Chapter C.4

Terrestrial Data Assimilation

Paul Houser · Michael F. Hutchinson · Pedro Viterbo · Hervé Douville · Steven W. Running

Accurate assessment of the spatial and temporal variation of terrestrial system storages (energy and mass) is essential for addressing a wide variety of highly socially relevant science, education, application, and management issues. Improved land-surface state estimates find direct application in agriculture, forest ecology, civil engineering, water resources management, crop system modelling, rainfall-runoff prediction, atmospheric process studies, and climate and ecosystem prediction. Data assimilation is a method by which observations and modelling are combined to create a continuous dataset in space and time, devoid of gaps. Pioneered for the atmosphere in operational Numerical Weather Prediction (NWP) centres (the National Centers for Environmental Prediction, Kalnay et al. 1996; the European Centre for Medium-Range Weather Forecasts, Gibson et al. 1994; the NASA Data Assimilation Office, Schubert et al. 1993), the method is now being applied at the land surface. Spatially and temporally variable rainfall and available energy, combined with land-surface heterogeneity, cause complex variations in all processes related to surface hydrology between the scales of conventional measurement networks. The characterisation of the spatial and temporal variability of water and energy cycles are critical to improve our understanding of land-surface/atmosphere interaction and the impact of land-surface processes on climate extremes. Because the accurate knowledge of these processes and their variability is important for weather and climate predictions, most NWP centres have incorporated land-surface schemes in their models. However, errors in the NWP forcing accumulate in the surface water and energy stores, leading to incorrect surface water and energy partitioning and related processes. This has motivated the NWP centres to impose *ad hoc* corrections to the land-surface states to prevent this drift.

Land data assimilation entails the use of uncoupled land-surface models forced with near-surface meteorological observations, and is therefore not affected by NWP forcing biases. In practice, gaps in the meteorological observing network mean that land models are forced with output from atmospheric analyses produced with their own data assimilation techniques for gap-filling, as well as satellite data and radar precipitation measurements. Existing high-resolution vegetation and soil coverage data can be used to specify land-model parameters. By virtue of the use of a gridded regional or global land-surface model in the assimilation process, co-registration of all output data is achieved automatically. The land model, run at high resolution, produces results that can be aggregated to various scales to assess water and energy balances and validated with various *in situ* observations. Ultimately, observations of land-surface storages (soil moisture, temperature, snow) and fluxes (evaporation, sensible heat flux, runoff) can be used to further validate and constrain the land data assimilation predictions. By continuously confronting theoretical and observational knowledge, data assimilation presents a rich opportunity to better understand physical processes and observation quality in a structured, iterative and open-ended learning process.

Data assimilation is also an important tool to help us make sense of voluminous and disparate data types that are becoming available from new space-based Earth observation platforms, as well as traditional *in situ* observations. Inconsistencies between observations and predictions are easily identified in a data assimilation system, and demand explanation, providing a basis for observational quality control and validation. Finally, the data assimilation system can extend or "advect" the available observation information in time and space (e.g. surface soil moisture observations vertically into the root zone) to provide continuous fields for use in subsequent research and application. Essentially, data assimilation is used to consolidate disparate observational and model information operationally into a unified, complete description of the terrestrial system that can be used community-wide by scientists to study important phenomena, evaluate models and data, and enhance prediction.

As with oceanic or atmospheric data assimilation, there exist both statistical and dynamic approaches to fill the gaps between observations in space and time. Statistical methods apply assumed or estimated properties of a given variable to derive continuous fields or improve resolution. These may be as simple as linear interpolation or regression, or may employ sophisticated techniques such as empirical consideration of the physical terrain or use of neural networks. Stochastic techniques also may be used to generate "weather" where only time-averaged data exist, thereby downscaling in time and producing more realistic forcing for individual modelling efforts.

Dynamic data assimilation employs physically-based models to fill gaps between observations, and produces a physically-consistent estimate of the space-time evolution (Fig. C.7). For land data assimilation, either standalone land-surface schemes or coupled land-atmosphere models may be used. However, both statistical and dynamic land data assimilation represent only an approximation to the true evolution of the physical state – the more ground truth data that can be brought to bear on the problem, the better the result.

When compared to its atmospheric counterpart, terrestrial data assimilation is in its infancy (Mahfouf and Viterbo 1998). This is due to a lack of an operational exchange of observed data, the small-scale structure of many land-surface variables, and the relatively crude specification of the surface in the atmospheric host models used by the operational centres. Moreover, the operational data assimilation centres, mainly supporting weather forecasts, were slow to recognise the importance of the land surface. Finally, surface observables are re-

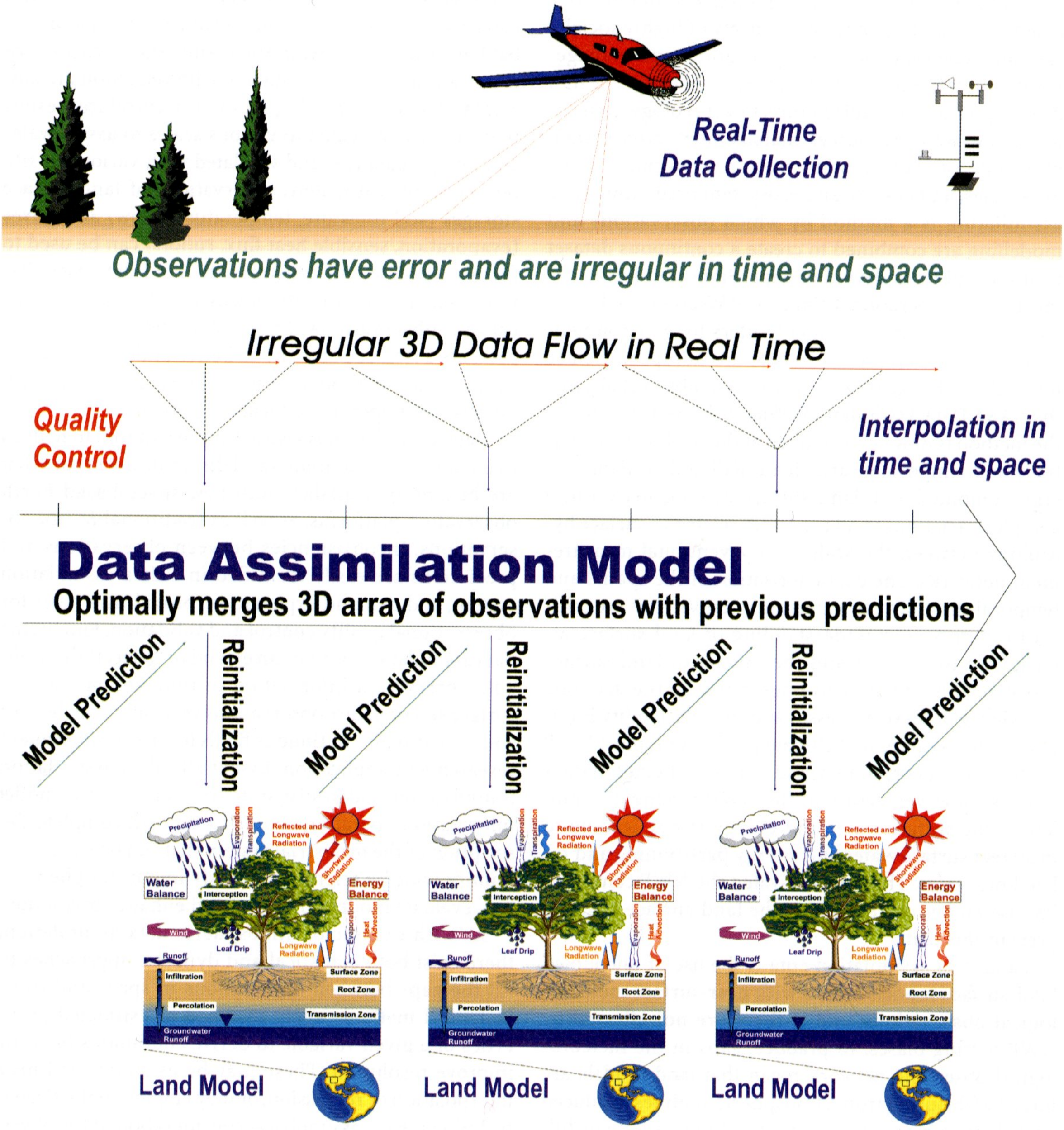


Fig. C.7. The land-surface data assimilation process lated in a non-linear way to the model state variables and the structure of model and observations errors is poorly known.

