumes independence of the three residual series. In general, n will be less than the number of observations because of serial correlation and cross-correlation in the residuals but, for the number of time steps used in this study, will still be expected to be large. Such a function therefore tends to provide an overly stringent measure of the acceptability of a given parameter set, especially given an expectation that even the best model parameterisation will not simulate all aspects of the observations. A power transformation can, however, be used as an empirical shaping function to control the stringency/leniency of the required objective function. Whilst the choice of this scaling parameter is necessarily subjective, once assigned, prediction bounds can be derived representing a consistent test of the model's ability to reproduce further evaluation data sets. The final likelihood measure used in this study is therefore derived as;

$$L(\underline{\Theta}_i|\underline{Y}, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\chi^2) \propto \left(\frac{\sigma_\alpha^2}{\hat{\sigma}_\alpha^2} \frac{\sigma_\beta^2}{\hat{\sigma}_\beta^2} \frac{\sigma_\chi^2}{\hat{\sigma}_\chi^2}\right)^{-N} \quad (13)$$

where N represents the empirical scaling factor. For constant error terms, increasing the value of N effectively increases the number of time steps and hence the assumed informative content of the data (i.e. towards fully independent observations). This has the effect of adding more weight to the parameter sets that produce the smallest residuals. The effect of this parameter on the derived prediction bounds will be assessed below.

As the derived likelihood measure requires the specification of the minimum estimated variance of errors for each modelling objective, then an iterative procedure is required. Initially, the optimal variance of the errors for each of the variables of interest is calculated according to the best reproduction from *any* parameter set. These optimal error variances are then applied in Eqn. 13 to identify the minimum error variance (optimal) set (i.e. the parameter set with minimum error variances for all fluxes). The final stage requires the re-application of Eqn. 13 using the identified minimum error variance (optimal) set.

Results

Following the methodology outlined above, 10000 parameter sets were selected from the feasible parameter space (Table 1) and the model was run for each parameter set, forced with the ABRACOS data set. This data set was divided into sections of equal length. The first period of data was used as the conditioning period, hereafter referred to as the calibration data set. Using these data, likelihood values were derived for each parameter set following the reproduction of the latent, sensible and ground heat fluxes records, for both single and multiple objective conditioning. The second period of data was used as an evaluation data set whereby predictive uncertainty, associated with the conditioning on the latent heat fluxes, and then all fluxes, could be assessed.

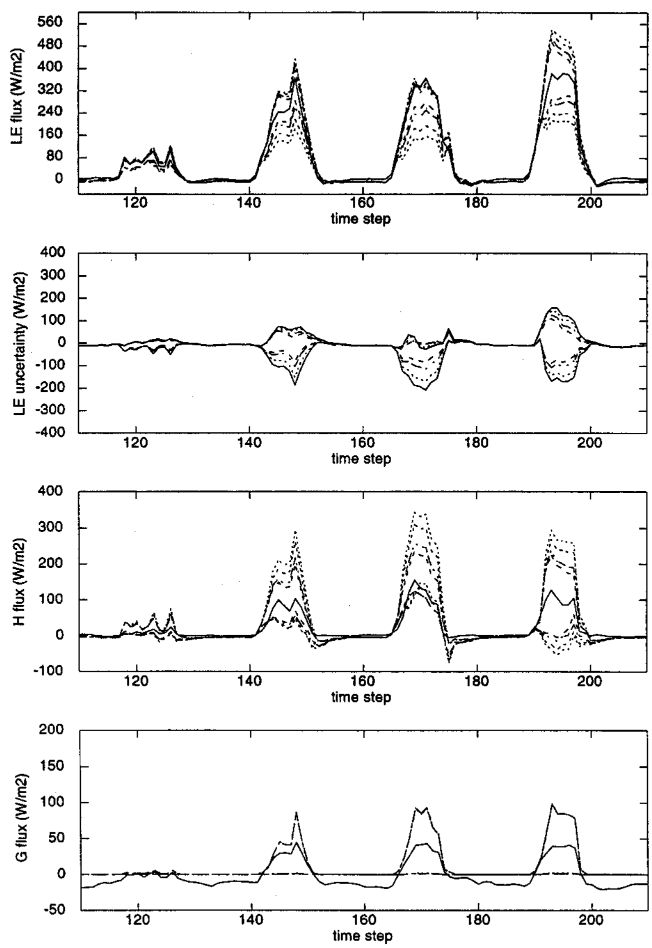

Fig. 2. *Derived prediction bounds for a section of the calibration data set of (a) latent heat flux, (b) the residual differences between the observed and the prediction bounds, (c) the sensible heat flux and (d) the ground heat flux. Note that successively constrained prediction bounds correspond to N = 1, 2.5, 5, and 7.5.*

Assessing the effect of the shaping parameter, N, on derived prediction bounds

Figure 2 shows the prediction bounds associated with the prediction of all heat fluxes for a section of the calibration data set following conditioning on the latent heat flux alone. For clarity these have been scaled by subtracting the observed data value for each time step so that zero represents a zero error. Figure 2 shows that the effect of increasing N is to narrow or constrain the uncertainty envelope. Figure 3 shows the 95% prediction bounds limits for the evaluation data period, conditioned on the latent heat flux measurements alone. As such, for a 'validated' model structure and calibration, then the residual prediction bounds should centre on zero for the diurnal predictions, if the prediction bounds encompass the observed flux.

