USE OF FIELD DATA TO DIAGNOSE LAND-SURFACE INTERACTION.

Alan K. Betts
Atmospheric Research, Pittsford, Vermont
akbetts@aol.com
Pedro Viterbo, Anton Beljaars (ECMWF) and Bart van den Hurk (KNMI)

ECMWF Semina, Sept. 6-10, 1999 **/Updated 1/5/2000**

1. INTRODUCTION

The land-surface interaction is as important to the climate of a global model as the sea surface boundary condition. However over land, there is no measured input field analogous to the sea surface temperature, which controls the surface sensible and latent heat fluxes (together with surface wind, air temperature and humidity). The fluxes over land in contrast are driven on diurnal time scales by the net radiation, and the partition into latent and sensible heat, which depends on the availability of water (either in the soil or in surface reservoirs) for evaporation. The accuracy of the down-welling fluxes depends on the model radiation physics, and the specification of clouds and aerosols. The outgoing fluxes depend on the calculated surface radiation or skin temperature, and the albedo, which in turn depends on snow cover, vegetation type and season (for example, leaf-out). Thus even the calculation of net radiation at the surface involves many physical parameterizations. The availability of water for evaporation depends on both a realistic hydrological model (balancing model precipitation, surface and soil water reservoirs and runoff) and a realistic vegetation model (to extract soil water for transpiration as a function of photosynthetic processes). At higher latitudes, soil water only becomes available after the ground melts, so the soil thermal balance processes, and the timing of snow melt (snow insulates the ground) also control the seasonal cycle of transpiration. However global fields for soil water and soil temperature are not available for analysis, so a model must derive them from its own physical parameterizations, using near surface atmospheric measurements as constraints (e.g. *Douville et al.*, 1999).

However the coupling between surface fluxes and the convective boundary layer is so tight, that errors in any of the physical parameterizations, whether sub-surface hydrology or thermal transfers, vegetation parameterizations, or stable/unstable boundary layer parameterizations can all interact to give an erroneous diurnal cycle of the mixed layer, and in turn an erroneous diurnal cycle of convective precipitation in a model. This paper will address efforts to use data from both field programs and on the scale of river basins to assess errors in the model formulation of the land surface interaction.

2. VALIDATION TOOLS AND CRITERIA.

We shall consider three general classes of validation tools.

a) Off-line comparisons with field data of model land surface physics, driven by observed near-surface meteorological forcing. These are useful tests of the ability of the model land-surface physics to reproduce the diurnal and seasonal cycle of the surface fluxes given known meteorological forcing.

b) Comparison of model output time series at a single grid-point with field data time series. These test the model's ability to reproduce the surface diurnal and seasonal cycles with a fully coupled surface and boundary layer.

c) Comparison of basin averaged model surface fluxes with basin averaged precipitation and runoff.

These give insights into key aspects of the model hydrology on regional scales: the accuracy of the precipitation in the analysis cycle and in short-term forecasts, and over the diurnal cycle: the realism of the runoff parameterization and of spatially averaged evaporation. In addition, at high latitudes, we can see how well the model handles the seasonal aspects of the frozen and liquid hydrology.

Several general issues are worth discussing. One perennial question is field data representivity. Current high resolution global models have effective grid resolutions of 50 km or greater. Can a grid-point time-series from a global model be usefully compared with a time-series of field data, which are representative of only a few square km? Field sites are usually located over carefully chosen vegetation types with adequate fetch for eddy correlation measurements to give representative fluxes on the hourly timescale. There are now 50-100 towers globally measuring continually the surface radiation, energy, CO_2 and hydrological balances. Most are on towers over forests typical of a region, some over grassland and crops. They are a rich source of validation information for global models, which are not yet fully utilized. Many sites were installed primarily to study the global carbon balance for climate purposes, and not short-term meteorological issues for which the sensible and latent heat fluxes are primarily surface drivers. However it is much easier to compare point time series measurements with a forecast model analysis or short-term forecast than say with a climate model output, because the forecast model may represent quite well the local advective processes associated with precipitation systems. A model represents a grid average, not a single vegetation type, but if a model has a tiled structure, then a single representative vegetation tile can be compared with the tower flux site data. However exact agreement between measurements and model cannot be expected as flux sites are representative of much smaller areas than a GCM grid.

Instead we ask these questions.

i) How well does the model represent the observed diurnal cycle of the soil temperatures, surface temperature, humidity and the surface fluxes?

ii) How well does the model represent the corresponding observed seasonal cycles, and in addition the seasonal cycle of soil moisture, snow depth, snow melt and soil melt, albedo, green up and leaf-fall?

iii) How well does the model represent the transitory response to precipitation events, such as runoff and soil dry downs?

iv) What do the model error fields, including the nudging of soil water, indicate about the model physics errors?

For forecast model studies, we can use the same validation dates (for example for (i)), while for climate model comparisons, longer term averages must be compared, such as the monthly (or seasonal) mean diurnal cycles, bearing in mind any differences in the observed monthly precipitation and the climate model precipitation. A study of the diurnal cycle is essential since day-night differences between the stable and unstable BL play a major role. The stable BL in particular is often poorly modeled, and night-time measurements are also less reliable, particularly under low wind conditions. Indeed more research on the stable BL is needed. If only daily or monthly averaged model output are analyzed, compensating nighttime and daytime errors can be missed. In addition, the diurnal temperature and humidity range are key climate parameters (although unfortunately the humidity range is not always measured at routine climate stations).

The global network of flux measurements is a rich resource, and I urge a major international cooperative effort to process and analyze this data for forecast and climate model validation and development.

3. OFF-LINE COMPARISONS WITH FIELD DATA

The illustration I shall use is the development of the ERA-40 tiled land surface model. This is work in progress (which will be published as *Van den Hurk et al.*, 2000): the figures in this section were updated from those presented at the semina . The data set is from the Northern Study Area (NSA) of BOREAS (Boreal forest Ecosystem-Atmosphere Study) from (55.9°N 98.5 °W) near Thompson, Manitoba in Canada for the time period 1994-1996. A 30-min. driver data set for three years was assembled primarily from two mesonet stations, taking measurements a few metres above the forest canopy. The two sites were typical of the region. One was an old jack pine stand on sandy soil, covered with lichen. Tree heights were 10-13 m and the measurements were at 16 m, about 3 m above the canopy. The second site was a mixed forest of spruce and

poplar on a clay and peat layer soil with a thick layer of moss; with tree heights also to 13 m, and tower measurements at 19 m about 6 m above the canopy.

The measured driver data set variables were above canopy values of

Wind-speed, pressure, air temperature, and specific humidity
Incoming shortwave and long wave radiation
Precipitation.

Precipitation required additional data and special processing. In winter, two weighing Belfort precipitation gauges were used to measure snowfall. In summer an additional array of 10 precipitation gauges (installed to study the hydrology of a small basin in the NSA) were included, together with rain-gauges at four tower sites to get a better representative average.

For validation data we have primarily a nearly continuous set of flux measurements over a black spruce site (*Goulden et al.*, 1997: *Betts et al.*, 1999a). Black spruce is the dominant land cover, although other conifers are present (such as jack pine) and deciduous species such as aspen; in addition to fens and lakes. For the summers of 1994 and 1996 flux measurements are also available over fen and jack-pine sites in this study area. Aircraft flew a grid patterns to assess the relation between spatially averaged fluxes, and those measured at the flux tower sites. (*Barr et al.*, 1997; *Barr and Betts*, 1997).

3.1 Comparison of model and old black spruce site data.

Figure 1 compares the three year cycle (from January,1994 to November, 1996) of monthly mean data from the ERA-40 model [version dated 11/2/1999] and the NSA old Black Spruce site (NOBS). Net radiation (R_{net}) is shown positive and the sensible (SH) and latent heat (LH) fluxes are shown negative for clarity. The model ground heat flux is shown positive in summer when it is downward. There was no measured ground heat flux for the data. The monthly mean net radiation in the model is generally biased slightly low relative to the data. There are two reasons for this. One is albedo: the model albedo in winter with snow under the trees is about 20% for the grid average, while the spruce site measured albedo rarely reaches 0.15 (Betts and Ball, 1997). In summer, the model climate albedo field at this grid point is 0.12, while the spruce site has a very low albedo (about 0.08). Errors in the skin temperature of the model also introduce small biases in the outgoing long-wave temperature. The characteristic feature of the observed SH and LH fluxes (solid lines) at these latitudes is the phase lag in Spring of the LH flux. SH typically has a double maximum in April and June; while evaporation is very low in April (because the ground is still frozen), and reaches its summer maximum in June (July in 1996). The model captures this feature well, although the model has generally a lower SH and higher LH than the data.