C.4.1 Topographic Coherence of Weather – The Role of Statistical Assimilation

C.4.1.1 Stochastic Weather Models for Scenario Generation

A major requirement of many IGBP Programmes, including BAHC, GCTE, LUCC and GAIM, is the development of techniques for the simulation of spatially and temporally detailed atmospheric inputs to hydrological and ecological models. These are required at the spatial and temporal scales needed by real management systems. The spatial scales asked for are normally not coarser than a few kilometres. The temporal scale usually required is at least the daily time scale, incorporating mean behaviour as well as measures of variation and extremes. The spatial scale is much finer than the spatial resolution of general circulation and regional atmospheric models. A commonly accepted way of generating fine-scale scenarios is to perturb parameters of stochastic models of the observed current weather, for example, in accordance with broad-scale simulations of general circulation models (IGBP 1993). This approach has been used widely in climate impact research, both for the current weather and for projected future climates (Zorita et al. 1992; Kittel et al. 1995; Mearns et al. 1997; Semenov and Barrow 1997; Wilby et al. 1998). It has been called semi-empirical downscaling by Giorgi and Mearns (1991) and rests on three principal assumptions:

1. The validity of the broad-scale scenarios generated by GCMs;
2. The maintenance in changed climates of the observed links between the broad-scale atmospheric behaviour and stochastic weather model parameters;
3. The ability of stochastic weather models, with parameters essentially consisting of first- and second-order summary atmospheric statistics, to simulate accurately spatially and temporally detailed atmospheric inputs to ecological and hydrological models.

These assumptions are ordered by their degree of validity. Assumption 1 is most open to question, particularly with regard to significant interactions between the atmosphere, the land-surface and the ocean system and therefore a very significant source of uncertainty in scenario generation. This is underlined by significant differences in the broad scale precipitation scenarios provided by different GCMs (CSIRO 1996). Nevertheless, GCMs provide a starting point for one type of generation of a range of climate change scenarios. Given the uncertainties associated with GCM predictions it is now recognised that broad scale scenarios should also be generated in the light of perceived vulnerabilities of land-surface systems. This is illustrated below in Fig. C.8 and further discussed in Part E.

Assumption 2 is also open to question but is supported by the observed broad scale temporal and spatial coherence of the current atmosphere system. This coherence appears to be founded on physical principles, and is therefore likely to be maintained in changed climates. The validity of the links identified under assumption 2 will be enhanced if the stochastic weather models are simply parameterised, so that they can be calibrated robustly from minimal data. This can facilitate the identification of the parameters most likely to respond to climate change, especially if the stochastic models incorporate physically-based structures that can be identified from observed data (Hutchinson 1995a).

The validity of assumption 3 is generally accepted. This has formed the basis for the development of stochastic weather models to provide inputs to ecological and hydrological models, beginning with the work of Jones et al. (1972), Richardson (1981) and Srikanthan and McMahon (1984), through to more recent work by Racsko et al. (1991), Shah et al. (1996) and Wilks (1999b). First- and second-order long-term weather statistics, from which stochastic simulation is a direct corollary, can well calibrate atmospheric variability, including probabilities of extreme events for the current weather. The main ongoing issue here is the practical one of identifying and calibrating space-time stochastic models that incorporate adequately observed spatial and temporal behaviour. This is made difficult by the relative sparseness of measured surface weather data and the need to calibrate observed longer term weather variations, over decades or more.

There has been steady progress in the development of methods for observing and interpreting spatially detailed atmospheric inputs by remote sensing and consequent potential for integrating remotely-sensed data with ground-based data (Georgakakos and Kavvas 1987; Stewart and Finch 1993; Fo and Crawford 1999). However, there are ongoing difficulties in calibrating remotely-sensed surface weather data accurately, particularly rainfall data (O'Connell and Todini 1996). Radar estimates of rainfall amounts can be in error by as much as a factor of two (Barros and Kuligowski 1998). As recognised in IGBP (1993), the main role for remotely-sensed data in the context of stochastic weather model development appears to be in providing insight into the nature of the spatial variability of surface weather. The ground-based meteorological data network remains the only source of daily surface weather data with extensive temporal coverage over the 20th century and near complete global land coverage. It is important that this network be maintained and, where appropriate, upgraded.

C.4.1.2 Scheme for Generating Stochastic Weather Scenarios

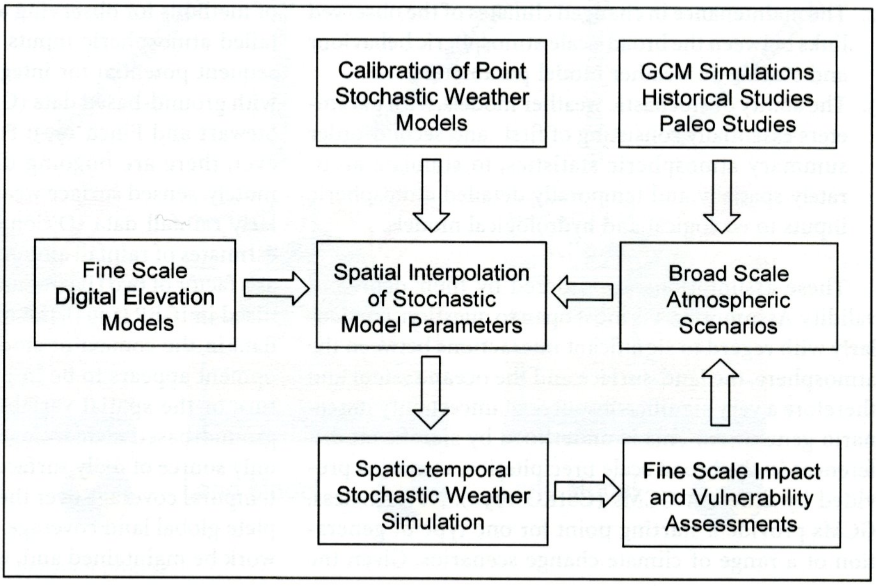
If stochastic models with global coverage are to be developed from the ground-based meteorological data network, there is a critical need for accurate spatial interpolation of appropriate weather statistics across the Earth's land surface. This has been achieved by incorporating topographic dependencies into the interpolation process. The task of developing stochastic space-time weather models from standard meteorological networks can then be conveniently divided into three steps (Hutchinson 1995a), as shown in Fig. C.8. This approach conveniently isolates different aspects of weather model development. In particular, it helps to identify the differing spatial dependencies and spatial scales of various model parameters and weather anomalies, with consequent differing requirements with regard to model complexity and observation network density.

The first step in this scheme is the development of stochastic models at single points where recorded weather data are available. It is at this stage that questions of appropriate temporal scale for ecological and hydrological models are addressed, as well as the incorporation of simple physically-based model structures, as discussed above. Both daily and monthly time scales have been used, although the daily time step is the most common. It is sufficient to model rainfall extremes, soil erosion and time-critical temperature-dependent events in plant growth and yield. The monthly time step is sufficient to model drought, natural vegetation, and broad scale hydrology. The variables most commonly required for ecological and hydrological models are precipitation, daily maximum and minimum temperature, solar radiation, atmospheric humidity and potential evaporation (IGBP 1993).

The second step in Fig. C.8 is the development of techniques to extend the parameters of point stochastic models spatially across the landscape. This is usually done by incorporating the effects of fine scale topography, as provided by the GTOPO30 and other elevation models. It is after parameters have been interpolated across the landscape that broad scale long-term weather change perturbations are applied. As both daily and monthly point stochastic weather models are normally calibrated on a month-by-month basis (Richardson 1981; Georgakakos and Kavvas 1987), the spatial interpolation of model parameters is related strongly to the interpolation of summary monthly weather statistics.

The most critical parameters in stochastic weather model calibration are the means of the different variables. Measures of variance and serial dependence are often less well defined by the data, and less critical to the performance of the fitted weather model. Variances are normally required to calibrate probabilities of extreme events, but Hutchinson (1995a), building on the work of Stidd (1973), has shown that daily rainfall distributions can be calibrated accurately with a truncated normal distribution using essentially just two first-order statistics, the mean rainfall amount and the mean number of dry days. Similarly, Chia and Hutchinson (1991) have shown that the variance of daily sunshine duration can be estimated reasonably from the daily mean and Geng (1986) has demonstrated empirical relationships between mean and variance parameters of the Richardson (1981) weather model. Mean weather parameters typically display the most complex spatial patterns, often strongly modulated by topography. Thus effective methods for the spatial interpolation of monthly mean weather statistics should be sufficient to enable the interpolation of all of the parameters of suitably parameterised stochastic point models.