As can be seen, as the value of N is increased, for values of $N = 5$ and 7.5, then the prediction bounds do not enclose the observed data (as indicated by the residual prediction bounds not being centred on zero). This indicates that the observed data is outside the uncertainty envelope and hence that the model is apparently biased in its predictions during some of the simulation period.

By increasing the value of N, the likelihood measure is made more stringent in that 'better' simulations are given higher likelihoods than those performing less well. As stated earlier, the N shaping parameter is used in this study to reflect the degree of belief in the information content of the data, relative to the limitations of the model and the observations. Use of a high value of N is only really justified where it can be assumed that a true model of the process exists, otherwise it will result in an over-conditioning of the parameter space towards a single optimum parameter set. The limitations of any SVAT model suggest that more leniency is required in the definition of the likelihood measure. Therefore, for the rest of this study a value of $N = 2.5$ is adopted representing a reasonable compromise between constraining the prediction limits of the simulations and encompassing of the observed data. As

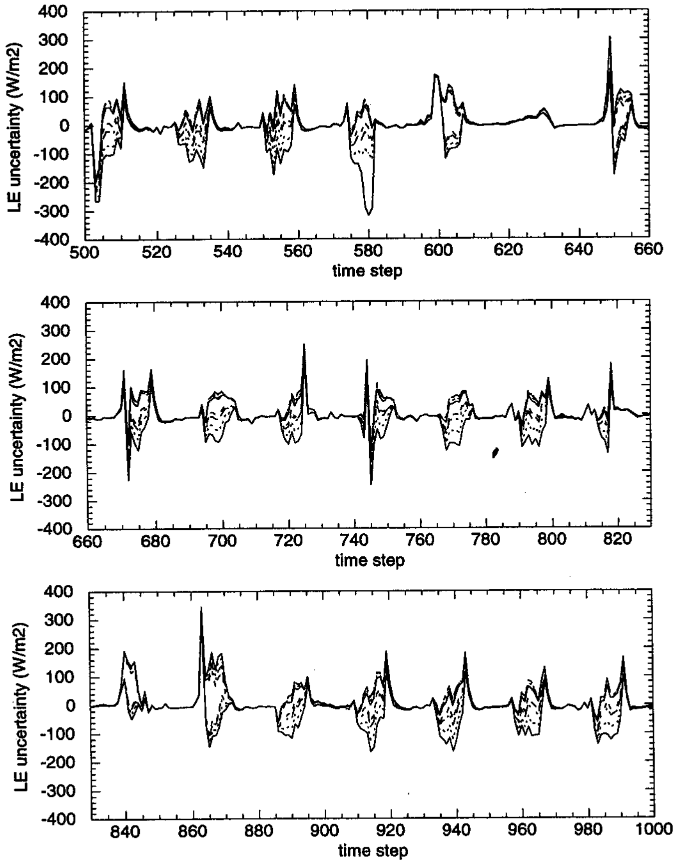

Fig. 3. *Latent heat flux residual prediction bounds propagated for the evaluation data set. Note that successively constrained prediction bounds correspond to N = 1, 2.5, 5, and 7.5.*

such, the N parameter has not been arbitrarily prescribed but assigned on the basis of the implicit effects of limitations of the data and the simplified model structure.

Assessing the constraint of the feasible parameter space

To assess the effect of conditioning on the different heat flux variables on the feasible parameter space, cumulative likelihood plots for each parameter may be used (see Franks *et al.*, 1997a). As the *a priori* parameter sets were selected from uniform distributions, then the prior cumulative likelihood plots should show a straight line. If, after conditioning, there exists a tendency towards a particular sub-range of that parameter from the initial distribution, then the reproduction of the single or multiple flux variables is revealed as showing some sensitivity to that parameter (remembering that it is the parameter set that gives a good or bad fit to the data). A posterior straight line cumulative distribution therefore represents a lack of effective conditioning of a particular parameter, whilst a significant deviation from a straight line demonstrates that the feasible values of that parameter have been constrained despite potential parameter interactions. Such plots may therefore be used to infer significant conditioning of the parameters.

Figure 4 shows such cumulative likelihood plots for four of the TOPUP-SVAT model parameters related to the moisture stores and surface resistances. These parameters are SRMAX, the root zone store, MAXINT, the interception canopy store, RSMIN, the minimum surface resistance, and RSMAX, the maximum surface resistance. These plots present the posterior parameter distributions after conditioning on latent heat fluxes alone, multi-objective conditioning on the latent and sensible heat fluxes, and finally the multi-objective conditioning on all heat fluxes.

Figure 5 shows the cumulative likelihood plots for four of the TOPUP-SVAT model parameters relating to the aerodynamic characteristics and the ground heat flux parameterisation. These parameters are z_o, the roughness length for momentum, $\ln(z_o/z_{oh})$, the natural log of the ratio of the roughness lengths for momentum and heat, d, the displacement height, and b, the ground heat flux parameter (Eqn. 6). From Fig. 4, it can be seen that only the RSMIN parameter remains uniform in each of the posterior cumulative distributions following each of the conditioning schemes. As such, it can be stated that the results are insensitive to this parameter.