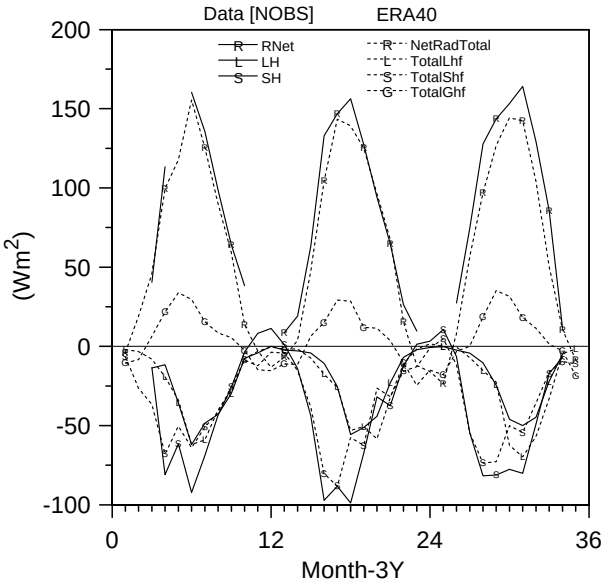


Fig. 1. 3-year monthly mean fluxes from data and ERA-40 (version 11/2/99)

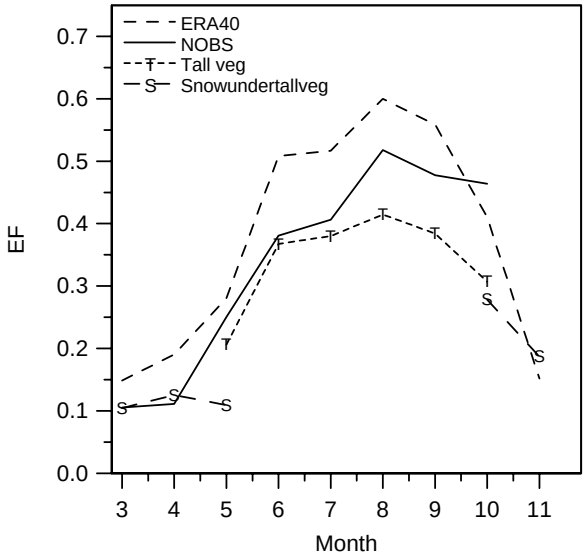


Figure 2 shows the mean annual cycle of evaporative fraction (EF) from March to November from the model (dashed) and the NOBS flux site (solid), as well as two of the model tiles. The "snow under tall vegetation" tile is shown in Spring and Fall, for periods when the ground is snow covered and the "tall vegetation" tile is shown from May to October. In the months of partial snow cover (May and October), the tile averages are for the days with and without snow. The model reproduces well the mean rise of EF from Spring to Fall that is observed, and the two model tiles for the dominant vegetation type agree well with the black spruce observations. Note that the average over all the model tiles exceeds the EF for the NOBS data in almost all months (see later).

Fig. 2 Mean annual cycle of EF from March to November.

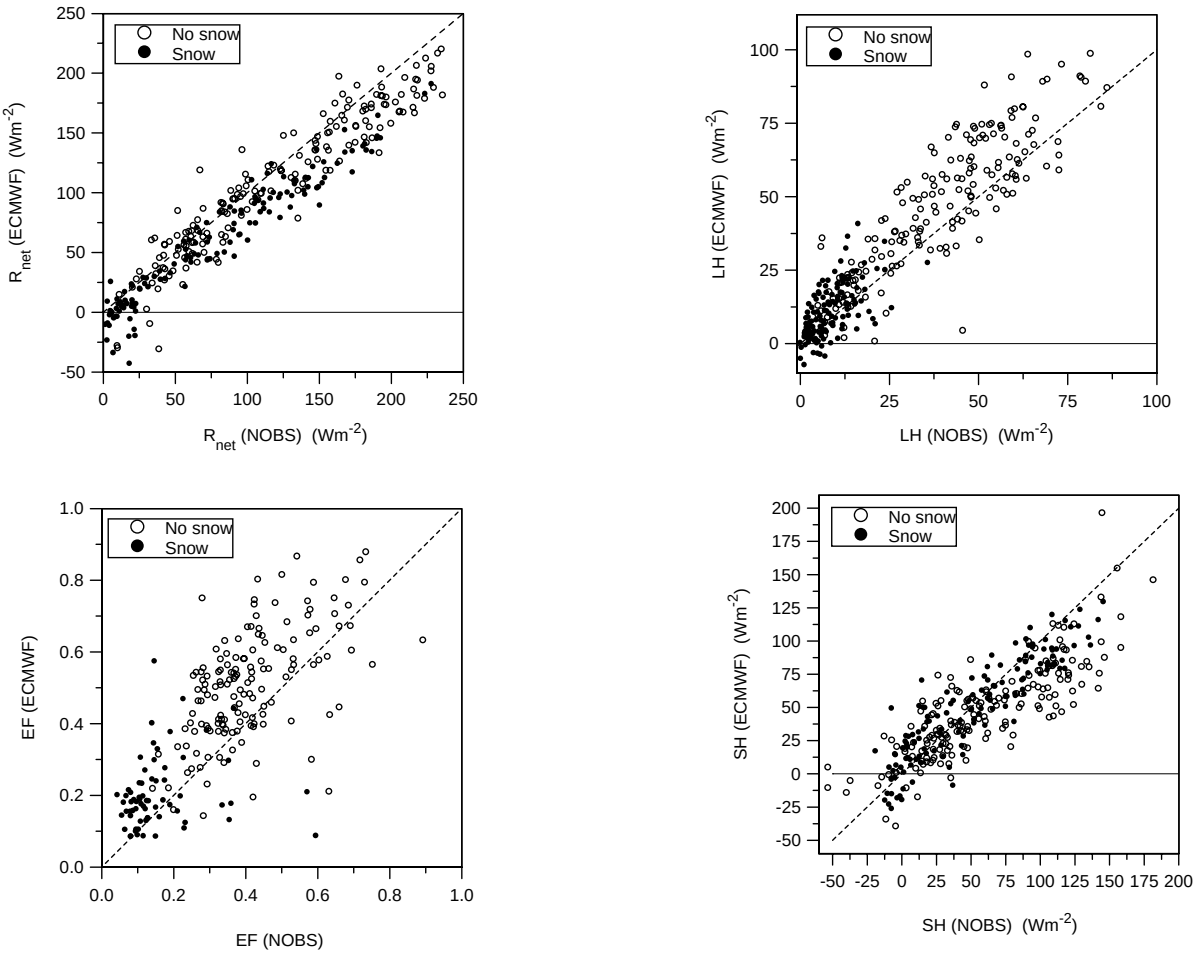


Fig. 3. Comparison of R_{net}, LH, SH and EF (clockwise order)

Figure 3 compares daily mean fluxes and evaporative fraction (EF), partitioning the data into days with snow on the ground (generally from November to mid-May) and those with no snow. The upper left panel is the R_{net} comparison: it is clear that the model low bias of R_{net} is larger when there is snow on the ground. The panels on the right compare LH and SH fluxes. The clear separation in evaporation between warm and cold seasons is reproduced by the model, with low evaporation when there is snow on the ground, and the ground is frozen. The high bias of the model evaporation (which represents a grid average (see section 3.3) is mainly in summer. The correlation between model and observed LH fluxes is high. The SH comparison shows a low model bias in the warm season, with less bias in the cool season. The final panel of EF is a useful summary, showing that although the model properly distinguishes between the warm and cold seasons, EF in the model, though well correlated with the black spruce site data, is generally biased high. The few outliers in winter (when the data shows high EF and the model low EF) are primarily days of snowfall: perhaps because the model has no model for the evaporation of snow on the canopy (the interception reservoir is only a liquid reservoir).