Fig. C.8. Scheme for fine scale stochastic space-time simulation of weather from ground-based data and broad scale atmospheric scenarios for climate impact and vulnerability assessments



Developments in the first two steps in Fig. C.8 are guided by the requirements of the third step, the development of coordinated space-time models that adequately respect both spatial and temporal dependencies of daily or monthly weather anomalies. These anomalies are essentially normalised differences between actual daily (or monthly) weather values and the monthly mean values. Normalised anomalies show a high degree of spatial and temporal coherence and can be modelled effectively by a first-order multivariate autoregressive model. Richardson (1981) showed that such a model was appropriate for normalised daily temperature and solar radiation anomalies, and that the parameters defining the first-order temporal correlation structure could be assumed to be constant across all sites examined. Similarly, Schneider and Griffies (1999) found that a first-order autoregressive model was sufficient to model GCM output and could be used as a basis for predictability studies of area-averaged variables. Autoregressive structures can also be simply extended to account for interannual variability, such as in the twofold autoregressive Markov model described by Karner and Rannik (1996). More complex autoregressive integrated moving average models have also been used to model space-time precipitation (Shah et al. 1996).

C.4.1.3 Stochastic Models

Stochastic weather model development has been almost totally dominated by developments in stochastic models of precipitation, for which there is a vast literature (Georgakakos and Kavvas 1987). This reflects both the dominant control exerted by precipitation on ecological and hydrological processes and the difficulty in developing complete weather models. Only relatively simple daily and monthly precipitation models, however, can be used for long-term weather-scenario applications. More complex precipitation models have been developed for restricted types of precipitation and have complex structures that are acknowledged to be difficult to calibrate and validate over large areas (Valdes et al. 1985; Sivapalan and Wood 1987). However, there has been progress in simplifying the parameters of rainfall models based on cluster processes and in regionalising these parameters for environmental impact studies (Cowpertwait et al. 1996).

Two classes of daily stochastic precipitation models have achieved common use in scenario generation. The traditional approach uses a first-order two-state Markov chain to model rainfall occurrence and models rainfall amounts on wet days independently using a gamma distribution. This approach has been extended by fitting more complex models, but the basic separation between rainfall occurrence and rainfall amount has remained (Wilks 1999a). The occurrence structure has been extended by fitting different distributions to lengths of wet and dry spells and by extending the order of the Markov chain. Models for rainfall amount have been extended to include the mixed exponential distribution and by conditioning the parameters of the gamma distribution on the wet or dry status of preceding days. These extensions are obtained at the expense of fitting additional parameters. A simple occurrence model based on the normal distribution was suggested by Hutchinson (1995a). This model uses no additional parameters but was found to simulate dry spell lengths better than a first-order Markov chain model at locations across the whole USA.

The second approach consists of conditional stochastic precipitation models that incorporate physically-based controls by having model parameters conditioned on classified broad scale atmospheric circulation patterns. Such models have been developed by Bardossy and Plate (1992), Hay et al. (1991) and Wilson et al. (1992). These models have been found to match various observed mean and extreme rainfall statistics and to respect observed spatial dependencies. These models offer fairly direct links to large-scale circulation patterns. Their utility in generating future weather scenarios, with respect to Assumption 2 above, depends on the maintenance of the observed classes of atmospheric circulation patterns, and the constancy of their links with the rainfall model parameters. These models have provided valuable insight into the non-linear sensitivity of hydrological processes to postulated changes in global weather.

Complete stochastic weather models, such as the point model proposed by Racsko et al. (1991) and the multi-site model proposed by Wilks (1999b), have typically been constructed along the lines of the model proposed by Richardson (1981). These models have been used widely in scenario applications. In these models, weather variables, including daily maximum and minimum temperature and daily total solar radiation, are conditioned on the daily occurrence and non-occurrence of precipitation. However, this conditioning is weakly defined, since the *process* that affects temperature and solar radiation directly is the occurrence of significant cloud. This is not always associated with precipitation. More direct process-based relationships between the commonly required surface weather variables need to be explored. Such developments may overcome some of the reservations expressed by Katz (1996) about the use of such conditional stochastic weather models in generating long-term weather change and variability scenarios.

These models also need to be extended to better account for observed interannual variability. As widely recognised, and discussed by Katz and Parlange (1998), stochastic weather models fitted to daily statistics tend to underestimate interannual variability. This can be partially addressed by fitting more complex models, but these models tend to vary with geographic location and this can make the construction of multi-site models difficult Wilks (1999a). It is also likely that this phenomenon is due in part to non-stationarities in the weather system, such as those associated with ENSO (Phillips et al. 1998). More generic methods are called for to address this issue.

C.4.1.4 Topographic Dependent Interpolation of Weather Model Parameters

Statistical interpolation techniques appear to be best suited to the task of spatially extending the parameters of point simulation models. The techniques include kriging (Cressie 1991) and thin plate smoothing splines (Wahba and Wendelberger 1980). These methods have similar accuracy although splines tend to be more easily calibrated (Hutchinson and Gessler 1994). Thin plate smoothing splines have been used to interpolate monthly mean weather parameters across the Australian continent (Hutchinson 1991), England (Semenov and Brooks 1999) and Canada (Price et al. 2000), at spatial resolutions of a few kilometres. The PRISM method (Daly et al. 1994) fits local elevation-based regressions to weather data. It has been used to interpolate long-term weather statistics across the USA, also at a spatial resolution of a few kilometres. Thin plate smoothing splines have also been used to interpolate weather means at coarser resolution across Europe (Hulme et al. 1995) and all continents except Antarctica (Leemans and Cramer 1991).

The major factor in the accuracy and spatial resolution of these interpolated weather surfaces has been the incorporation of dependences on elevation as indicated in Fig. C.8. This is well known in the case of temperature, where the dependence on elevation is almost linear. Monthly mean precipitation is also modulated strongly by topography, but its influence varies spatially. Both thin plate smoothing splines and the PRISM method can incorporate this spatially varying dependence. Hutchinson (1995b) and Running and Thornton (1996) have shown that the relative impact of elevation on precipitation patterns is two orders of magnitude greater than the impact of horizontal position. Thus precipitation patterns can be influenced significantly by relatively modest topographic features (Barros and Kuligowski 1998). The spatial resolution of this dependence has been estimated as 4–10 km (Daly et al. 1994; Thornton et al. 1997; Hutchinson 1998).

The number of data points and approximate standard errors of monthly mean weather surfaces fitted across Australia are given in Table C.4. These errors are typical for elevation-dependent surfaces derived from standard meteorological networks. As for most regions of the world, the number of stations that record precipitation exceeds by about an order of magnitude the number of stations that record other weather variables. This gives a reasonable indication of the relative spatial complexity associated with each variable. An example of a final interpolated surface for precipitation is shown in Fig. C.9. Further investigation is needed to clarify the scales of various topographic effects on precipitation (Hutchinson 1998) and to better define the topographic effects on temperature, such as topographically controlled inversions in winter and proximity to large water bodies (Hutchinson 1991).

Table C.4. Number of data points and approximate standard errors of fitted monthly mean climate surfaces across Australia (Hutchinson 1991)

Climate variable	Number of data points	Standard error
Solar radiation	150	3%
Daily maximum temperature	900	0.2–0.4 °C
Daily minimum temperature	900	0.3–0.5 °C
Precipitation	10000	5–15%
Pan evaporation	300	5%

C.4.1.5 Spatial Structure and Topographic Dependencies of Weather Anomalies

Since hydrological responses can be required simultaneously across larger regions, the spatial covariance structure of the weather anomalies generating these responses needs to be accommodated. Two applications of this analysis may be distinguished. The first is the spatial interpolation of values in real time from observation networks. This has particular relevance for calibration and validation of models of observed hydroecological response. Monthly rainfall anomalies can be interpolated reasonably well from standard meteorological networks (Lyons 1990), but additional remotely-sensed data are usually required to interpolate daily rainfall satisfactorily (Fo and Crawford 1999).