As can be seen, the SRMAX and RSMAX parameter distributions following the additional constraint on the sensible heat flux record are significantly different to those derived by calibration to latent heat fluxes alone. Figures 5a,b also show the same effect for the aerodynamic parameters [z_o, $\ln(z_o,z_{oh})$]. It can therefore be stated that conditioning on the sensible heat flux improves the constraint of these parameters—this is due to the greater sensitivity of the sensible heat fluxes to the aerodynamic parameters that produce appropriate aerodynamic resistances for solution of Eqn. 4.

As the aerodynamic resistance term is common to both calculations of sensible heat flux (Eqn. 4) and latent heat flux (Eqn. 2), then the constrained aerodynamic resistances produce a commensurate refinement of the distributions of those parameters which most strongly control the surface resistance terms (SRMAX and RSMAX; Franks *et al.*, 1997b). It is therefore apparent that additional conditioning of the TOPUP-SVAT model on the sensible heat fluxes produces a significant reduction in the parameter uncertainty. It should however be noted that whilst the parameter distributions are constrained, the feasible parameter ranges are still wide indicating non-uniqueness of parameter sets in reproducing the calibration data.

Figure 5c shows the distributions of the ground heat flux parameter, b. It can be seen that for the reproduction of the latent heat flux alone, the results show relative insensitivity. This is in marked contrast to the distribution of this parameter following multi-objective conditioning which displays a significant constraint on the range of values associated with high likelihood parameter sets. This

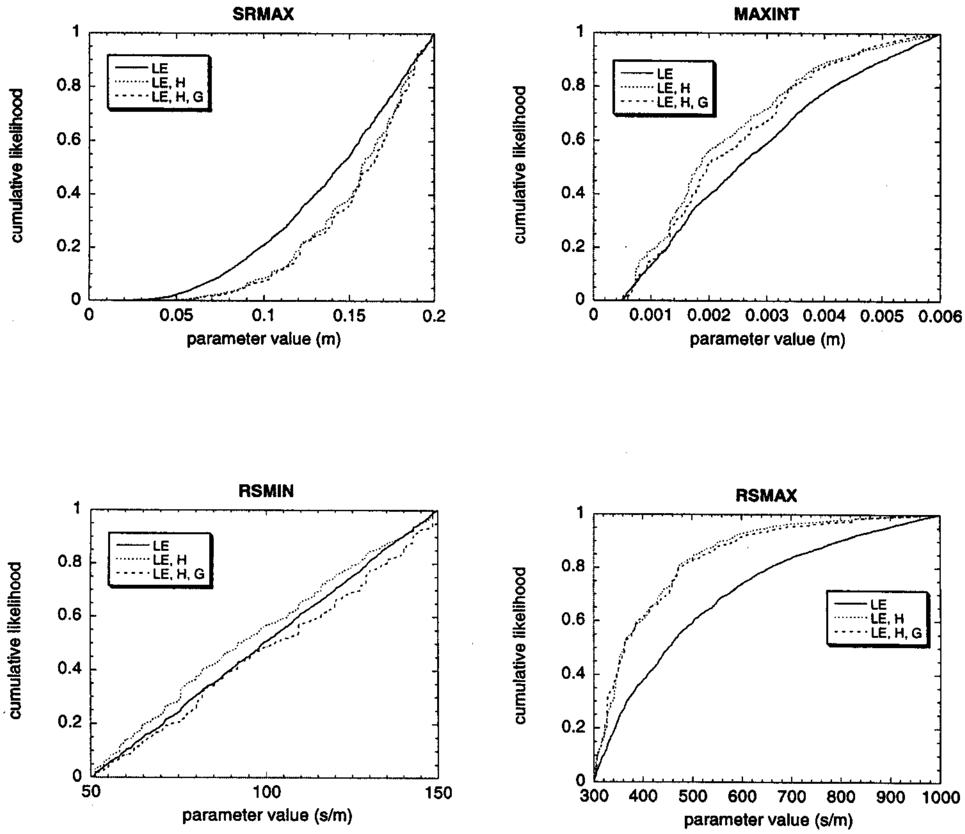

Fig. 4. *Cumulative likelihood plots for four of the model parameters related to the model moisture stores.*

can be interpreted as follows. The reproduction of the latent heat flux alone does not significantly constrain the ground heat flux; good reproduction of the latent heat flux can be achieved with any parameter value from the chosen *a priori* range (conditional on values of the other parameters). It is only when the model is required to reproduce additional heat fluxes (i.e. multi-objective conditioning) that this parameter displays significant conditioning . The ground heat flux parameter is therefore markedly constrained by the multi-objective conditioning. It should be noted that this is due partly to the fact that the energy balance is a 'bounded' problem in the sense that there is a strong dependence between each of the predicted fluxes as they must sum to equal the input net radiation. Without the additional observations, many different partitionings of modelled fluxes between sensible heat and ground heat flux could result in the similar predicted latent heat fluxes.

Assessing predictive uncertainty following multi-objective conditioning

To evaluate the potential constraint of the uncertainty associated with the prediction of the heat fluxes following multi-objective conditioning, prediction bounds were determined for the simulation of the evaluation data set. Figure 6 shows the derived prediction bounds propagated for the latent heat flux. In this plot, the solid lines represent conditioning on the latent heat flux alone whilst the dashes represent multi-objective conditioning on all heat fluxes. From this plot, it can be seen that significant improvement in the predictive uncertainty is attained through multi-objective conditioning. Indeed, the different prediction bounds can be visually differentiated on the upper bounds of the second diurnal response. Figures 7 and 8 show the prediction bounds propagated for the sensible and ground heat fluxes. In these plots, the solid lines represent the conditioning on latent heat flux alone and the dashes represent the multi-objective conditioning. As can be seen, the magnitude of the uncertainty envelopes show significant differences whilst largely encompassing the observed data. As such, it can be stated that multi-objective conditioning has been shown to be beneficial in terms of reducing the space of feasible models, and that this reduction in parametric uncertainty translates into significant reductions in the predictive uncertainty associated with the reproduction of the latent, sensible and ground heat fluxes.

