**Geoscientific
Model Development**

Simulation of land surface temperatures: comparison of two climate models and satellite retrievals

J. M. Edwards

Met Office, Exeter, UK

Abstract. Recently there has been significant progress in the retrieval of land surface temperature from satellite observations. Satellite retrievals of surface temperature offer several advantages, including broad spatial coverage, and such data are potentially of great value in assessing general circulation models of the atmosphere. Here, retrievals of the land surface temperature over the contiguous United States are compared with simulations from two climate models. It is found that these two models generally simulate the diurnal range of surface temperature realistically, but show significant warm biases during the summer. The models' diurnal cycle of surface temperature is related to their surface flux budgets. Differences in the diurnal cycle of the surface flux budget between the models are found to be more pronounced than those in the diurnal cycle of surface temperature.

1 Introduction

Near-surface air temperatures are of great interest in meteorology, where they are an important output of numerical weather prediction models, with especial relevance to the forecasting of energy consumption (e.g. Dempsey et al., 1998). They also constitute one of the principal indicators of climate change (Solomon et al., 2007). Consequently, there is a continuing need to assess and improve the representation of near-surface air temperatures in numerical weather forecasting and climate models.

Recently, increasing attention has been given to the retrieval of land surface temperature from satellites (e.g. Sun and Pinker, 2003; Inamdar et al., 2008), particularly with a view to characterising the diurnal range of temperature. Satellite data offer a number of potential advantages for this

purpose including homogeneous spatial coverage and good temporal resolution of the diurnal cycle. Sun and Pinker (2003) and Sun et al. (2006b) produced retrievals of land surface temperatures from geostationary satellites over the contiguous United States under clear-sky conditions: they estimated that the bias of their retrievals was less than 1 K with RMS errors of about 1–2 K. Sun et al. (2006b) used these retrievals to characterize the seasonal variation of the diurnal temperature range across the United States. Inamdar et al. (2008) proposed an algorithm using data from geostationary and polar-orbiting satellites to retrieve land surface temperatures at a resolution of 1 km. They validated their retrievals against surface observations at two sites in the United States and demonstrated an accuracy generally better than 2 K throughout the diurnal cycle.

It is natural to ask what value such data might have in the assessment and improvement of atmospheric models. From this perspective the land surface temperature is a more fundamental quantity than the near-surface air temperature, since it is directly linked to the surface flux budget. By contrast, the relationship between the surface (skin) temperature and the near-surface air temperature is not fully understood. Large temperature gradients often exist just above the surface, which are characterized in models using a thermal roughness length; however, thermal roughness lengths are highly uncertain (Mahrt and Vickers, 2004; Sun, 1999). In very stable conditions, the surface may become decoupled from the atmosphere (Coulter and Doran, 2002): decoupling is not well represented in models, which typically employ parameterizations with enhanced mixing in stable conditions (Cuxart et al., 2006) that may therefore tie the surface and near-surface air temperatures too closely together. Additionally, questions have been raised about our understanding of surface transfer in strongly convective conditions (Zilitinkevich et al., 2006). Thus, an assessment of the simulation of land surface temperatures in atmospheric models may offer a complementary perspective to that obtained from assessing

Table 1. Sources of data.

Source of data	Resolution (EW×NS)	Period
GOES-8 satellite	0.5°×0.5°	1996–2000
GFDL-AM2D	2.5°×2.0°	1983–1998
HadGEM2-A	1.875°×1.25°	1983–1998

near-surface air temperatures, with additional benefits coming from the homogeneous spatial coverage and good temporal resolution available from satellites. Indeed, in a study of mesoscale flows over the island of Majorca, Jimenez et al. (2008) showed that satellite surface temperatures were substantially less affected by local effects than traditional point observations.

Sun et al. (2006b) have made their data for the period 1996–2000 available on 1/2 and 1/8 degree grids at http://www.meto.umd.edu/~srb/gcip and the aim of the present study is to use these data assessing the diurnal cycle of the land-surface temperature and the surface flux budget in the climate models of the Geophysical Fluid Dynamics Laboratory (GFDL) and the Met Office (UK). In the following section the data and climate models are introduced, after which the processing of the model data is discussed. Section 3 is concerned with assessing the extremal temperatures and diurnal ranges of the model data across the United States. In Sect. 4 diurnal and seasonal composites for three distinct geographical regions are considered and links are made to the surface flux budget. Conclusions are presented in the final section.