3.2 Verification of diurnal cycle

In these off-line comparisons, the radiative and atmospheric forcing data are specified. However, the diurnal cycle of the surface heat fluxes and skin temperature of the model are independent checks on the model physics. We shall show two illustrations of this. Figure 4 shows the mean summer diurnal cycle of air temperature (specified) above the canopy, the radiometric skin temperature from two nearby BOREAS NSA mesonet sites (labeled #8 for the site at Thompson and #9 for the northern old jack pine site). A characteristic of these forest sites is that the canopy skin temperature is quite closely coupled to the air temperature, even at night. The heavy dashed line is the ERA-40 model. Although the model skin temperature has a tighter coupling to the ground at night than during the day (*Van den Hurk et al.*, 2000), the model skin temperature drops further than the observed skin temperatures at night. In day-time the skin temperatures agree well, although there is a small lag after sunrise while the model skin temperature rises. What physics is poorly represented in the model that in nature prevents the drop of skin temperature at night? Figure 5, which shows the corresponding summer diurnal cycle of the model and observed heat fluxes, gives a possible clue. The SH and LH fluxes are plotted negative in the daytime (local noon is about 1800 UTC). First note that the primary balance in the model at night is between outgoing net radiation and ground heat flux. Unfortunately we do not have ground heat flux measurements at night and percent errors in the measurement of R_{net} at night are larger than in the daytime,

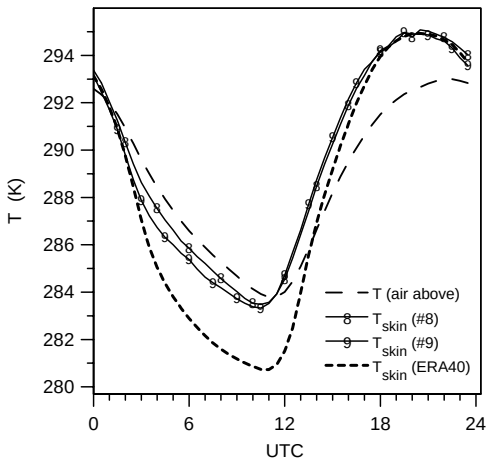


Fig. 4. Comparison of mean summer diurnal cycle of air and skin temperature

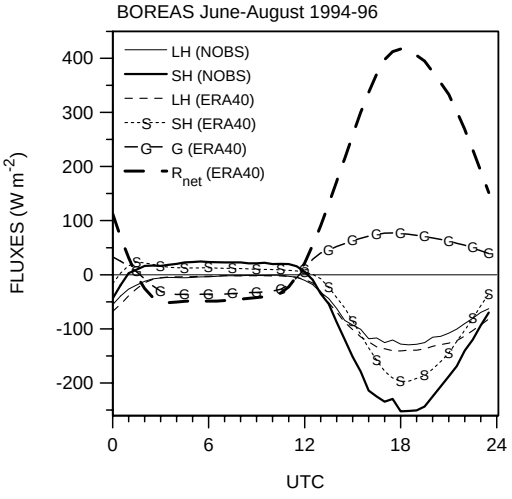


Fig. 5. Comparison of mean summer diurnal cycle of surface fluxes.

so we have no good observational check on this. At night the LH in both model and data are small as expected. However the *observed* SH flux at night, while small, is greater than in the model. This is significant as the flux estimates in the stable boundary layer at night are *underestimates* except at higher wind speeds (Goulden et al. 1997). It seems likely, as implied by the smaller drop in observed skin temperature at night in Figure 4, that the thermal coupling between the forest canopy and the atmosphere is much larger than in the model, that is the downward SH flux in the model under stable conditions at night is probably too small, perhaps by a factor of two.

3.3 Comparison of model, tower and aircraft data

In this section, we show a comparison for the BOREAS NSA between the ERA-40 off-line simulation, the NOBS site and spatially averaged data from flights of the Canadian Twin Otter aircraft. Two types of aircraft patterns are included. The primary one was a grid pattern, consisting of 9 legs 16 km in length, which mapped a 16 by 16 km square, situated over the tower flux sites of the NSA. This pattern was flown twice, the second time with a reverse heading. This grid gives a representative average for a 16 by 16 km area (Desjardins et al., 1997; Ogunjemiyo et al., 1997). This area is smaller than an ECMWF grid-square, but it is still much larger than the flux footprint of a typical forest tower. On some days, instead of the grid pattern or in addition to it, repeated flights were made past the tower flux (TF) sites over patches of relatively homogeneous forest. These runs were typically 10 km in length but were repeated 6 by 8 times to give a representative flux. These TF runs were flown past the NOBS tower, the NSA old jack pine tower (NOJP) and the young jack pine tower (NYJP). No runs was made past the northern Fen site, because of limited fetch, but instead a site was chosen for repeated runs, where the vegetation was recovering from a recent burn (with characteristically more deciduous vegetation). The aircraft patterns took 2-3 hours to fly and were generally in the 1600-1900 time period (local noon is near 1800 UTC). We generated averages from the model off-line run and the NOBS tower data for time-periods corresponding to the aircraft pattern times. For days when the aircraft did not fly, we show the 1600-1900 UTC average.

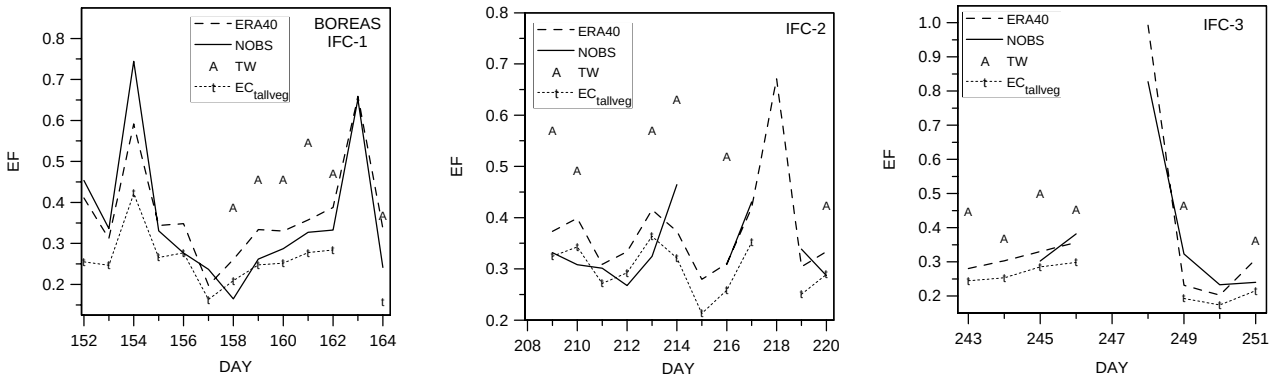


Fig.6 Near-noon comparison of model with NOBS and aircraft data

Figure 6 shows the model-data comparison of EF for parts of three intensive field campaigns (IFC's) in 1994, while the Twin Otter (TW) was in the NSA. The first " IFC-1", starting on day 152 (June 1), is shown on the left and the following IFC-2 and IFC-3 are later in the summer. Model and the NOBS data track quite well, with the model (dashed) usually higher than the old black spruce site (as shown in Figure 3). The peaks in EF are days with low net radiation. The symbols A mark the averages for the Twin Otter aircraft flights for just those days when either the grid pattern was flown or a representative set of the TF and burn site tracks. While the aircraft spatial average tracks the NOBS tower site reasonably well, it is higher by about 0.14, and the aircraft is also higher than the model. In addition, we show the $EF_{tallveg}$ of the ERA-40 tile representing tall

vegetation for this grid square, excluding days when the interception reservoir fraction exceeds 20%. This is lower than the ERA40 tile average, but generally agrees quite well with the NOBS data or is a little lower on some days. Why is the aircraft spacial average of EF higher than the NOBS site? One reason is that the EF from the NOBS site is the lowest of all the flux tower sites. Desjardins et al. (1997) also speculated from the study of the flight tracks past individual towers (which also show higher aircraft EF than at the towers) that the tower sites might have been chosen in drier locations.

Barr et al. (1997) made a separate estimate of spatially averaged EF from boundary layer rawinsonde budgets for the BOREAS study areas. Their estimate, shown in Table 1, for undisturbed days (when sequential rawinsondes were launched) over all IFC's, was a little less than that from the Twin Otter flights. Table 1 shows that given the difference in EF between spatial averages and the NOBS site (the dominant vegetation type), it appears that the ERA-40 model, which is a little higher than the NOBS site and lower than the spatial averages, may not be significantly biased with respect to the data.