Assessing the constraint of the predicted aerodynamic surface temperatures

The results presented above indicate that conditioning SVAT models on additional objectives leads to some improvement in the prediction uncertainty associated with feasible model parameterisations. Additional sources of data should therefore yield increasingly constrained feasible parameter sets, through the rejection of further parameter sets (in terms of reproducing the additional observations).

Additional sources of data that might prove useful in this respect include surface temperature measurements. Figure 9 shows the prediction bounds of the surface temperature derived by multi-objective conditioning for a section of the evaluation data set. As can be seen, the resultant uncertainty envelopes are typically approximately 5 degrees wide, with one diurnal peak uncertainty at approximately 10 degrees. This figure therefore indicates that if appropriate measurements can be achieved with an accuracy of better than 5 degrees, then further constraint of the parameter space can be achieved and hence possible reductions in the predictive uncertainty associated with the model.

However, it must be remembered that additional complications may be introduced because thermal infra-red thermometry measures the radiative surface temperature at the effective scale of a particular instrument, whereas the model predicted variable is the effective (patch scale) aerodynamic temperature. A conceptual SVAT model effective temperature conditional on a particular 'big leaf' model structure and a conceptual radiative temperature after processing with the interpretative model of the digital numbers in the imager, are different quantities. This issue of commensurability means that we must expect the information content of the measured variable to contain inherent uncertainty.

One method of approaching such data to extract pertinent information is to use a 'temperature change' approach. The change of radiative temperature over a fixed period may be compared to model predicted changes in the aerodynamic surface temperature over the same fixed period, due to the approximately constant relationship between radiometric and aerodynamic temperatures (as observed by Huband and Monteith, 1985).

The actual conversion of radiometric to aerodynamic surface temperature requires the specification of the surface emissivity. Humes *et al.* (1994) have shown that for low emissivities (e.g. $\varepsilon = 0.96$), the required correction is

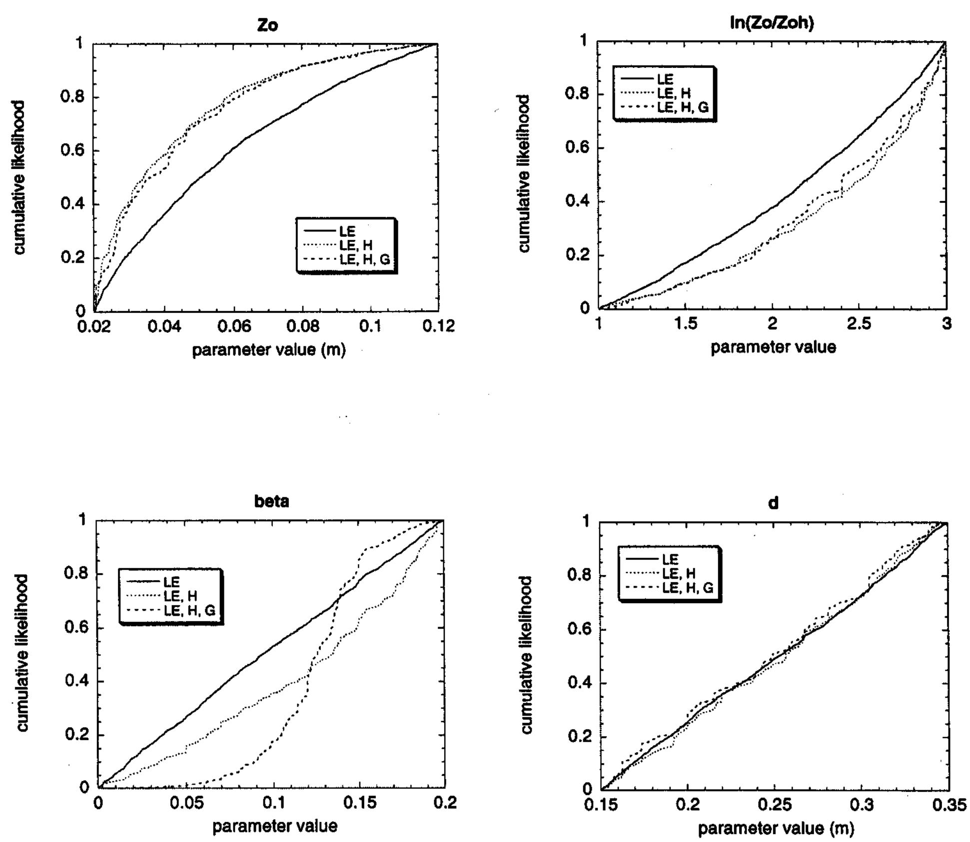

Fig. 5. *As figure 4, but for four model parameters relating to aerodynamic characteristics and the ground heat flux parameter (β).*