The second application is the statistical simulation of spatially distributed weather anomalies for scenario applications. Normalised weather anomalies tend to display broad spatial patterns that are controlled by broad synoptic patterns. These patterns tend to be independent of topography, unlike the corresponding non-normalised quantities. This can simplify the task of analysing spatial covariance structure, which can be addressed using standard multivariate statistical interpolation and analysis methods (Bell 1987; Daley 1991). However, while relatively simple models can be used, rainfall anomaly structure can be complicated by anisotropy (Obled and Creutin 1986), by systematic differences between inten-

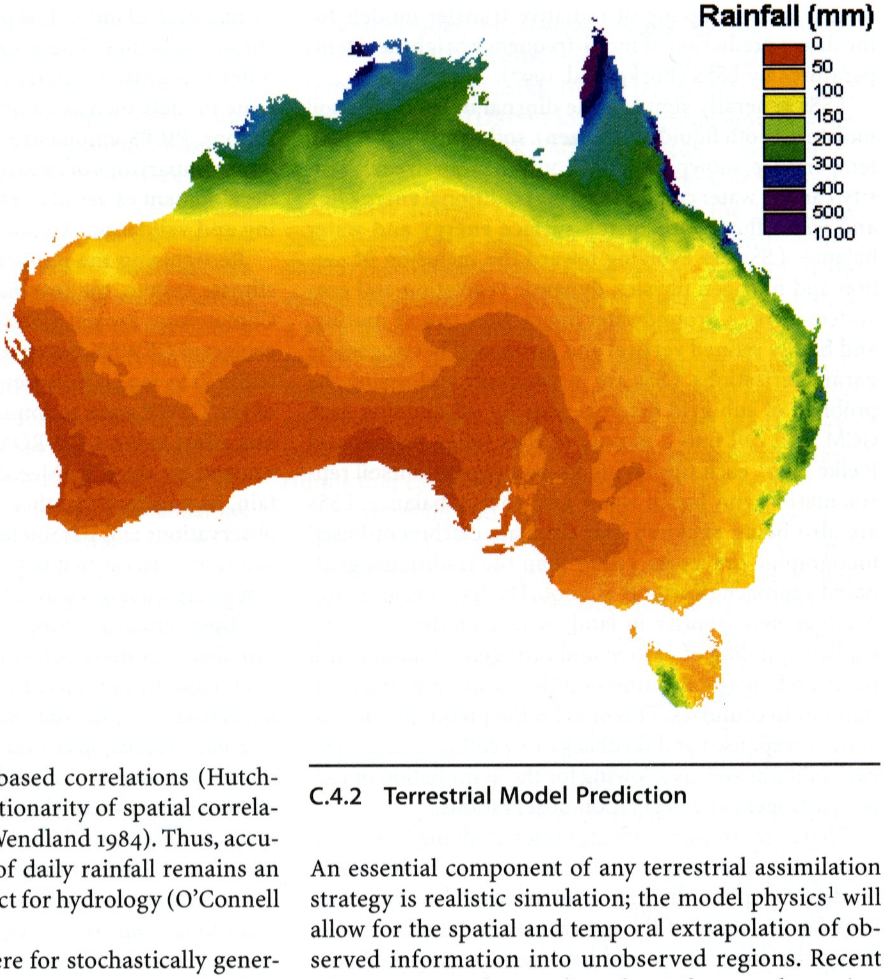


Fig. C.9.
January mean precipitation interpolated from 16 000 stations across Australia across a 2.5 km grid

sity-based and occurrence-based correlations (Hutchinson 1995a) and by non-stationarity of spatial correlations over time (Jones and Wendland 1984). Thus, accurate space-time simulation of daily rainfall remains an active and challenging subject for hydrology (O'Connell and Todini 1996).

The scheme described here for stochastically generating spatially and temporally detailed weather values from ground-based meteorological data has been widely adopted for environmental impact research. Its strength is its reliance on well-established methods for spatial and temporal analysis. These methods reflect a high degree of spatial and temporal coherence in the surface weather, particularly when dependences on topography are incorporated. The methods are subject to ongoing development and are critically dependent on the maintenance, and upgrade where appropriate, of the ground-based meteorological network. Equally important issues are the identification and calibration of simple, physically-based model structures, and whether these structures can be used with confidence in changed weather (Hutchinson 1995a). There is a similar search for parsimonious, physically-based model structures in hydrology (O'Connell and Todini 1996). That such stochastic model structures can be used in a changed weather can only be truly ascertained after such a change occurs. Nonetheless, these structures are well founded in existing methods for calibrating current and past weather, and are an appropriate assumption for scenario generation.

C.4.2 Terrestrial Model Prediction

An essential component of any terrestrial assimilation strategy is realistic simulation; the model physics[1] will allow for the spatial and temporal extrapolation of observed information into unobserved regions. Recent advances in understanding the soil water dynamics, plant physiology, micrometeorology and hydrology that control biosphere-atmosphere interactions have spurred the development of Land Surface Schemes (LSSs), whose aim is to represent simply yet realistically the transfer of mass, energy and momentum between a vegetated surface and the atmosphere (e.g. Dickinson et al. 1993; Sellers et al. 1986). LSS predictions are regular in time and space but these predictions are influenced by model structure, errors in input variables and model parameters, and inadequate treatment of sub-grid scale spatial variability. Several studies have shown that models with physically observable parameters and states are preferred in data assimilation studies (Entekhabi et al. 1994; Houser et al. 1998). Therefore, LSSs must be developed that replace the current conceptual parameters and states that are not observable with more physically-realistic representations. An important development in this

[1] Throughout this part, the term "physics" is used to represent all physical processes, including chemical and biological as well as thermodynamic, hydrodynamic, and other "diabatic" processes that produce or consume non-mechanical energy.

regard is the coupling of radiative transfer models for the direct prediction of multi-frequency brightness temperatures by LSSs (Burke et al. 1997).

LSSs generally simulate the diurnal dynamics of soil moisture (both liquid and frozen), soil temperature, skin temperature, snowpack water equivalent, snowpack density, canopy water content, and the traditional energy flux and water flux terms of the surface energy and water balance. LSSs are evolving toward the inclusion of carbon and nitrogen physics, dynamic vegetation and ecosystem physics, groundwater interaction, runoff routing, and highly refined vertical and horizontal heterogeneity parameterisations. LSSs are increasingly addressing the problem of sub-grid heterogeneity by subdividing each GCM grid cell into a mosaic of tiles (after Avissar and Pielke 1989), each tile having its own vegetation/soil representation and hence water and energy balance. LSSs are also being explored that simulate catchment-based topographic processes, rather than the traditional grid-based approach (Koster et al. 2000; Ducharne et al. 2000). Another new frontier in land-surface modelling is the explicit coupling of carbon and nitrogen dynamics with water and energy dynamics on timescales ranging from minutes to centuries. This enables the prediction of ecosystem responses and feedbacks to weather and climate variability, as well as allowing for the assimilation of carbon, nitrogen, and vegetation observations.

There are strong justifications for studying LSSs, both uncoupled and coupled with atmospheric and ocean models. Coupling the LSS to an atmospheric model allows for the study of the interaction and feedbacks between the atmosphere and land surface. However, coupled modelling also imposes strong land-surface forcing biases predicted by the atmospheric model on the LSS. These biases in precipitation and radiation can overwhelm the behaviour of LSS physics (Dirmeyer 2001). In fact, several weather prediction centres must "correctively nudge" their LSS soil moisture toward climatological values to offset its drift. An uncoupled LSS can use observed land-surface forcing, use less computational resources, and still address many relevant scientific questions. Arguably, the most critical terrestrial coupling with the atmosphere is through precipitation. There are currently large efforts to derive precipitation (e.g. Global Precipitation Climatology Project) and to assimilate precipitation estimates into coupled models (Hou et al. 2000), which may improve the forcing of land models in coupled systems significantly.

PILPS has been responsible for a series of complementary experiments that focuses on identifying parameterisation strengths and inadequacies in about 30 land-surface process models. PILPS is a project designed to improve the parameterisation of the continental surface, especially hydrological, energy, momentum and carbon exchanges with the atmosphere. This is an important exercise because there are significant differences in the formulation of individual processes in the available land-surface schemes. These differences are comparable to other recognised differences among current global climate models such as cloud and convection parameterisations. PILPS emphasizes sensitivity studies with and intercomparisons of existing land-surface codes and the development of areally extensive datasets for their testing and validation (Henderson-Sellers et al. 1993).