Table 1. Comparison of BOREAS NSA Evaporative Fraction Estimates (Fig 6)

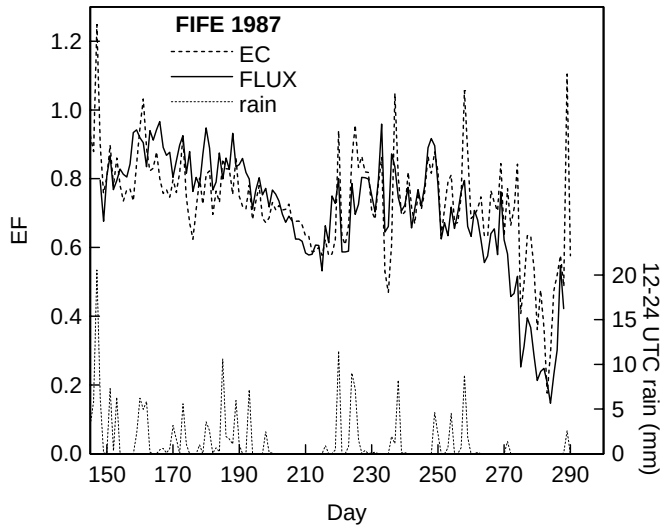
Source	EF (mean)
ERA-40	0.35
Twin Otter	0.47
Barr et al. (1997)	0.45
$EC_{\text{tall veg}}$	0.26
NOBS	0.33

4. VALIDATION OF MODEL OUTPUT GRIDPOINT TIME-SERIES

Although the off-line validation of model surface physics can illuminate many of the physical processes in the land-surface and vegetation model, it does not tell us how the interaction with the model boundary layer processes will affect near-surface atmospheric parameters as these are specified. To illustrate these we will show a few figures from *Betts et al.* (1998a), which compared output from ERA-15 with an average time-series for the FIFE site in Kansas, near 39°N, 96.25°W (*Sellers et al.*, 1988; *Sellers and Hall*, 1992).

4.1 Seasonal variation of EF and soil water

Figure 7 shows daytime evaporative fraction $(EF = LH/(LH + SH))$ over the summer season in 1987 for the ERA-15, labeled EC, and the FIFE flux data, a site average labelled FLUX. The model tracks the observations quite well seasonally, although it is biased low in Spring and high in Fall. However, there are notable model peaks on many days, when it rains in the model (lower curve on right hand scale). We concluded that the re-evaporation off the wet canopy in the model may be too high (or the data may suffer from some low



bias on wetter days). The success of the model in tracking EF seasonally, even without a seasonal variation of the vegetation is probably because of the tight link between evaporation and soil water. The ERA-15 model carries soil moisture as a variable for 4 layers 0-7, 7-28, 28-100 cm and a base layer of 100-289 cm. The first three contain the rooting zone (*Viterbo and Beljaars*, 1995). We calculated mean values for equivalent soil layers from the site-averaged gravimetric and neutron probe soil moisture data (see *Betts and Ball*, 1998).

Fig. 7 Comparison of daytime EF (above) with model precipitation (below)

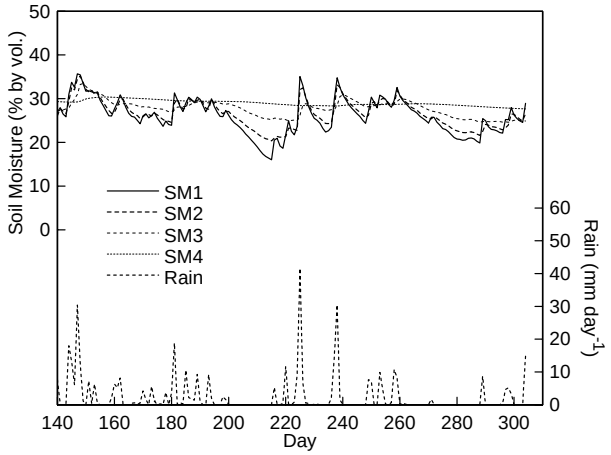


Fig. 8a Daily mean volumetric soil water and daily precipitation for ECMWF model (4 layers).

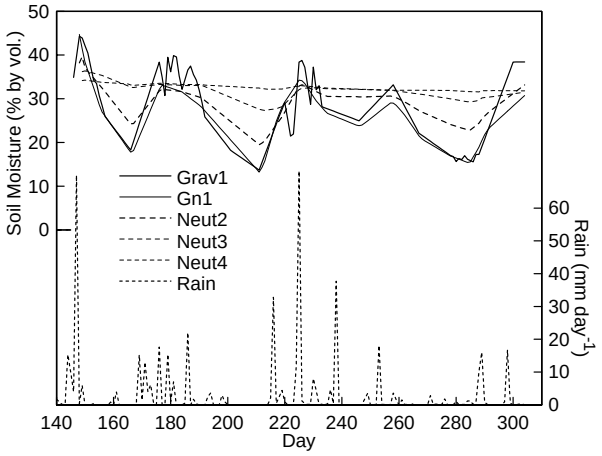


Fig 8b As Fig 8a for FIFE data.

Figures 8a,b show the comparison for the season. The left panel shows the model data: soil water for the 4 ERA-15 model layers, and precipitation in the model. We see the impact of the major rainfall events in the model (on right-hand-scale) on soil moisture. The large rainfall events bring the 4 soil layers to 30-35% water by volume. During the dry spells, particularly late July, the upper 3 layers dry out, and the upper layer falls to 16% soil moisture by August 3, when the first rain falls after 2 dry weeks. The response to rainfall events, and the dry-downs between them, seem satisfactory, as shown in *Viterbo and Beljaars* (1995) in off-line tests, although we will comment on the impact of soil moisture nudging in the model below. The right-hand panel is the corresponding set of graphs for the FIFE site-average of soil moisture and rainfall. The general pattern is similar, although there are many differences. The observed intense rainfall events on days 148 and 225 had more precipitation than in the model; not too surprising perhaps as the model is representative of a larger grid square of order 100 km, while the FIFE site is only 15x15 km. The deep soil moisture (for the 100-200 cm

depth layer) at the FIFE site had a value around 33%, slightly higher than in the model. In general, the amplitude of the fall in soil moisture by volume, for the surface (0-7 cm) layer is larger in the data than the model. We show 2 curves for this level: the light solid line is for all the gravimetric sites (converted to a volumetric value), while the light dashed line is for that subset of samples near and at the times of the neutron probe data (which are in a merged neutron probe data file in *Strebel et al.*, 1994). Irregular time sampling is a problem with the FIFE soil moisture data. During the 4 IFC's, almost daily gravimetric measurements were made, as can be seen from the high frequency fluctuations with rainfall events for the upper layer. However the neutron probe sites were sampled less frequently, so the corresponding curves are smoother. Between IFC's, sampling was least frequent, and as a result, some rainfall events, such as on day 238, were not seen at all by the soil moisture data. This sampling problem must be born in mind in comparing Figures 8a, and 8b. The data corresponding to the second model level (7-28 cm) are the poorest, since the neutron probe data at 20 cm is not reliable (see appendix to *Betts and Ball*, 1998), and we have simply averaged the gravimetric data at 7.5 cm (converted to volumetric soil moisture) and the 30-cm neutron probe data. The trough in observed soil moisture on day 166 in June is somewhat puzzling. There are measurements on this one day, but the model (which admittedly shows more precipitation than the data during this dry June spell) has no correspondingly low value, and curiously enough the observed evaporation remains high through this period.

Our main conclusion was that the ERA-15 model is reproducing the main features of the seasonal soil moisture behavior, but with reduced amplitude in the near surface layers. The model was developed using this same FIFE data (*Viterbo and Beljaars*, 1995), and when we compare with Figure 9 of that paper, we see that the amplitude in the reanalysis model is a little less than in the off-line simulation. The probable reason is that the subsequent addition of soil moisture nudging to the model (*Viterbo and Courtier*, 1995) has 'added' some soil moisture during extended dry periods. For the 10 dry days 205-214 (see next section), the loss of water due to evaporation is 4.35 cm (126 Wm^{-2} for 10 days), while soil moisture in the first 3 layers falls only 3.33 cm, considerably reduced, because nudging adds 0.80 cm of water to the same 3 layers. The small imbalance is probably diffusion from the base layer.

4.2 Diurnal variation of temperature and mixing ratio.

The diurnal variation of near surface variables in coupled model runs tells us whether the land-surface and BL schemes are interacting properly. The FIFE comparison for the period from Julian Day 205-215 (July 24-August 2, 1987) is of interest because there was no rain in either the model or the FIFE data. Figure 9 shows the sequence of temperature, *T*, and mixing ratio, *q*, (lower curves) at a 3-hrly time resolution. The time axis is a decimal Julian Day. The daily maximum temperatures steadily rise as the soil dries, and the surface *EF* falls (not shown). The reanalysis short-term forecast tracks the 2-m daytime temperature maximum very well, but the model T minimum is typically low by 3 K. The mixing ratio comparison shows 2 anomalies in the model, although the mean mixing ratio is approximately correct. There is a mid-morning (1500 UTC) peak of *q* in the model every day, and an evening (2400 UTC) *q* minimum (which is generally greater than the minimum observed). This diurnal cycle error in mixing ratio is probably the cause of the near noon-peak in the diurnal cycle of convective precipitation in ERA-15 (see section 5).