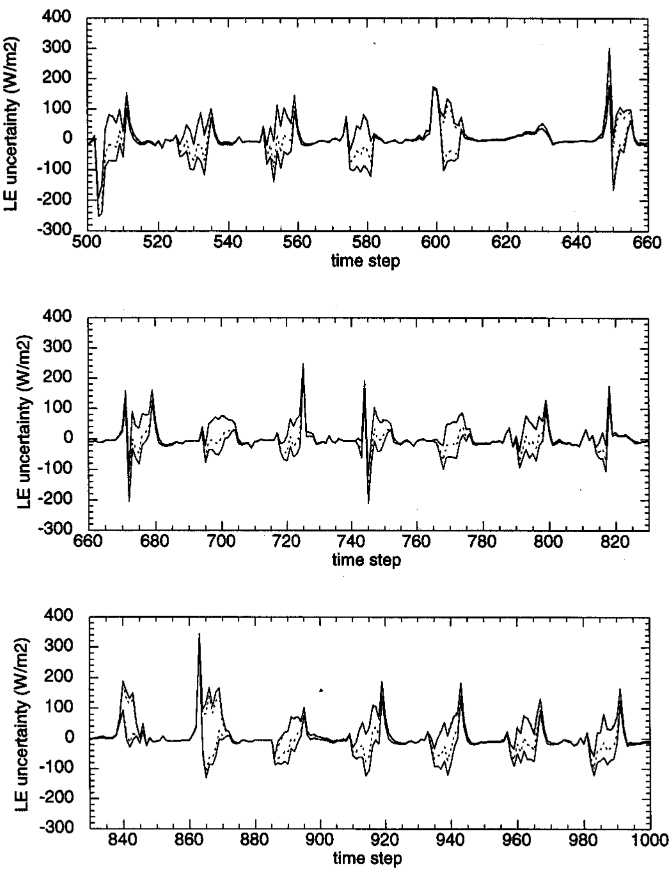

Fig. 6. *Comparison of uncertainty bound residuals for latent heat flux following multi-objective conditioning. The solid line represents single objective conditioning, whilst dashes represent multi-objective conditioning.*

1K at a radiometric temperature of 15 degrees celsius, increasing to 2K at approximately 50 degrees celsius, though this also depends to a lesser extent upon longwave radiation. The required correction is lower for higher emissivities (e.g. $\varepsilon > 0.96$), and less variable for the range of radiometric temperatures. Hence an approximately constant relationship may be assumed between radiometric and aerodynamic temperatures with limited uncertainty. The acknowledgement of intrinsic uncertainty in such data necessitates the employment of a Bayesian uncertainty framework such as GLUE to evaluate the potential worth of such data sources. Given the wide range of uncertainty associated with the posterior multiple-objective surface temperatures, shown in Fig. 9, it can be expected that measured temperatures will provide some additional information to constrain the feasible model space further.

Discussion

The results presented here indicate that conditioning on the measured latent fluxes alone can yield reasonable simulations of that variable when the retained models are used to simulate an additional evaluation period. It has been shown that multiple parameter sets can reproduce adequately the calibration and evaluation data sets used in this study.

It has also been found that the additional requirement to simulate the sensible heat flux produces significant additional constraint of some of the parameters to which model output is sensitive (Franks *et al.*, 1997b). Parameters SRMAX, RSMAX, z_o, and $\ln(z_o/z_{oh})$ are all significantly constrained when the observed sensible heat fluxes are used to condition the model. The results also demonstrate that ground heat flux parameter is not sufficiently constrained through calibration to the latent heat flux and/or both dominant heat fluxes. As indicated by Fig. 5, conditioning on the latent or both latent and sensible heat fluxes does not identify, adequately, suitably robust parameter values for the ground heat flux component. By conditioning the model on the latent, sensible and ground heat fluxes, it was found that a markedly constrained range of the ground heat flux parameter was achieved.

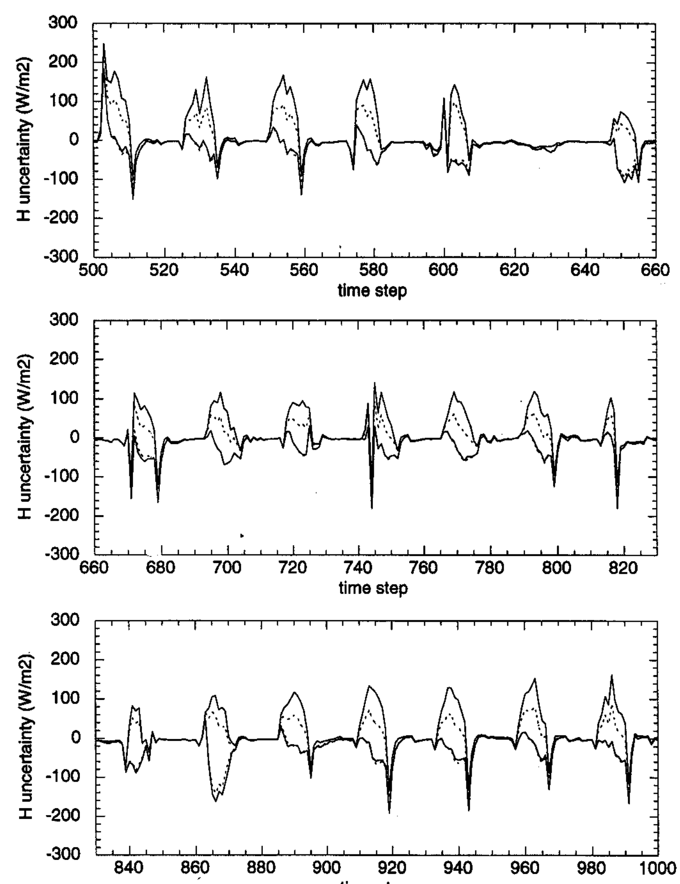

Fig. 7. *Comparison of uncertainty bound residuals for the sensible heat flux following single and multi-objective conditioning. The solid line represents single objective conditioning, whilst dashes represent multi-objective conditioning.*