2   Satellite and model data

2.1   Satellite data

The retrieval of land surface temperature from satellite observations is discussed by Sun and Pinker (2003) and Sun et al. (2006b). The algorithm makes use of the 10.8 and 12.0 micron channels on the GOES-8 satellite. Pixels are classified as clear if the retrieved cloud fraction is less than 10% and the skin temperature in pixels which fail to meet this criterion is set as missing data. The final product consists of 5 years of hourly values of the surface skin temperature covering the region 25 to 50 N and 125 to 70 W. Data are available on either a 1/2 degree or a 1/8 degree grid. For the present purpose, data on a 1/2 degree grid are selected since the resolution of both models is coarser than this. For comparison with the model data the salient characteristics of the data are listed in Table 1.

To enable comparison with data from climate models, the satellite data have been used to construct monthly composites of 24 hourly values for each 1/2-degree grid-square. Although 5 years of data are available, they are not sufficient

to establish complete coverage of the whole land area of the United States throughout the diurnal cycle in all months. The data are, however, sufficient to characterise most of the diurnal cycle and in particular to establish extrema at each point, being present for at least one hour around the time of the extremal value.

2.2   Models and model data

Two climate models are employed in this study:

1. The GFDL climate model: The GFDL climate model is described by Global Atmospheric Model Development Team (2004) (hereafter GAMDT). The data used here were obtained from a configuration based on the standard AM2 version, but with the deep convection scheme of Donner (1993), implemented as described by Wilcox and Donner (2007), and accompanied by a retuning of the cloud scheme to balance the radiation budget at the top of the atmosphere. This model will be referred to as AM2D.

2. The Met Office Unified Model[®]: The version of this model used in the present study is the recently defined HadGEM2-A atmospheric configuration, which is a further development of the configuration HadGEM1-A, described by Martin et al. (2006).

Each model was integrated through the period 1983–1998, with instantaneous diagnostics for an area covering the contiguous United States being produced at hourly intervals, in the case of AM2D, and half-hourly intervals in the case of HadGEM2-A. The diagnostics comprise, principally, the surface temperature and cloud fields to enable masking of the data in a way that is consistent with the satellite product, together with the surface fluxes. The principal characteristics of the data are presented in Table 1.

These instantaneous data are subsequently formed into monthly composites for the analysis presented below. Composites are formed with and without cloud masking to enable comparison with the retrievals in clear-sky conditions and also to permit comparisons between the models in all-sky conditions. Clear-sky composites of modelled data are formed by applying a cloud clearing threshold of 10% (also rejecting foggy grid-boxes). The sampling procedure is applied independently to each model and is completely independent of the satellite data. It is then expected that the resulting climatological composites will be statistically comparable with the satellite data. This assumption is now investigated by considering the various sources of error in the data and the degree of consistency which should be expected between the retrieved and modelled data.

2.3 Consistency between the retrievals and model data

The estimated bias in the retrievals is less than 1 K (Sun et al., 2006b). Since only five years of retrievals are available, giving a maximum potential number of samples for each hour of the diurnal cycle of about 150, the resulting composites will be subject to a measure of statistical error. From the standard deviation of the retrieved clear-sky surface temperatures (around 2–3 K across the eastern US in July and around 5 K in the west), it is estimated that this error is less than 0.5 K, with a slightly lower estimate for the composited model data, which cover a longer period.

In the retrieval process a threshold of 10% is used for cloud-clearing and the same threshold is adopted for cloud-clearing of the model data. Underlying this choice is the assumption that the surface temperature is determined locally by the surface fluxes and that the cloud fraction serves a proxy for the radiative flux at the surface. (One might, however, expect some dependence of the appropriate threshold on resolution on the grounds that imposing a fixed cloud threshold on coarser grid-boxes represents an increase in the stringency of the criterion.)

To assess the sensitivity of the clear-sky surface temperature to the choice of cloud-clearing, composites of model data diagnosed with different thresholds are briefly considered. The land surface temperature would be expected to be most sensitive to the cloud fraction in conditions of high incident solar radiation. Fig. 1a shows the change in the diagnosed clear-sky land surface temperature in HadGEM2-A at 18:00 UTC (close to local noon across the eastern US) in July resulting from a reduction in the tolerance for cloud clearing from 10 to 5%. With a tolerance of 5% only the clearest grid-boxes are used to define the temperature and since clear skies will be associated with warm temperatures at this time, an increase in the diagnosed clear-sky temperature is to be expected. Across the eastern US the diagnosed clear-sky surface temperature does show an increase, generally not exceeding 0.5 K.