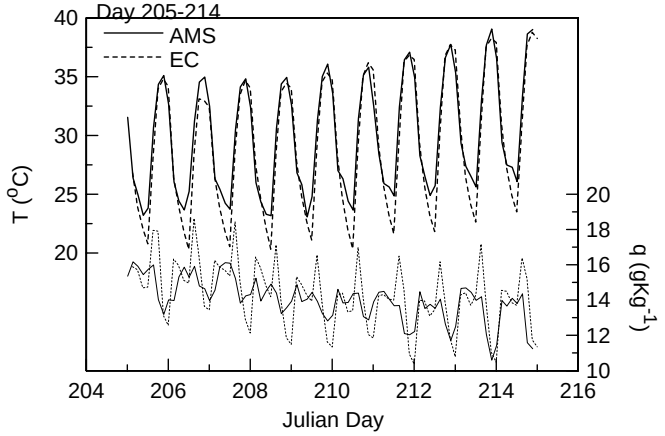


Fig. 9 Comparison of 2-m temperature and mixing ratio for late July drydown.

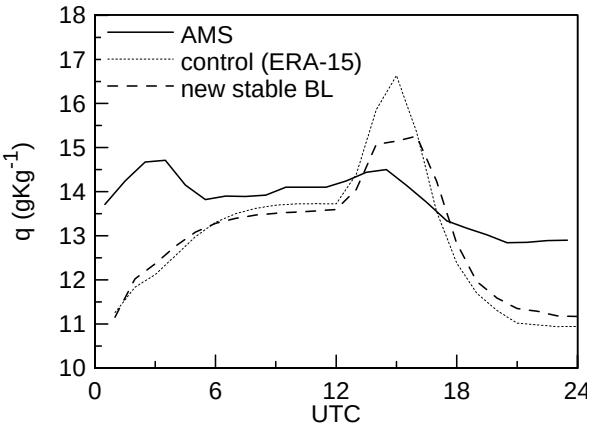


Fig. 10a Two-day average diurnal cycle of mixing ratio for data and two model versions

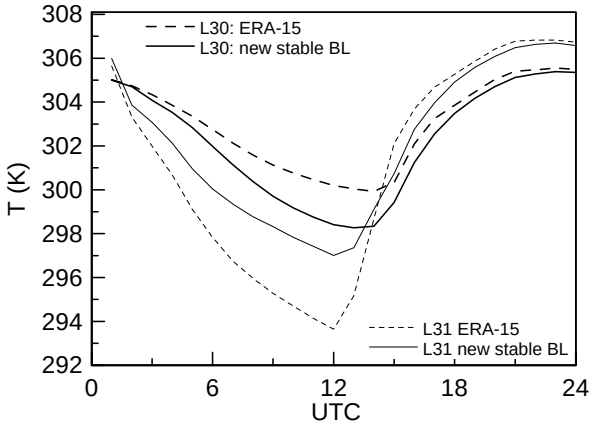


Fig. 10b Two-day average diurnal cycle of temperature at 2 lowest model levels for two model versions.

Changes were made to the stable BL parameterization from the ERA-15 model (which became operational in September, 1996, *Viterbo et al*, 1997) The cold bias of the model at night was reduced by increasing the coupling of the land surface to both the stable BL above, and to the ground below. One important effect of this change is to weaken the nocturnal inversion at sunrise, and this in turn affects the BL development in the morning, reducing the mid-morning peak in *q*. Figure 10a shows 3 curves of the diurnal cycle (at hourly resolution) for a 2-day average (from 0000 UTC on Day 209 to 0000 UTC on Day 211). The solid line is the FIFE AMS data at 2 m. The dotted and heavy dashed lines are for the lowest model level (level 31 at about 30 m above the surface) extracted from two 120-hr forecasts initialized at 1200 UTC on Day 208 (July 27, 1987), one a control using a model with the same land-physics as ERA-15, and the second with the new stable BL and ground flux scheme (heavy dashes). The control shows a mid-morning peak at 1500 UTC, followed by a much steeper fall than the data to a low afternoon value (as in Figure 9). For the forecast with the new stable BL scheme, the mid-morning peak of *q* is sharply truncated, a significant improvement, as the BL deepens sooner to the next model level 30 (not shown). The reason is that there is a significant difference in the stability near the surface at sunrise. Figure 10b shows the corresponding average diurnal cycle of temperature for the two model versions at level 31 and the next level 30 (about 150 m above the surface). Whereas the model with the improved stable BL is more than 3 K *warmer* near sunrise at L31, it is 1.5 K *cooler* at L30. Consequently as the lowest level 31 warms after sunrise, it starts mixing upwards to the next level 30 much sooner that in the reanalysis control model. This upward mixing sharply truncates the morning rise of q, as seen in Figure 10a. Note that the sharp fall of *q* in the model near local noon is not reduced much, and the late afternoon low bias is only slightly improved. It is clear that the diurnal cycle of the near surface meteorology over land results from the interaction of the surface fluxes over the full 24-hour period with the stable and unstable boundary layers.

5. VALIDATION OF MODEL BASIN AVERAGES

Basin-scale averages of the surface energy and water budgets on the scale of river basins cannot be determined with any reliability from the few scattered time-series measurements of the surface radiation budget and the surface sensible and latent heat fluxes. On the other hand, basin-scale averages of precipitation and streamflow can be used as evaluation data to help identify surface processes that are poorly represented in the forecast model, and thus lead to improvements in the modeling of the surface evaporation and hydrological response. One of the objectives of the GEWEX Continental International Project (GCIP) was to assess the ability of our forecast models to estimate the energy and hydrological balances for the Mississippi basin, which covers a total drainage area of $3.16\ 10^6$ km^2 (*Couglan and Avissar*, 1996). Here we will show results from the ECMWF reanalysis (*Gibson et al.*, 1997). ERA-15 has a 6-hr analysis cycle, and from every analysis short-term

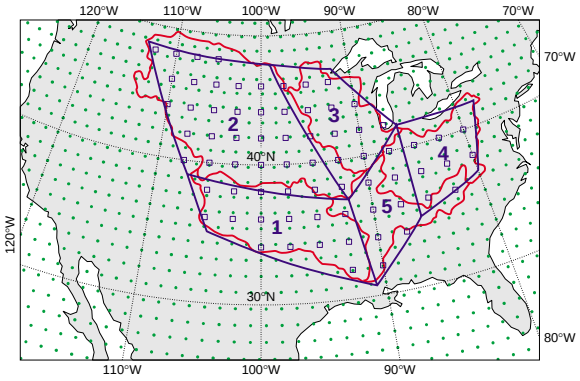


Fig.11 Mississippi sub-basins. Shaded dots are ECMWFphysics grid-points. Open squares are Higgins precipitation data points.

forecasts are run. The standard global grid-point archive contains the meteorological state variables and averaged surface fluxes every 3 hours. In addition, meteorological and averaged surface flux variables for selected grid-points and averaged quadrilaterals of grid-points were archived at an hourly time resolution as single column data sets. For the last nine years of the reanalysis from 1985-1993, average quadrilaterals were included for the five main sub-basins of the Mississippi, and here we show a few figures from *Betts et al.*, (1998, 1999), who analyzed the surface water and energy budgets. Figure 11 shows the physics grid-points (shaded dots) of the ECMWF T-106 ERA-15 model for the US: superimposed are the outlines of the 5 major Mississippi sub-basins and their approximation in the reanalysis model by quadrilaterals. In clockwise sequence, basin 1 comprises the Arkansas-Red rivers, basin 2 the upper Missouri, basin 3 the upper Mississippi, basin 4 the Ohio, and 5 the lower Mississippi and Tennessee rivers (rather more poorly represented than the others, as the online integration scheme could handle only a few simple quadrilaterals). This online domain integration capability is a unique aspect of the ECMWF data assimilation system.

5.1 Model precipitation spin-up

The hourly archive includes both the short-term forecasts (hourly to 6-hr) used in the reanalysis cycle, and twice-daily 24-hr forecasts from 0000 and 1200 UTC (also archived hourly), so that issues relating to the diurnal cycle and the spin-up in the precipitation field can be addressed. We shall show model precipitation for the same verifying times from both the 0-6 hr analysis cycle (this we shall refer to as analysis precipitation), and from the 12-24 hr sections of the twice-daily 24 hr forecasts (which we shall refer to as 12-24FX precipitation). The model has a significant initial spin-up of precipitation from the analysis to the 12-24 hr forecast, and we shall find that the observed precipitation generally lies between these two model estimates in summer.