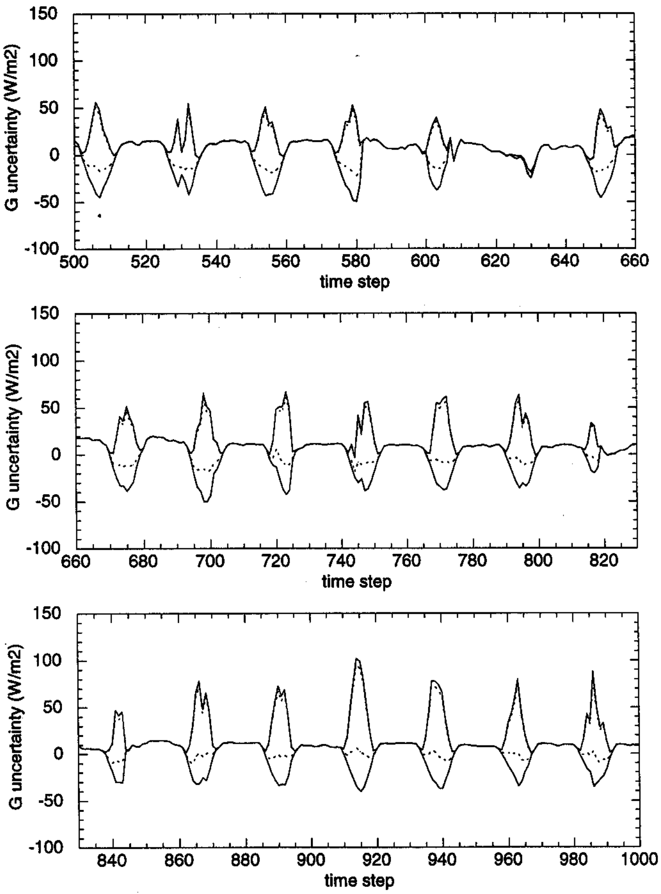

Fig. 8. *Comparison of uncertainty bound residuals for the ground heat flux following single and multi-objective conditioning. The solid line represents single objective conditioning, whilst dashes represent multi-objective conditioning.*

The utility of multi-objective conditioning has therefore been demonstrated through demonstrably constrained parameter ranges representing a significant reduction in the parametric uncertainty associated with the application of the TOPUP SVAT model. More significantly, however, this reduction in the parametric uncertainty has been shown to lead to a significant reduction in the uncertainty for all of the predicted fluxes.

These results have a number of implications for the application and calibration of SVAT models; they indicate that such SVAT models are over-parameterised with respect to the data typically available for their calibration. However, the additional constraint provided by the conditioning to the sensible heat flux results in a more constrained acceptable parameter space, with a consequent reduction of the predictive uncertainty associated with the reproduction of the heat fluxes. It therefore follows that calibration of a SVAT model can be more robustly achieved through such multi-objective conditioning. This is particularly important for applications of SVAT as a lower boundary for atmospheric modelling, as SVAT models will simulate more accurately all aspects of the energy balance if the model is conditioned on independently measured heat fluxes (though note that the problem of the uncertainty associated with the spatial heterogeneity of the energy budget components is not addressed here, but see Franks and Beven, 1997b, 1999).

However, it remains unclear how such a SVAT model may be applied at the large scales required for atmospheric modelling. As has been shown, SVAT models cannot be calibrated uniquely at the local patch (or plot) scale where a single parameterisation is appropriate. Consideration of the variability of pertinent land surface—atmosphere parameters and land surface fluxes at the typical resolution scales of atmospheric models indicates that such large-scale calibration/validation exercises are not currently feasible. Large-scale measurements of fluxes such as are achievable from airborne or satellite platforms may be representative over larger areas but then only provide a few measurements in time (Franks and Beven, 1999). Scintillometry offers some potential to provide continuous, large-scale measures of fluxes; however this technique is still under development (de Bruin *et al.*, 1995; Lagouarde *et al.*, 1997). The utility of evaluating SVAT models at local scales with regard to their large-

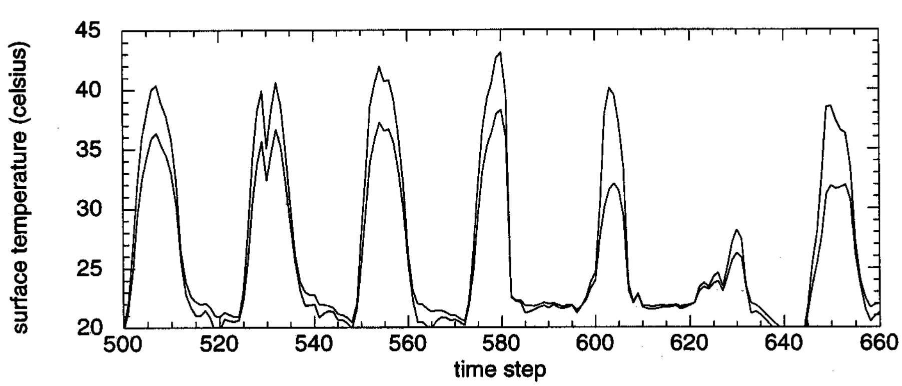

Fig. 9. *Prediction bounds propagated for the modelled aerodynamic surface temperature following multi-objective conditioning.*

scale application in such atmospheric models must therefore be questionable.