Recognising the importance of soil moisture in the climate system, the International Satellite Land Surface Climatology Project (ISLSCP), which is a contributing project of GEWEX, began the Global Soil Wetness Project (GSWP) in 1994 (Dirmeyer et al. 1999). The initial efforts of the GSWP were to implement a land-surface modelling effort using a CD-ROM set of land-surface data developed by ISLSCP (Meeson et al. 1995). The CDs contain, in addition to other information, meteorological observations and parameter datasets sufficient to obtain soil moisture estimates for 1987–1988 for a $1° \times 1°$ grid. Ten groups, using various LSSs, including BATS, Mosaic, multiple versions of SiB, and others, produced soil moisture fields for these two years. Through the GSWP, Entin et al. (1999) attempted to validate these soil moisture fields using various soil moisture observations from the Northern Hemisphere mid-latitudes. They found that no model was able to recreate the actual soil moisture for all the areas studied. They also discovered that no model was able to recreate the seasonal cycle of soil moisture in Illinois and Russia, though all the models were deficient in recreating the changes of soil moisture in Mongolia and China, some of the few locations where routine soil moisture observations are made. The quality of the forcing data vary greatly from place to place, and may be a factor in the poor performance of the LSSs over certain regions (Oki et al. 1999). A second CD-ROM set is planned by ISLSCP, which will contain data for at least ten years (1986–1995). The GSWP will use these data to force LSSs which should then address another of the main issues raised when citing the difficulty of performing soil moisture validation, namely that of simulating interannual variability.

C.4.3 Terrestrial Observations

Another essential component of terrestrial data assimilation is the regular provision of land observations with known error characteristics. The data assimilation problem is best posed when the state observations being assimilated have similar physical complements in the LSS, and these states have significant memory or inertia so that an improvement is preserved and (hopefully positively) impacts subsequent predictions. Observations of significance to terrestrial data assimilation include temperature (air temperature, surface skin temperature, canopy temperature, and soil temperature), moisture (near-surface humidity, surface and profile soil moisture content, surface saturation, total water storage, plant water content, depression storage, lakes and rivers), snow (aerial extent, snow water equivalent, depth), carbon and nitrogen (plants and soil), and vegetation biomass (height, leaf area index, greenness). Land-surface fluxes, such as runoff, latent and sensible heat flux, carbon and nitrogen flux, and radiative fluxes, can be used in terrestrial data assimilation in the context of backing out a mass or energy state correction through conservation equations. Generally, it is more robust to perform a multi-variate analysis or assimilation, where an observation is used to constrain multiple relevant LSS states (this is further improved when observations of several different states are used). Data assimilation methods are designed to merge predictions and observations depending on the perceived errors of each. Establishing these errors can be the most complex and subjective part of data assimilation. Therefore, it is critical that observation error characteristics be well established through instrument calibration and validation. Large-scale terrestrial data assimilation development has lagged behind atmospheric data assimilation, primarily due to a lack of suitable observations available regularly in time and space. However, with the deployment of several new Earth system remote sensing platforms, this situation is quickly changing. The status of a few particularly critical terrestrial observations is described in more detail below.

Remote sensing of surface temperature is a relatively mature technology (see Chapt. B.8). The land surface emits thermal infrared radiation at an intensity directly related to its emissivity and temperature. The absorption of this radiation by atmospheric constituents is smallest in the 3–5 and 8–14 μm wavelength ranges, making them the best windows for sensing land-surface temperature. Some errors due to atmospheric absorption and improperly specified surface emissivity are possible, and the presence of clouds can contaminate or obscure the signal. Generally, surface-temperature remote sensing can be considered an operational technology, with many spaceborne sensors making regular observations (i.e. Landsat TM, AVHRR, MODIS, and ASTER) (Lillesand and Kiefer 1987). The evolution of land-surface temperature is linked to all other land-surface processes through physical relationships, so it is an ideal observation to assimilate.

Remote sensing of near-surface soil moisture content is a developing technology, although the theory and methods are well established (Eley 1992). Long-wave passive microwave remote sensing is ideal for soil moisture observation, but there are technical challenges in correcting for the effects of vegetation and roughness. Microwave soil moisture remote sensing has been limited previously to aircraft campaigns (e.g. Jackson 1997a). There are several current or future space-borne passive and active (radar) microwave sensors that may be useful to derive soil moisture information in a data-assimilation context, including the Defense Meteorological Satellite Program (DMSP) SSM/I (Engman 1995; Jackson 1997b), the EOS-AMSR (Advanced Microwave Sounding Unit), the Tropical Rainfall Measurement Mission – Microwave Imager (TRMM-TMI), and the European Space Agency Soil Moisture and Ocean Salinity (ESA-SMOS) instruments. All of these sensors have adequate spatial resolution for land-surface applications but have a very limited quantitative measurement capacity, especially over dense vegetation and topographic relief. Because of the large error in remotely-sensed microwave observations of soil moisture, there is a real need to maximise its information by using data assimilation algorithms that can potentially account for this error.

An important and emerging technology with respect to terrestrial data assimilation is the potential to monitor variations in total water storage (ground-water, soil water, surface waters, water stored in vegetation, snow and ice) using satellite observations of the time variable gravity field. The Gravity Recovery and Climate Experiment (GRACE), an Earth System Science Pathfinder mission, will provide highly accurate estimates of changes in terrestrial water storage in large basins when it is fully operational after it has been launched successfully in 2002. Wahr et al. (1998) note that GRACE will provide estimates of variations in water storage to within 5 mm on a monthly basis (Rodell and Famiglietti 1999). Birkett (1995, 1998) demonstrated the potential of satellite radar altimeters to monitor height variations over inland waters, including climatically sensitive lakes and large rivers and wetlands. Such altimeters are currently operational on the ERS-2, ENVISAT and TOPEX/POSEIDON satellites, and are planned for the JASON-1 satellites.

Finally, snow aerial coverage and snow water equivalent can be monitored routinely by many operational platforms, including the Advanced Very High Resolution Radiometer (AVHRR), Geostationary Operational Environmental Satellites (GOES) and SSM/I. Recent algorithm developments even permit the determination of the fraction of snow cover within Landsat-TM pixels (Rosenthal and Dozier 1996). Cline et al. (1998), describe an approach to retrieve snow water equivalent from the joint use of remote sensing and energy balance modelling.

C.4.4 Data Assimilation Concepts and Methods

Charney et al. (1969) first suggested combining current and past data in an explicit dynamic model, using the model's prognostic equations to provide time continuity and dynamic coupling amongst the fields. This concept has evolved into a family of techniques known as *four-dimensional data assimilation* (4DDA). "Assimilation is the process of finding the model representation which is most consistent with the observations" (Lorenc 1995). In essence, data assimilation merges a range of diverse data fields with a model prediction to provide that model with the best estimate of the current state of the natural environment so that it can then make more accurate predictions (see Fig. C.7). The application of data assimilation in land-surface studies has been limited to a few one-dimensional, largely theoretical studies (i.e. Entekhabi et al. 1994; Milly 1986) primarily due to the lack of sufficient spatially-distributed hydrological observations (McLaughlin 1995). However, the feasibility of synthesising distributed fields of soil moisture by the novel application of 4DDA applied in a hydrological model was demonstrated by Houser et al. (1998). Most land data assimilation schemes include the following steps:

- An *error checking* procedure is used to correct or eliminate erroneous data (Bengtsson 1985). Observations can contain different types of error, including errors due to faulty instruments, improper processing, or unsatisfactory communication of the data.
- Observations can rarely be assumed to be physically and spatially identical to the modelled state. Therefore, an *observation operator* is often employed to facilitate bias correction, space and time interpolation, and range matching.
- The actual *analysis or merging* of observations with model predictions is performed using a data assimilation algorithm. Common data assimilation methods include direct insertion, Newtonian nudging, optimal or statistical interpolation, Kalman filtering, and variational approaches (often using an adjoint model).

"The process of replacing model values by 'observed' ones is called direct insertion" (Daley 1990) or updating. This method assumes "perfect" observations, or observations with no error. Thus, model predictions that are known to contain error are totally rejected and replaced with the perfect observation. Any spatial or temporal information advection is performed entirely through the model physics.

Newtonian nudging continuously adds a forcing function to the model's prognostic equations to "nudge" the model state gradually toward the observations. These small forcing terms, based on the difference between the simulated and observed state, gradually correct the model fields, which are assumed to remain in approximate balance at each time step (Stauffer and Seaman 1990).

Statistical interpolation is a minimum variance method that is closely related to kriging (Bhargava and Danard 1994). The technique can be traced back to Kolmogorov (1941) and Wiener (1949), who applied it to various areas of science and engineering. With the development of computer power, and through the inspiration of Gandin's publication, *Objective Analysis of Meteorological Fields* (Gandin 1963), most major western meteorological services were using statistical interpolation operationally by the mid-1970s.