Table 2 summarizes this spin-up in the convective, large-scale and total precipitation in the model in terms of the ratio of the 12-24FX precipitation to the 0-6 hour analysis cycle precipitation for the annual average precipitation. The spin-up of large-scale precipitation is significantly larger (a mean of 1.38) than the

Table 2. Ratio of 12-24FX precipitation to analysis precipitation: model spin-up.

Basin	Convective precipitation	Large-scale precipitation	Total precipitation
Arkansas-Red	1.18	1.39	1.29
Missouri	1.22	1.50	1.40
Upper Mississippi	1.17	1.38	1.30
Ohio	1.24	1.31	1.28
Lower Mississippi	1.15	1.32	1.24
MEAN	**1.19**	**1.38**	**1.30**

convection spin-up (1.19), so that the spin-up of the total precipitation is 1.30 in the mean. The spin-up is largest for the most north-western and driest basin, the Missouri, and least for the lower Mississippi, basin 5. In the cool season, most model precipitation is large-scale. In contrast, in the warm season, convective and

large-scale precipitation is comparable for all basins. Consequently the spin-up of the model precipitation is greater in winter than in summer.

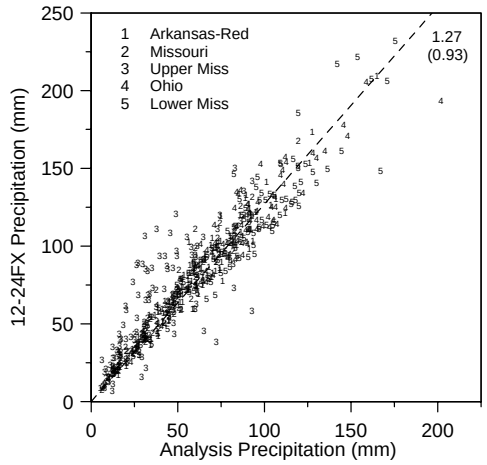


Fig. 12 Monthly 12-24FX precipitation against analysis precipitation

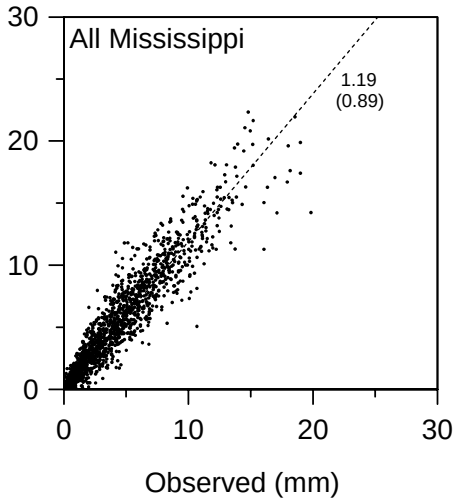


Fig. 13 Two-day ECMWF precipitation against *Higgins et al., (1996) precipitation*

Figure 12 shows the scatter of the monthly 12-24FX total precipitation against the analysis precipitation. The dashed line is the regression fit through the origin to all the data from all basins. It has a slope of 1.27 and an R^2 coefficient of 0.93. The scatter appears larger for the upper Mississippi basin 3.

5.2 Comparison of two-day precipitation with observations.

Figure 13 shows the 2-day total 12-24 FX precipitation from the ECMWF reanalysis against the basin average for the whole Mississippi basin from the *Higgins et al.* (1996) gridded data. The regression line through the origin is shown dashed, together with their slope and R^2 correlation coefficient in parentheses. For the whole Mississippi basin the correlation coefficient is high (0.89) and the standard error is quite small, only 1.4mm for this regression of 2-day data. There are some differences among the basins (not shown here: see *Betts et al.*, 1999). The drier Missouri basin, which has the largest spin-up (1.40: see Table 2), has the largest slope (1.35), and the wetter lower Mississippi, which has the smallest spin-up (1.24) has the smallest slope (0.93).

5.3 Precipitation on the diurnal timescale

Figure 13 shows that on the two-day timescale the model and observed precipitation agree quite well, but on the diurnal timescale the ECMWF summer precipitation matches the observations poorly, because the model has an erroneous convective precipitation maximum near local noon. Figure 14 shows that this is true for all basins in the warm season, defined here as March to September. On the left is the diurnal cycle of the model 12-24FX precipitation (we have repeated the 0.5 UTC value at 24.5 UTC for convenience). All basins have a peak near 1800 UTC (close to local noon), exactly the time when the observed precipitation is near a minimum. The observed warm season diurnal cycles, from the Higgins data shown on the right are quite different. Most basins have a late afternoon convective precipitation peak as well as peak near midnight (0600 UTC), and some basins have another peak near sunrise (*Higgins et al.*, 1996), none of which are reproduced

by the model. We believe this model convection peak near local noon is linked to an erroneous late-morning

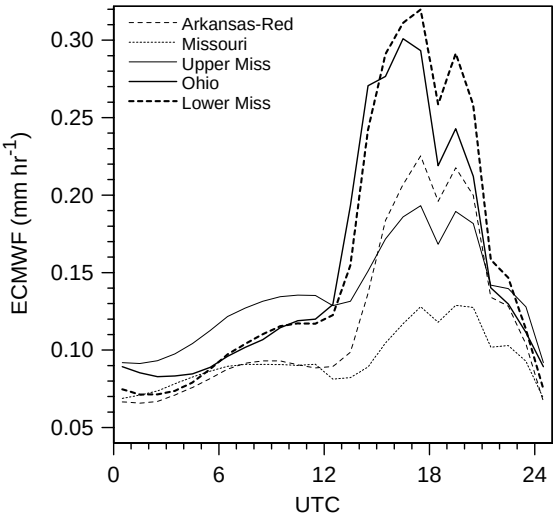


Fig. 14a Diurnal cycle of warm season precipitation for ECMWF model.

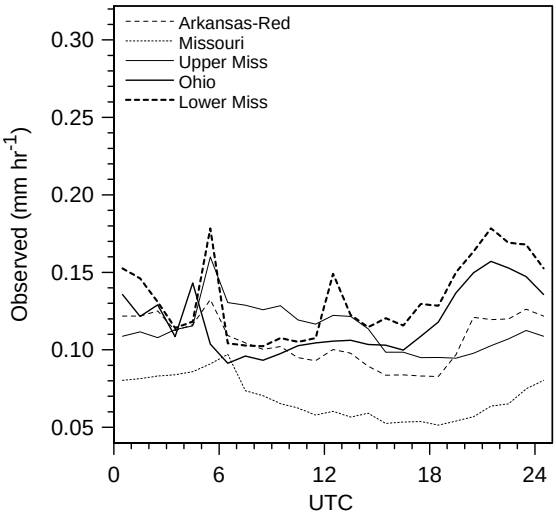


Fig. 14b As Fig 14a for *Higgins et al.* precipitation.

maximum (1600-1700 UTC) in the diurnal cycle of boundary layer mixing ratio, shown in Figure 10.. Despite this large diurnal error in the model, on time-scales longer than a day, the bias between the model 12-24FX precipitation and the observations is quite small for all basins (eg. Figure 13). In the cold season, defined as October to March, when large-scale precipitation dominates, the diurnal cycle in the model agrees much better with the observations, although the model does not capture the sharp precipitation maxima observed near local midnight (not shown: see *Betts et al.*, 1999)

5.4 Runoff and Stream flow

Model runoff is an important component of the surface model hydrology. Inadequate parameterizations for surface and deep runoff must be compensated by other model processes or (in the ECMWF model) by soil water nudging. Figure 15 compares the seasonal cycle of the nine-year average model runoff with the corresponding observed stream flow (both in mm month^{-1}) for the Arkansas-Red rivers, the upper Missouri,

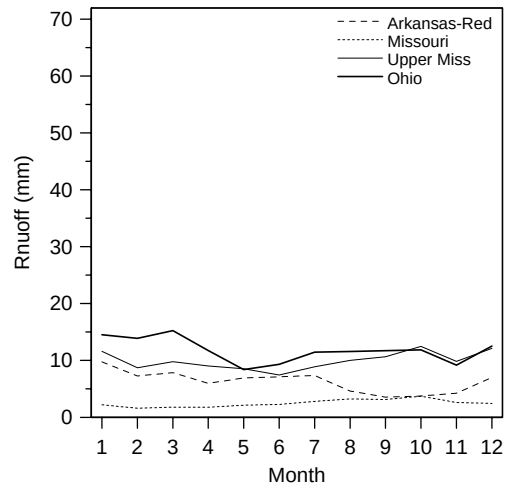


Fig.15a Monthly averaged ECMWF runoff (1985-1993)