Figure 1b shows the variation of the average diagnosed clear-sky surface temperature across the eastern US (here taken as 32–40 N and 100–77.5 W) with the tolerance for cloud-clearing at 12:00 UTC and 18:00 UTC in July. For each model, composites of the clear-sky surface temperatures at these times were formed from the data for July using cloud-clearing thresholds of 5, 10, 15, 20 and 25%. The figure shows the differences in the average diagnosed clear-sky surface temperature relative to that obtained with the standard tolerance of 10%. There is little sensitivity to the threshold in the data for 12:00 UTC, corresponding to conditions early in the morning, in either model but HadGEM2-A shows more sensitivity around 18:00 UTC. The diagnosed temperature increases as the threshold is decreased, as would be expected around local noon. (Although the reasons for the differences in sensitivity between the models have not been explored in detail, it seems likely that the greater sen-
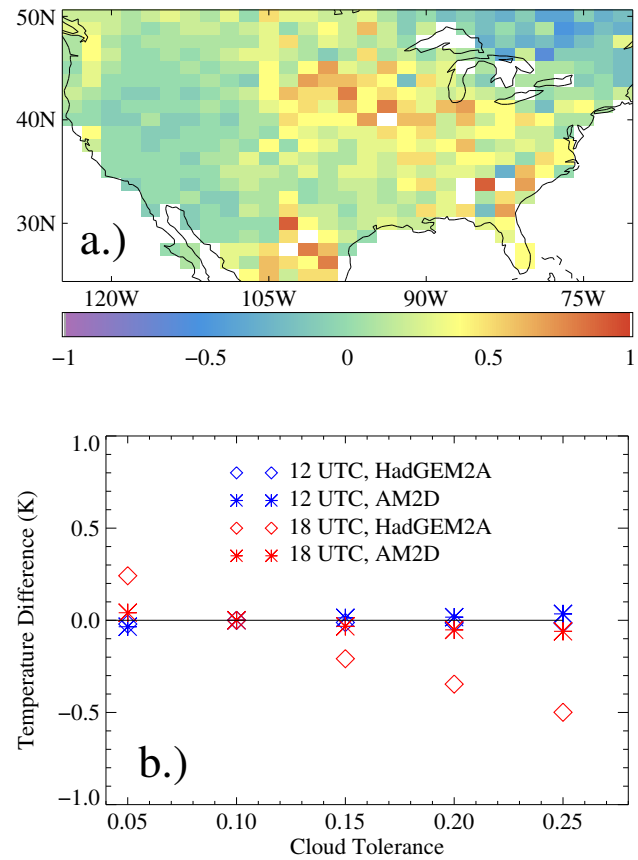

Fig. 1. The sensitivity of the diagnosed clear-sky land surface temperature to the tolerance for cloud-clearing. **(a)** The change in the composite diagnosed clear-sky surface temperature (K) at 18:00 UTC in July resulting from a reduction of the tolerance from 10% to 5% in HadGEM2-A. **(b)** The sensitivity of the diagnosed clear-sky surface temperatures (K) across the eastern US to the tolerance for cloud clearing at 12:00 and 18:00 UTC in both HadGEM2-A and AM2D. See text for details.

sitivity of HadGEM2-A results from the rather close link between cloud amount and cloud water in this model, both for large-scale cloud (Smith, 1990) and convective cloud. By contrast, AM2D incorporates a version of the scheme of Tiedtke (1993) to treat both large-scale and convective cloud, in which cloud amount and cloud condensate are determined from separate prognostic equations.) Extrapolating these data suggests that for a threshold of 10% the diagnosed land surface temperature may be about 0.5 K colder than that which would be obtained for perfect cloud clearing in HadGEM2-A, with a smaller bias for AM2D.

To assess the influence of the threshold on the retrieved data is harder, but a simple test involves forming composites subject to the criterion that each grid-box and its 8 surrounding boxes have cloud fractions below the threshold of 10%. For 18:00 UTC in July, the composite surface temperature defined in this way is 0.7 K warmer over the eastern United

States than that obtained by compositing with no reference to the neighbouring grid-boxes. This represents a fairly stringent cloud-clearing criterion and is tentatively interpreted as suggesting that the retrievals are probably less than 1 K cold relative to perfect cloud clearing.

Attention must also be given to the impact of variations of the real land surface temperature within the area represented by a model grid-box. It could be argued that if the model is to represent the surface flux budget correctly it is more important that its upward longwave flux should match the average observed flux than that its land surface temperature should match the observed mean over the area of the grid-box. Inamdar et al. (2008) discuss the variation of their 1-km data within a region 100×50 km, giving a standard deviation of the land surface temperature of 5.85 K for a mean temperature of 326.2 K. In this case, the difference between the mean land surface temperature and a mean inferred by averaging the upward longwave flux from each 1-km pixel, then extracting an equivalent temperature is less than 0.2 K. This suggests that variations on scales unresolved within the models will not significantly affect the following comparisons.