The Kalman filter has been extensively utilised in data assimilation research (Ghil et al. 1981; Cohn 1982). The Kalman filter assimilation scheme is a linearised statistical approach that provides a statistically optimal update of the system states, based on the relative magnitudes of the covariances of both the model system state estimate and the observations. The principal advantage of this approach is that the Kalman filter provides a framework within which the entire system is modified, with covariances representing the reliability of the observations and model prediction.

Variational methods were first introduced by Sasaki in 1958, and their use proved effective because they can incorporate many constraints easily (Ikawa 1984). "The variational algorithm requires the computation of the gradient of the distance function to be minimised with respect to the model state at the beginning of the assimilation period" (Courtier and Talagrand 1990). Thus, variational assimilation is principally very simple. One first defines a scalar function that describes the distance between the observations and the model prediction. Then, one simply seeks the model solution that minimises this function (Courtier and Talagrand 1990). The complexity comes from the generally large size of the minimisation problem.

One of the major components of any assimilation system is quality control of the input data stream. Quality control (QC) refers to the process by which observational data and their attributes are analysed to identify data items which are likely to contain gross errors and the attempts to correct or remove such errors. Observation errors are usually of two types: *natural error* (instrument or representativeness error), and *gross* or *rough errors* (improperly calibrated instruments, incorrect spatial/temporal registration, incorrect coding of observations, or telecommunication errors). These errors can be either random or spatially and/or temporally correlated with each other. Clearly, QC for any single observation must involve information other than the observational datum itself. Common QC algorithms can be categorised as follows:

- *Theory*, *realism*, or *sanity checks* see if the observation absolute value or time rate of change is physically realistic. This check filters such things as observations outside the expected range, unit conversion problems, etc.
- *Buddy checks* compare the observation with comparable nearby (space and time) observations of the same type and reject the questioned observation if it exceeds a predefined level of difference.
- *Background checks* examine if the observation is changing similarly to the model prediction.

According to a 1991 National Research Council report, "to produce research-quality data from a new satellite mission, the observed data should be subjected to a critical evaluation by an assimilation system in order to identify error characteristics of the instruments and the algorithms" (National Research Council 1991). The assimilation system can provide a systematic and powerful means of merging new, remotely-sensed observations with all earlier and current *in situ* and remotely sensed measurements. In a real-time context, data assimilation can provide quality assurance and validation of the observations, and can provide rapid identification and diagnosis of problems that might otherwise go unnoticed for longer periods. The data assimilation system can extend the available observations in time and space to provide continuous fields for use in subsequent research and application.

The continuous confrontation of theoretical and observational knowledge in a data assimilation system presents a rich opportunity to better understand physical processes and observation quality in a structured, iterative, and open-ended learning process. Data assimilation is also an important tool to help us make sense of voluminous and disparate data types that are becoming available from new space-based Earth observation platforms. Inconsistencies between observations and predictions are easily identified in a data assimilation system, providing a basis for observational quality control and validation. Modern data assimilation techniques use relevant observations and a state-of-the-art land-physics model to estimate the state of the land surface. For each observation, a background value is derived from the model forecast for comparison. Systematic differences between observations and model predictions can identify systematic error. Thus, the consistency of the model provides guidance to identify observation problems in a data assimilation context. This methodology clearly illustrates the importance of a good quality forecast and an analysis that is reasonably faithful to the observations. If the land model makes reasonably good predictions, then the analysis must only make small changes to an accurate background field (Hollingsworth et al. 1986). In many cases the analysis fields can provide guidance for identifying observational problems that can be compared with carefully chosen *in situ* observations to provide conclusive proof.

C.4.5 Current Projects

Subsurface moisture and energy stores exhibit persistence on various time scales that have important implications for extended climatic and hydrological predictions. Because these stores are time-integrated, errors in NWP forcing accumulate in them, which leads to incorrect surface water and energy partitioning. Land Data Assimilation Systems (LDAS) which are uncoupled LSSs that are forced primarily by observations and are therefore not affected by NWP forcing biases, are currently under development (Brutsaert 1998). The implementation of a LDAS also provides the opportunity to correct the model's trajectory using remotely-sensed observations of soil temperature, soil moisture and snow using data assimilation methods.

A multi-institutional LDAS research effort involving NASA, NOAA, Princeton University, the University of Washington, Rutgers University and the University of Maryland is currently under way. This LDAS operates in both retrospective and real-time modes at a $1/8°$ resolution over the continental United States using several different land-surface models. Project information and a real-time image generator are located at the LDAS web site: *http://ldas.gsfc.nasa.gov/*. Model parameters are taken from the high-resolution AVHRR-derived vegetation and soil survey classifications (Mitchell et al. 1999). Figure C.10 shows July average downwelling surface short-wave derived from GOES and the Eta model, total monthly precipitation derived from NEXRAD radar, gauges and the Eta model, experimental LDAS average skin temperature predictions, and experimental average near-surface soil moisture. A more complete description is given above.

NASA and NOAA are currently extending the North American LDAS project described above to all global land. This high-resolution, near real-time Global Land Data Assimilation Scheme (GLDAS) will use all relevant remotely-sensed and *in situ* observations within a land data assimilation framework. This development will increase greatly our skill in land surface, weather, and climate prediction, as well as provide high-quality, global land-surface *assimilated data fields* that are useful for subsequent research and applications.

Loosely linked to GLDAS are other projects, like the European project called ELDAS (Development of a European Land Data Assimilation System to predict Floods and Droughts), which is supported by the European Union 5th Framework Programme (*http://www.knmi.nl/samenw/eldas/*). ELDAS has been designed to develop a general data assimilation infrastructure for estimating soil moisture fields on the regional (continental) scale, and to assess the added value of these fields for the prediction of the land-surface hydrology in models used for numerical weather prediction and climate studies.

ELDAS uses a common infrastructure implemented at three participating institutes: ECMWF, DWD and CNRM/Météo France. The procedure followed will be able to generate soil moisture fields for a suite of land-surface schemes – respectively TESSEL (Tiled ECMWF Surface Scheme of Exchange processes at the Landsurface), TERRA (SVAT from the DWD) and ISBA (Interaction Sol Biosphère Atmosphère of CNRM) – but also others. The analysis method follows on the work by Rhodin

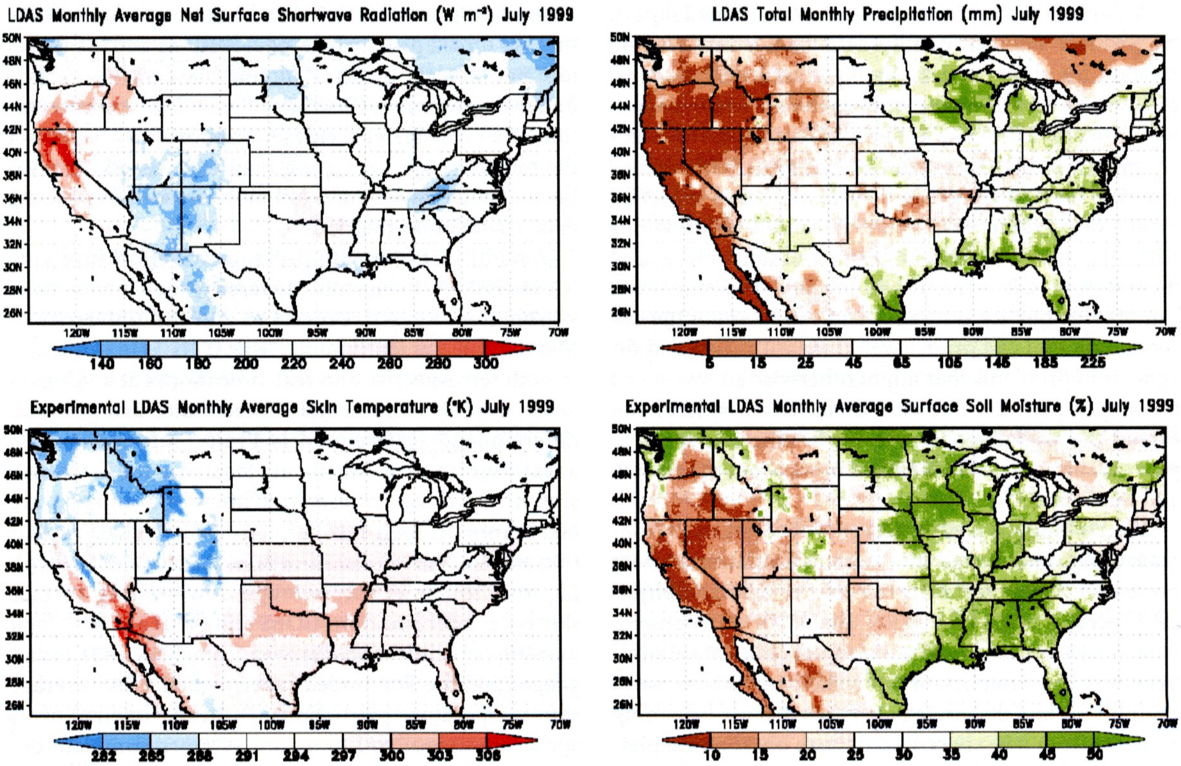


Fig. C.10. An example of atmospheric forcing and land-surface state fields from the North American LDAS project

et al. (1999) and Hess (2001). Simulations will be carried out with the full atmospheric model, but with model precipitation and radiation replaced by the observed data. Daily soil moisture field will be generated for a grid covering Europe in a sequential, cycled way, updating the atmospheric initial fields using analyses, and propagating the soil fields as first guess. The model grid will be different for different case- and validation studies, but use a common set of up/down scaling procedures.