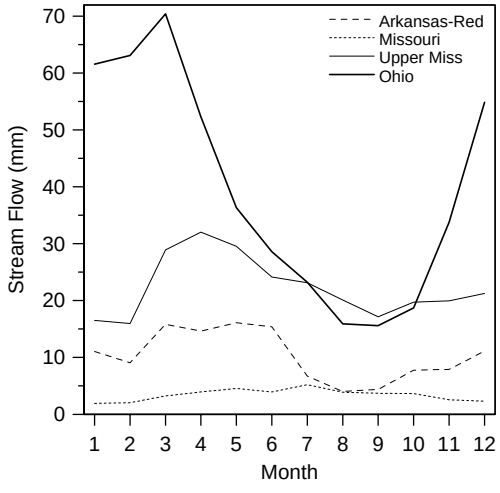


Fig. 15b As Fig. 15a for observed streamflow.

the upper Mississippi, and the Ohio. Although the model runoff does increase from the dry Missouri basin to the Ohio, the model runoff has no Spring or Fall runoff peaks, so that on an annual basis the model runoff is only a third to a half of the observed stream flow. This is a clear model deficiency, that is related to the fact that the ERA-15 model runoff is all drainage from the deep soil layer. This deep runoff increases exponentially as the soil water in the model base layer (100-289cm) approaches a threshold value. The surface runoff model is hardly ever activated, because of an inadequate representation of sub-grid-scale precipitation. The model has no routing scheme for runoff, but runoff does lag surface precipitation by a few weeks, because of the time it takes for the soil water to drain through to the base model layer. Despite this large underestimate of model runoff, Roads and Betts (2000) show that the ERA-15 runoff fluctuations are well correlated with streamflow fluctuations on the monthly timescale (not shown here). This lack of model runoff is most significant for the Ohio basin, which has the largest cool season stream flow. The error is partly compensated in the model by an increase in the removal of soil water by the nudging scheme during Spring in the analysis cycle.

5.5 MacKenzie basin comparisons

Another similar study (*Betts and Viterbo*, 2000) analyzes the hydrology and surface energy balance of seven sub-basins of the Mackenzie River in western Canada, which flows into the Arctic (and is the focus of the MAGS project). For this basin observations are sparser than for the Mississippi, but they are sufficient to indicate model biases at this high latitude and give some insight into the model frozen hydrology. Many of the high latitude seasonal processes discussed in section 3 are relevant, as most of the basin is boreal forest. We will just show a few illustrations. The model data we show is from the operational ECMWF model in 1996-1997, which had two significant updates in the land-surface scheme from the ERA-15 model. The first was the change to the stable boundary layer (*Viterbo et al.*, 1997: briefly discussed in section 4), which was introduced together with soil freezing in Sptember 1996. The second was the reduction of the albedo of snow under the boreal forests (*Viterbo and Betts*, 1999), which had a large impact on the surface energy balance.

5.5.1 Monthly precipitation comparisons.

Figure 16a compares precipitation from basin averages of the11-35 hour forecasts with corrected monthly observations (from *Hogg et al*, 1996) for the Mackenzie and 6 sub-basins for the 16 months, September, 1996 -December, 1997. (Corrected observations for 1998 are not yet available.) The upper panel for the whole Mackenzie shows that the model has considerably more precipitation than the observations, consistent with the climatological comparison. The ratio of model to observed precipitation (dotted) peaks at 2.5 in April, 1997 (when the snow evaporation peaks in the model), and averages about 1.4 for the rest of 1997. The overestimate of precipitation is consistent with the model having too much evaporation in spring and summer (*Betts et al.*, 1998b) over the boreal forest, and there being significant evaporation-precipitation feedback. During 1996, before the reduction in the model forest albedo with snow in December (which affected the surface energy balance), there is some indication that the model precipitation bias was smaller. This is consistent with the comparison of the precipitation in ECMWF reanalysis over the Mackenzie river basin with rain gauge based climatologies in *Stendel and Arpe* (1997).

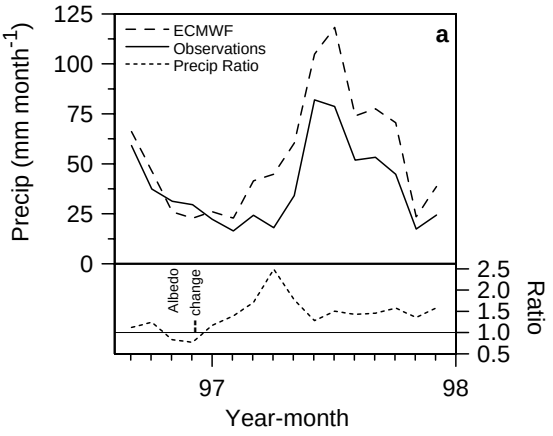


Fig. 16a Comparison of ECMWF precipitation with MacKenzie observations and ratio of two.

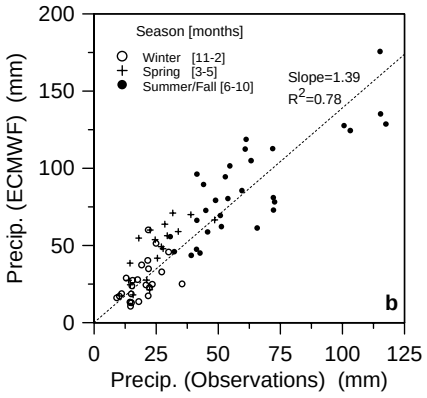


Fig. 16b Scatterplot of monthly ECMWF precipitation against observations for the sub-basins, by season.

Figure 16b is a scatter plot of the monthly basin precipitation ECMWF versus observations for 1997, grouping the data into winter, spring and summer/fall. The linear regression line has a slope of 1.39 ±0.04 with an R^2 coefficient of 0.78. In general, we could not distinguish any significant difference in bias between basins, or between winter or summer with this one year comparison. However, the spring points (for March, April,May) (except for the colder northern basins 1and 2 in March) lie on or above the regression line,confirming that the larger model bias in spring was seen in all basins. Compared to the Mississippi basins with the ERA-15 model, the model precipitation bias in 1997 appears to be larger for the MacKenzie (nearly 40%).

5.5.2 Basin comparison of model runoff and evaporation with annual stream flow and estimated evaporation

How does runoff compare with streamflow for this northern basin? A detailed monthly comparison between the ECMWF runoff and the observed stream flow is not worthwhile, as the present model has only drainage from the lowest model layer and no river routing scheme. At these latitudes the spring runoff peak is large as snow melts. (The new ERA-40 model under development has improved surface runoff with snow melt in Spring.) Figure 17a shows annual model runoff and observed streamflow (both plotted positive) for the "water year", September 1996-August 1997. Even though for the Mackenzie as a whole (point M), model runoff and observed stream flow are close(unlike the Mississippi), it is clear that the variability in model runoff on an annual basis across the basins bears no relationship to the stream flow differences.

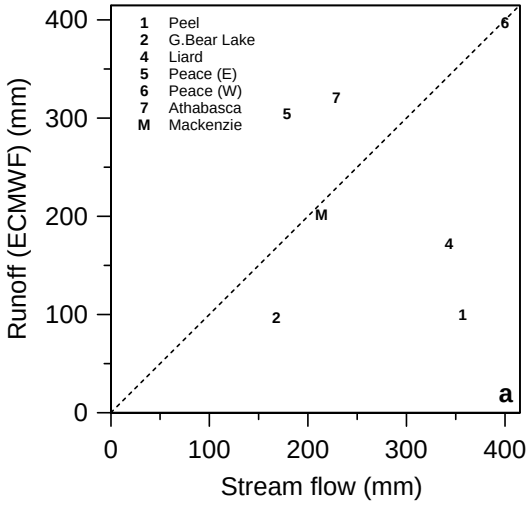


Fig.17a ECMWF runoff versus observed streamflow for MacKenzie and sub-basins for 1997 water year.

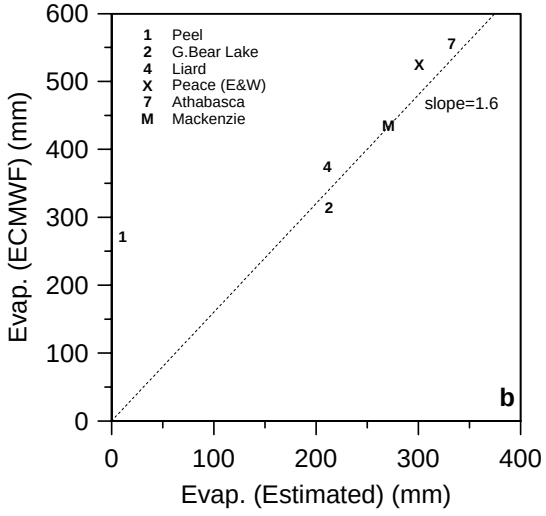


Fig. 17b Comparison of ECMWF annual evaporation with estimate from observed precipitation and streamflow.