In all cases, it is clear that the complexity of current SVAT models inhibits their robust calibration and that significant uncertainty must be associated with their predictions. SVAT models need to be simplified to a level that is commensurate with the data available for their calibration. Ways must be sought to simplify the complexity of model structures and these models must be tested with additional sources of data to constrain their parameterisations.

A methodology to test the utility of such data through multi-objective conditioning has been presented. This methodology tests the achieved conditioning through the evaluation of the consequent parameter and predictive uncertainty, through the use of the Bayesian GLUE uncertainty framework. It may be inferred that the equifinality of parameter sets demonstrated here will also apply to multiple model structures (Beven, 1999). The GLUE approach can be extended easily to multiple SVAT model structures subject only to the limitation that each model simulates the variables that are compared with the available observations.

Acknowledgements

This study was funded under The University of Newcastle, Australia, Research Management Committee New Staff Grant 'Inter-comparison of SVAT models within a multi-objective uncertainty framework', December 1997. ABRACOS was an Anglo-Brazilian collaboration between the United Kingdom Overseas Development Agency and the Agência Brasileira de Cooperação. This work benefitted substantially from many discussions with George Kuczera and Jetse Kalma, University of Newcastle. Studies such as this are not possible without the intensive effort involved in the careful collection of the field observations under difficult conditions. Sincere thanks are due to all involved in the collection of these data sets, particularly staff from the Institute of Hydrology, the Brazilian Instituto Nacional de Pesquisas da Amazonia, and the Instituto Nacional de Pesquisas Espaciais.

References

Beven, K.J., 1989. Changing ideas in hydrology: the case of physically-based models. *J. Hydrol.*, **105**, 157–172.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, **16**, 41–51.

Beven, K.J., 1999. On uniqueness of place and process representations in hydrology. *Hydrol. Earth System Sci.*, in press.

Beven, K.J. and Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Processes*, **6**, 279–298.

Beven, K.J. and Quinn, P.F., 1994. *Similarity and scale effects in the water balance of heterogeneous areas*, Proc. AGMET conference on The Balance of Water—Present and Future, AGMET, Dublin. September 1994.

Beven, K.J., Lamb, R., Quinn, P.F., Romanowicz, R. and Freer, J., 1995. TOPMODEL, in Singh, V.P., (ed.), *Computer models of watershed hydrology*, Water Resource Publications, Highlands Ranch, Co.

Box, G.E.P. and Tiao, G.C., 1973. *Bayesian inference in statistical analysis*, Addison-Wesley, Reading: Massachusetts.

Calder, I.R., 1990. *Evaporation in the uplands*, Wiley, Chichester, UK, pp.148.

Cameron, D., Beven, K.J., Tawn, J., Blazkova, S. and Naden, P., Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty), *J. Hydrol.*, **219**, 169–187

Chappell, N.A., Franks, S.W., and Larenus, J. 1998. Multi-scale permeability estimation in a tropical catchment. *Hydrol. Processes*, **13**, 1507–1523.

Clothier, B.E., Clawson, K.L., Pinter, P.J., Moran, M.S., Reginato, R.J. and Jackson, R.D., 1986. Estimation of soil heat flux from net radiation during the growth of alfalfa, *Agric. For. Meteorol.*, **37**, 319–329.

de Bruin, H.A.R., van den Hurk, B.J.J.M. and Kohsiek, W., 1995. The scintillation method tested over a dry vineyard area, *Boundary Layer Meteorol.*, **76**, 25–40.

Duan, Q., Sorooshian, S. and Gupta, V., 1992. Effective and efficient global optimisation for conceptual rainfall-runoff models, *Wat. Resour. Res.*, **28**, 1015–1031.

Franks, S.W. and Beven, K.J., 1997a. Bayesian estimation of uncertainty in land surface—atmosphere flux predictions, *J. Geophys. Res.*, **D20**, 23,991–23,999.

Franks, S. and Beven, K.J., 1997b, Estimation of evapotranspiration at the landscape scale: a fuzzy disaggregation approach, *Wat. Resour. Res.*, **33**, 2929–2938.

Franks, S.W. and Beven, K.J., 1999. Conditioning a multiple patch SVAT model using uncertain time-space estimates of latent heat fluxes as inferred from remotely-sensed data, *Water Resour. Res.*, **35**, 2751–2761.

Franks, S.W., Beven, K.J., Chappell, N.A. and Gineste, P., 1997a. The utility of multi-objective conditioning of a distributed hydrological model using uncertain estimates of saturated areas. In A. D. McDonald, and M. McAleer (eds.), *Proc. Int. Congress on Modelling and Simulation (MODSIM '97)*, Hobart, 8–11 December 1997, Vol. 1, 335–340.