Integral to the project are validation studies meant to assess the quality of the soil moisture fields using independent data from the GSWP 2000 dataset (Global Soil Wetness Project of GEWEX), from SSM/I (Special Sensor Microwave/Imager) or AMSR (Advanced Microwave Sounding Unit) validation (comparing these to computed top-of-the-atmosphere microwave radiation, generated from the surface state in the data assimilation modules), from GLDAS output (GLDAS forcings will be fed to the ELDAS data assimilation system) and from French river basin datasets. The usefulness of the dataproduct will be assessed in flood forecasting systems for UK rivers and the Rhine, and in European numerical weather predictions.

At the European Centre for Medium-Range Weather Forecasts, observations of screen-level temperature (SLT) and humidity are assimilated using an optimal interpolation technique (Douville et al. 2000). Over land, screen-level winds are not included in the assimilation because it is felt that the observations reflect local circulations, poorly described in the assimilating model. Snow-depth observations are combined with a model snow density field and a short-term forecast background to produce an analysis of snow mass (water equivalent). An analysis of soil water is performed, based on the SLT analysis (Fig. C.11). The errors in short-term forecasts of SLT are combined linearly in an optimal way to produce soil water corrections (Douville et al. 2000). Note that, in sharp contrast to the atmosphere, no remote sensing information is used in the surface data assimilation, hampering the quality of the analysed products in data-void areas. All the variables described above are analysed in the 40-year ECMWF re-analysis.

Terrestrial data assimilation systems are also under development outside the immediate meteorological context. The CAMELS (Carbon Assimilation and Modelling of the European Land-Surface) project recently started the development of a prototype carbon cycle data assimilation system (CCDAS) in order to produce operational estimates of "Kyoto sinks". To produce a best estimate of carbon uptake CAMELS will use all of the constraints implied by the different data sources, as well as of the physiological and ecological constraints embodied in terrestrial ecosystem models (see Fig. C.12). This is essentially a data assimilation problem, requiring a system similar to those used to initialise weather fore-

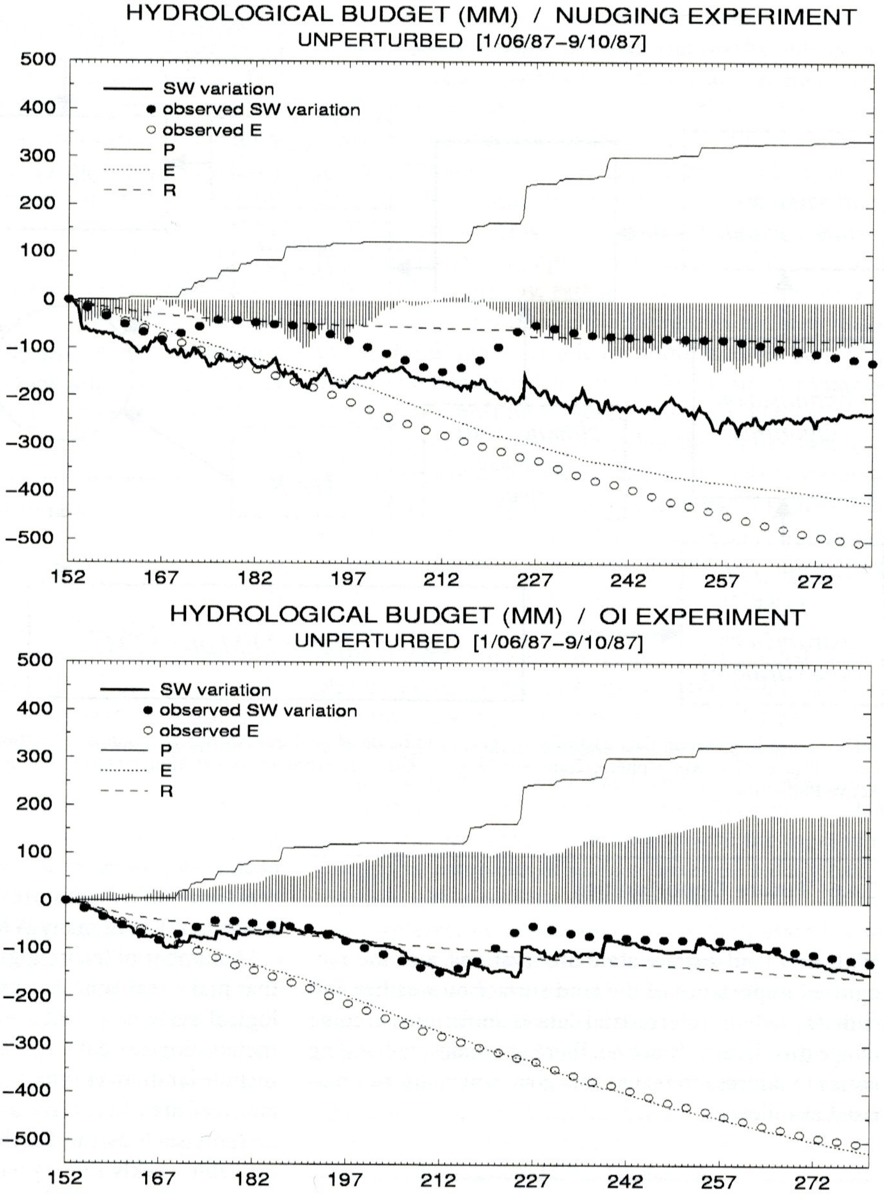


Fig. C.11. Hydrological budget simulated from 1 June to 9 October 1987 by the ECMWF single-column model forced with observed precipitation and radiative fluxes; *top*: nudging technique, *bottom*: OI technique. *P*, *E*, *R*, and *I* stand for the integrated precipitation, evaporation, runoff, and soil moisture increments respectively. The increments are represented by *vertical bars*. The observed soil moisture variation and evaporation are represented by *black disks and circles*

cast models. In this case observations are used to constrain the internal parameters of the terrestrial ecosystem models, while they are used to interpolate the observations.

The CCDAS scheme will use existing data sources (e.g. flux measurements, carbon inventory data, satellite products) and the latest terrestrial ecosystem models to produce operational estimates of the European land carbon sink. The terrestrial ecosystem models TRIFFID (terrestrial carbon cycle model from Hadley Centre, Cox et al. 2000), BETHY (Biosphere Energy Transfer Hydrology model from the Max Planck Institute for Meteorology, Knorr 2000; Knorr and Lakshmi 2001), ORCHIDEE (ORganizing Carbon and Hydrology In Dynamic EcosystEms from Laboratoire des Sciences du Climat et de l'Environnement, Viovy et al. 2001; Friedlingstein et al. 2001) will simulate the European land carbon sink at high resolution using operational analyses plus remote sensing products (e.g. seasonally-varying fAPAR). Atmospheric transport models will link terrestrial ecosystem model-predicted carbon fluxes to atmospheric CO_2 observations. Inverse models (e.g. Bousquet et al. 2000) will then be used to adjust both terrestrial ecosystem model parameters and prior estimates of carbon fluxes based on a 20–25 year simulation period. Eventually an online implementation of the CCDAS in a high-resolution atmospheric General Circulation Model (AGCM) will use the existing data assimilation structure in the AGCM plus the terrestrial ecosystem inverse model to nudge internal model variables, such as respiring soil carbon and leaf nitrogen, based on the atmospheric measurements.