On an annual basis, we can estimate basin evaporation from the observed precipitation and stream flow data as

$$\text{Evap.(Estimated)} = \text{Precip}_{obs} - \text{Streamflow}_{obs} \quad (1)$$

by ignoring basin-scale storage changes of water. Figure 17b compares Evap.(Estimated) with the annual model total evaporation

$$\text{Evap.(ECMWF)} = \text{Evap}_{liq} + \text{Evap}_{snow} \qquad (2)$$

The dotted line has a slope of 1.6, showing that on an annual basis, the model evaporation exceeds the observed estimate from (2) by about 60%. For basin 1, which lies far off the line, it is clear that the evaporation estimate from (2) is too low, perhaps because the catchment area for the stream flow estimate is smaller (and more mountainous) than the total area of basin 1 (which includes the Mackenzie delta), for which we have mean precipitation. As mentioned above, this is consistent with grid-point comparisons which show the model having too much evaporation in spring and summer (*Betts et al.,* 1998b) over the boreal forest.

6. CONCLUSIONS

In this brief review we have shown examples of model land-surface validation from the point scale to the scale of river basins. Only on the larger scale, can some aspects of the model surface hydrology and its interactive with the precipitation physics and model spin-up be assessed, but grid-point time-series comparisons are essential on a range of time-scales to understand how well the model parameterizations represent the observed physical processes. Data does not exist on all spatial scales, and all data has error signatures, some well known and some not, so care must be taken to distinguish model biases from uncertainties in the data or its representivity. Given however the still relatively primitive representation of land-surface physical processes in our numerical forecast models, the first test is to ask whether the processes that can be seen in the data are represented in the model on the diurnal and seasonal time-scales. The corollary of this is that complexity should not be added if it cannot be validated. The interactions between different parameterizations must be considered so that they are consistent with the known physical processes. The reader is referred to the references for further examples. There is a great deal of work still to be done to improve the land-surface physics in our earth system models.

Acknowledgments. Alan Betts acknowledges support from the National Science Foundation under Grant ATM-9505018, from NASA under Grant NAG5-7377, and from ECMWF for travel.

References

Barr, A.G. A.K. Betts, R. Desjardins and J. I. MacPherson, 1997: Comparison of regional surface fluxes from boundary-layer budgets and aircraft measurements above boreal forest. *J. Geophys. Res,* **102**, 29213-29218.

Barr, A.G. and A.K. Betts, 1997: Radiosonde Boundary-layer budgets above a Boreal forest. *J. Geophys. Res,* **102**, 29205-29212.

Betts A. K. and J. H. Ball, 1997: Albedo over the Boreal forest. *J. Geophys. Res.*, **102**, 28901-28910.

Betts A. K. and J. H. Ball, 1998: FIFE surface climate and site-average dataset:1987-1989. *J. Atmos Sci* , **55,** 1091-1108.

Betts A. K., P. Viterbo and A.C.M. Beljaars, 1998a: Comparison of the land-surface interaction in the ECMWF reanalysis model with the 1987 FIFE data.. *Mon. Wea. Rev.*, **126**, 186-198.

Betts A. K., P. Viterbo, A.C.M. Beljaars, H-L. Pan, S-Y. Hong, M. L.Goulden and S.C. Wofsy, 1998b: Evaluation of the land-surface interaction in the ECMWF and NCEP/NCAR reanalyses over grassland (FIFE)

and boreal forest (BOREAS). *J. Geophys. Res*., **103**, 23079-23085.

Betts, A. K., P. Viterbo and E. Wood, 1998c: Surface Energy and water balance for the Arkansas-Red river basin from the ECMWF reanalysis. *J. Climate*, **11**, 2881-2897.

Betts, A. K., M. L. Goulden, and S.C. Wofsy, 1999a: Controls on evaporation in a boreal spruce forest. *J. Climate*, **12**, 1601-1618.

Betts, A. K., J.H. Ball and P. Viterbo, 1999b: Basin-scale Surface Water and Energy Budgets for the Mississippi from the ECMWF Reanalysis. *J. Geophys. Res.*, **104**, 19293-19306

Betts, A. K. and P. Viterbo, 2000: Hydrological budgets and surface energy balance of seven subbasins of the Mackenzie River from the ECMWF model. *J. Hydrometeorol*., **1**, 47-60.

Coughlan, M. and R. Avissar, The global energy and water cycle experiment (GEWEX) continental-scale international project (GCIP): An overview. *J. Geophys. Res*., **101**, 7251-7268, 1996.

Douville, H., P. Viterbo, J.-F. Mahfouf and A.C.M. Beljaars, 1999: Evaluation of optimal interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Wea. Rev.* (in press).

Goulden, M.L., B.C. Daube, S.-M. Fan, D.J. Sutton, A. Bazzaz, J.W. Munger, and S.C. Wofsy,1997: Physiological Responses of a Black Spruce Forest to Weather, *J. Geophys. Res*., **102**, 28987-28996.

Higgins, R. W., J. E. Janowiak, and Y.-P. Yao, 1996: A gridded precipitation data base for the United States (1963-1993). NCEP/ Climate Prediction Center ATLAS No 1, 47pp. Available from Climate Prediction Center, NCEP, NWS, Camp Springs, MD 20746.

Hogg. W.D., P.Y.T. Louie, A. Niitsoo, E. Milewska and B. Routledge, 1996: Gridded Water Balance Climatology for the Canadian Mackenzie Basin GEWEX Study Area. Proceedings of the Workshop on the Implementation of the Arctic Precipitation Data Archive at the Global Precipitation Climatology Centre. 10-12 July 1996, Offenbach, Germany. p. 47-50. [Available from W.D. Hogg, Atmospheric Environment Service, 4905 Dufferin St., Downsview, ON, M3H 5T4, Canada.]

Roads, J. and A. Betts, 2000: NCEP/NCAR and ECMWF Reanalysis Surface Water and Energy Budgets for the GCIP Region. *J. Hydrometeorol.* , **1**, 88–94.

Sellers, P., F.G. Hall, G. Asrar, D.E. Strebel, and R.E. Murphy,1988:The First ISLSCP Field Experiment (FIFE), *Bull. Amer. Meteorol. Soc*., **69** (1), 22-27.

Sellers, P. and F.G. Hall, 1992: FIFE in 1992: Results, scientific gains, and future research directions, *J. Geophys. Res.*, **97**, 19,091-19,109.

Stendel, M. and K. Arpe, 1997: Evaluation of the hydrological cycle in reanalysis and observations. ECMWF Re-Analysis Project Report Series 6, 54 pp.

Strebel, D. E., D.R. Landis, K.F. Huemmrich, and B.W. Meeson, 1994: Collected data of the First ISLSCP Field Experiment, in *Surface Observations and Non-Image Data Sets*,. vol. 1, CD-ROM, NASA Goddard Space Flight Center, Greenbelt, MD 20771.

Van den Hurk, B.J.J.M., P. Viterbo, A.C.M. Beljaars and A. K. Betts, 2000: Offline validation of the ERA40 surface scheme. ECMWF Tech Memo, # 295.

Viterbo, P. and A.C.M. Beljaars, 1995. An improved land-surface parameterization in the ECMWF model and its validation. *J. Clim.*, **8**, 2716-2748.

Viterbo, P., and P. Courtier, 1995: The importance of soil water for medium-range weather forecasting. Implications for data assimilation, Workshop on Imbalance of Slowly Varying Components of Predictable Atmospheric Motions, World Meteorol. Org., Beijing, China. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, England]

Viterbo, P., A.C.M. Beljaars, J-F Mahfouf, and J. Teixeira, 1997: The role of soil water freezing and its interaction with the stable boundary layer. Research Activities in Atmospheric and Oceanic Modeling, ed. by A. Staniforth (pp 4.44-4.45), Rep. 25 of the CAS/JSC Working Group on Numerical Experimentation, WMO/TD-no. 792, Geneva.

Viterbo, P. and A.K. Betts, 1999: The impact on ECMWF forecasts of changes to the albedo of the boreal forests in the presence of snow. *J. Geophys. Res.*, **104**, 27 803-27 810.