Franks, S.W., Beven, K.J., Quinn, P.F. and Wright, I.R., 1997b. On the sensitivity of Soil Vegetation—Atmosphere Transfer (SVAT) schemes: equifinality and the problem of robust calibration, *Agric. For. Meteorol.*, **86**, 63–75.

Franks, S.W., Gineste, P., Beven, K.J. and Merot, P., 1998. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process, *Wat. Resour. Res.*, **34**, 787–797.

Freer, J., Ambroise, B. and Beven, K.J., 1996. Bayesian Estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Wat. Resour. Res.*, **32**, 2161–2173.

Gash, J.H.C., Nobre, C.A., Roberts, J.M. and Victoria, R.L., 1996. An overview of the ABRACOS project, in J.H.C. Gash, C.A. Nobre, J.M. Roberts, and Victoria, R.L., (eds.), *Amazonian Deforestation and Climate*, Wiley, Chichester, UK, pp 1–14.

Gash, J.H.C. and Nobre, C.A., 1997. Climatic effects of Amazonian deforestation: some results from ABRACOS, *Bull. Am. Meteorol. Soc.*, **78**, 823–830.

Gupta, H.V., Sorooshian, S. and Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and non-commensurable measures of information, *Wat. Resour. Res.*, **4**, 751–763.

Huband, N.D.S. and Monteith, J.L., 1986. Radiative surface temperature and energy balance of a wheat canopy. 1. Comparison of radiative and aerodynamic canopy temperature, *Boundary Layer Meteorol.*, **36**, 1–17.

Humes, K.S., Kustas, W.P., Moran, M.S., Nichols, W.D. and Weltz, M.A., 1994. Variability of emissivity and surface temperature over a sparsely vegetated surface, *Wat. Resour. Res.*, **30**, 1299–1310.

Klepper, O., 1997. Multivariate aspects of model uncertainty analysis: Tools for sensitivity analysis and calibration, *Ecol. Modelling*, **101**, 1–13.

Kuczera, G., 1983. Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty, *Wat. Resour. Res.*, **19**, 1151–1162.

Kuczera, G. and Mroczkowski, M., 1998. Assessment of hydrologic parameter uncertainty and the worth of multi-response data, *Wat. Resour. Res.*, **34**, 1481–1489

Kustas, W.P. and Daughtry, C.S.T., 1987. Estimation of soil heat flux/net radiation ratio from spectral data, *Agric. For. Meteorol.*, **49**, 205–233.

Lagouarde, J-P., McAneney, K.J. and Green, A.E., 1996. Scintillometer measurements of sensible heat flux over heterogeneous surfaces, In '*Scaling up in hydrology using remote sensing*', J.B. Stewart, E.T. Engman, R.A. Feddes, and Y. Kerr (eds.), Wiley, Chichester, UK.

Lamb, R., Bevan, K. and Myrabø, S., 1998. Use of spatially-distributed water table observations to constrain uncertainty in a rainfall-runoff model, *Adv. Wat. Resour.* **22**, 305–317.

Monteith, J.L., 1981. Evaporation and surface temperature, *Quart. J. Roy. Meteorol. Soc.*, **107**, 1–27.

Mroczkowski, M., Raper G.P. and Kuczera, G., 1997. The quest for more powerful validation of conceptual catchment models, *Wat. Resour. Res.*, 33, 2325–2335.

Romanowicz, R., Beven, K.J. and Tawn, J. 1996. Bayesian calibration of flood inundation models, in M.G. Anderson and D.E. Walling and P.D. Bates (eds.) *Floodplain Processes*, Wiley Chichester, UK.

Schulz, K., Beven, K. and Huwe, B., 1999. Equifinality and the problem of robust calibration in nitrogen budget simulations, *Soil Sci. Soc. Amer. J.*, in press.

Shuttleworth, W.J., Gash, J.H.C., Roberts, J.M., Nobre, C.A., Molion, L.C.B. and Ribeiro, M.N.G., 1991. Post-deforestation Amazonian climate: Anglo-Brazilian research to improve prediction, *J. Hydrol.*, **129**, 71–85.

Spear, R.C. and Hornberger, G.M., 1980. Eutrophication in Peel Inlet II. Identification of critical uncertainties via Generalised Sensitivity Analysis, *Wat. Res.*, **14**: 43–49.

Wright I.R., Gash, J.H.C., da Rocha, H.R., Shuttleworth, W.J., Nobre, C.A., Maitelli, G.T., Zamparoni, C.A.G.P. and Carvalho, P.R.A., 1992. Dry season micrometeorology of central Amazonian ranchland. *Quart. J. Roy. Meteorol. Soc.*, **118**: 1083–1099.

Wright, I.R., Manzi, A.O. and da Rocha, H.R., 1995. Surface conductance of Amazonian pasture: model application and calibration for canopy climate, *Agric. For. Meteorol.*, **75**, 51–70.

Yapo, P.O., Gupta, H.V. and Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models, *J. Hydrol.*, **204**, 83–97.

Zak, S., Beven, K.J. and Reynolds, B., 1997. Uncertainty in the estimation of critical loads: a practical methodology *Soil, Wat. Air Pollut.*, **98**, 297–316.

Zak, S. and Beven, K.J., 1999. Equifinality, sensitivity and uncertainty in the estimation of critical loads, *Sci. Total Environ.*, **236**, 191–214.