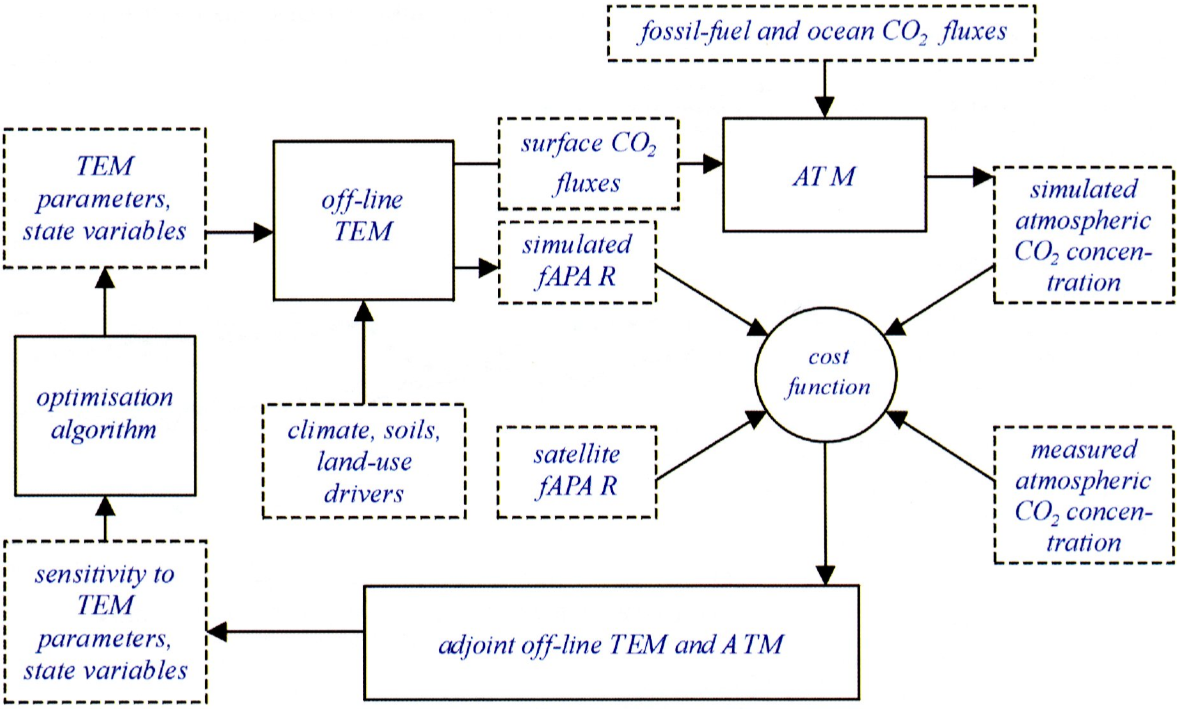


Fig. C.12. Off-line carbon data assimilation system to be developed in CAMELS (Carbon Assimilation and Modelling of the European Land Surface; *ATM*: Atmospheric Transport Model; *TEM*: Terrestrial Transport Model; *fAPAR*: fraction of Absorbed Photosynthetically Active Radiation)

C.4.6 Future Opportunities

With new land-surface state observations, and the recognised importance of the land surface on weather and climate prediction, terrestrial data assimilation will come of age this decade. However, there are many challenging issues to address to realise this goal, which are summarised as follows:

C.4.6.1 Terrestrial Observation

Profile soil moisture is arguably the most critical land state, and it remains largely unmeasured. Castelli et al. (1999) showed there is some soil water information in infrared radiative temperature measurements, but the best hope for operational soil moisture observations is with passive L-band microwave sensors. They provide information on soil water in the top 5–10 cm of soil, in areas free of contamination from the canopy water. An assimilation model can be used to solve the inversion problem and obtain a profile of water in the root zone (Calvet et al. 1998).

Subsurface soil temperatures have time scales of the order of days to years that can serve as very valuable diagnostics for anomalies in the surface forcing, in particular those related to cold season dynamics (Viterbo et al. 1999). An effort is needed to disseminate in near-real time soil temperatures, observed regularly down to a depth of 1 m in many WMO stations.

A number of technological capabilities are maturing that make real-time analysis of land-surface hydroecological systems possible – an ecological equivalent to meteorological data assimilation. Key ecological data include land-cover type, vegetation phenological status, and leaf area index are measured globally by satellite systems such as the Earth Observing System and are available weekly. Energy balances and carbon flux measurements from FLUXNET eddy covariance towers are now becoming available continuously (see Chapt. B.4).

Independent, continuous and comprehensive validation is critical to the success of any terrestrial data assimilation system. The structure of the assimilation scheme ensures the accurate reproduction of assimilated observations, but it is recognised that data assimilation constraints can cause other model predictions to diverge from reality.

C.4.6.2 Terrestrial Simulation

Comprehensive, physically-based terrestrial simulation models must be developed that include standard and accepted processes such as water and energy balance, evapotranspiration, soil moisture depletion, stream runoff routing, groundwater interaction, surface water and wetland processes, cold season dynamics, urban processes, photosynthesis, carbon and nitrogen cycling, plant and ecosystem dynamics, vegetation primary production, and coupling of all these processes with the overlying atmosphere. These models must also parameterise realistically the processes arising from sub-grid heterogeneity in precipitation, radiation, vegetation, soils, topography and atmospheric turbulence. These ecohydrological process models must have physically realistic states and parameters to ease their use in terrestrial data assimilation systems. Finally, parameter calibration methods must be fully developed and tested to specify robust parameters for large-scale terrestrial simulation.

C.4.6.3 Terrestrial Data Assimilation

Multi-variate terrestrial data assimilation methods must be further refined for operational use. These methods should include use of both *in situ* and remote measurements of soil temperature and moisture, and snow cover and depth. Longer-term goals should be the assimilation of runoff, groundwater and vegetation characteristics. The emphasis should be on combined use of observations and coupled models.

Hollingsworth et al. (1986) showed that assumptions on the bias and horizontal correlation structure of the model and observations can have a significant impact on error estimations. In practice, data assimilation is often implemented with the assumption that observations and predictions are unbiased and uncorrelated in space. These assumptions work reasonably well for *in situ* observations, but satellite observations are usually biased by inaccurate algorithms, and their errors are usually correlated horizontally because the same sensor is making all the observations. These assumptions must be evaluated, and observation and model errors must be defined better.

Terrestrial models have very little physics that act in the horizontal dimension, therefore it may be possible to limit the terrestrial data assimilation to one dimension. The benefits and drawbacks of using a one-dimensional (vertical) or a multi-dimensional (vertical, horizontal, and time) assimilation method must be evaluated. Multi-dimensional assimilation methods are much more computationally intensive, but maybe able to extend observations better into data-sparse regions. The exploration of sub-grid scale terrestrial data assimilation (i.e. into a tiled or mosaic model) and the ability of the data assimilation algorithm to downscale observations to fine resolutions also remain unexplored issues.

Data assimilation systems can include a radiative transfer observation operator, to allow the assimilation of radiances directly, rather than derived quantities. However, non-linearity in the forward model may make this method non-workable for some land-surface variables. The benefits and drawbacks of direct radiance assimilation must be explored more thoroughly.

Snow cover is a readily available observation that may be of great use in terrestrial data assimilation systems. However, terrestrial models usually use a snow water equivalent state variable. Assimilation and modelling methods need to be able to use snow cover information in order to infer snow mass (i.e. Liston 1999).

The development of off-line land data assimilation systems must be further developed and included in coupled prediction systems. A surface data assimilation using information from a variety of ground-based and remote-sensing observations will better produce realistic soil temperature, soil water and snow mass global fields. Moreover, short- to long-term forecasts in such a system will provide the community with the most reliable global estimates of surface fluxes on a daily basis, together with a realistic diurnal cycle.

An interesting aspect of LDAS is the possibility of multiple land-surface predictions made by several different LSSs, with various initialisations. These various land-surface prediction ensembles and super-ensembles will be intercompared and explored. It is quite possible that through this land-surface ensemble prediction strategy, spatially-varying confidence limits on predictions could be established and new ways to evaluate and improve predictions could be developed.

Finally, there is a need to perform additional long-term coupled land-atmosphere re-analyses that include improvement in terrestrial simulation, observation, and data assimilation. The resulting datasets would improve greatly our understanding, prediction, and practical application of terrestrial systems.