Since the retrieved data cover the period 1996–2000, while the model data cover the period 1983–1998, the impact of any climatic warming trend over recent decades should be considered. Globally, near-surface air temperatures have risen since the 1960s; but the warming has not been uniform. Trends in the near-surface air temperature across the United States during this period are discussed by Hansen et al. (2001), who conclude that anomalies in near surface air temperatures across the contiguous United States have not exceeded 0.5 K relative to the means from 1951–1980 during this period. This implies that the impact of any underlying warming trend is small relative to interannual variability and systematic errors in the models. This conclusion is reinforced by the analysis of Karl et al. (1993), who considered trends in diurnal extrema and temperature ranges over the US and found that the magnitudes of these trends did not exceed 0.3 K per decade; they suggested that the trends were most likely associated with changes in cloud clover.

3 The diurnal temperature range

Figure 2 shows the diurnal range of clear-sky surface temperature for January derived from composites of the satellite data (a) and as simulated by HadGEM2-A (b) and AM2D (c). The models agree with the satellite data in reproducing the pattern noted by Sun et al. (2006b), with higher diurnal ranges over the mountainous and desert areas to the west than over the eastern areas of the country. Towards the southern edge of the domain, over Mexico, the observed diurnal range exceeds 30 K, with both models reproducing this quite closely. Over the south-eastern United States, both models show a realistic diurnal range of just over 15 K. Towards the Canadian border the diurnal range is smaller in both the re-
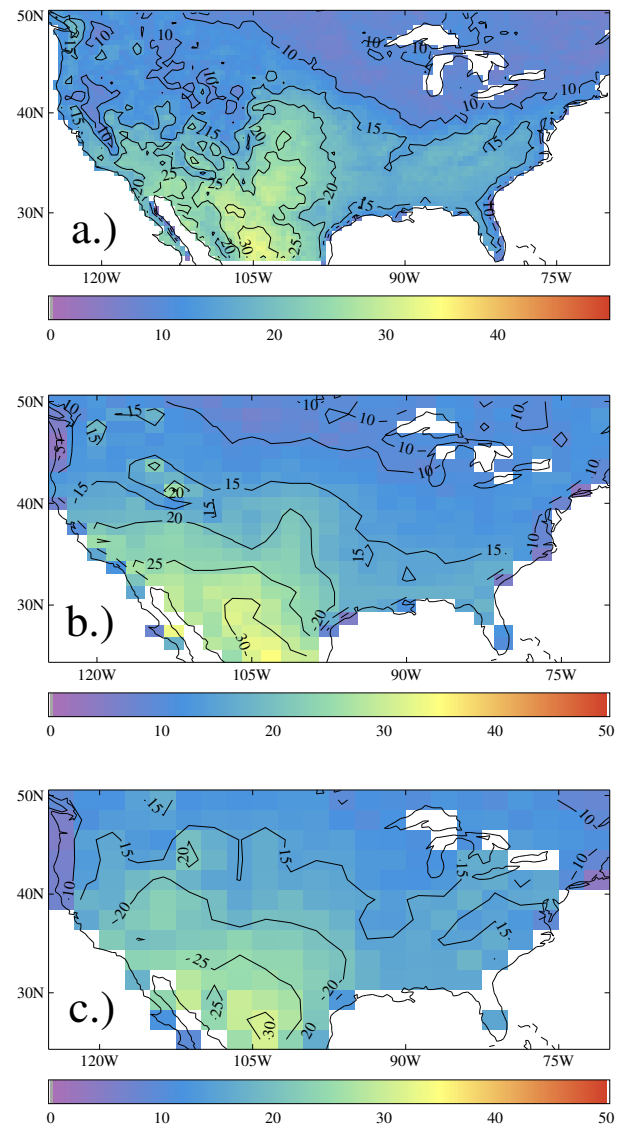

Fig. 2. Diurnal range of land surface temperature (K) under clear-skies in January. **(a)** retrieval. **(b)** HadGEM2-A. **(c)** AM2D.

trievals and the models, although here the models show a slight overestimate. Nevertheless, the good simulation of the diurnal range masks biases in the maxima and minima. Figure 3 shows the differences (model-retrieval) between the clear-sky maximum and clear-sky minimum temperatures for January. Both models show similar patterns of difference from the retrievals, with cold biases to the west and over the south-east and a warm bias over the northern Great Plains. The warm biases over the northern Great Plains are slightly more pronounced in AM2D, particularly in the maximum, leading to an overestimate of the diurnal range. It is interesting to relate these differences to biases in the near-surface air-temperature in the two models. Figure 3 of the paper of GAMDT (2004) shows that the underlying AM2 model has
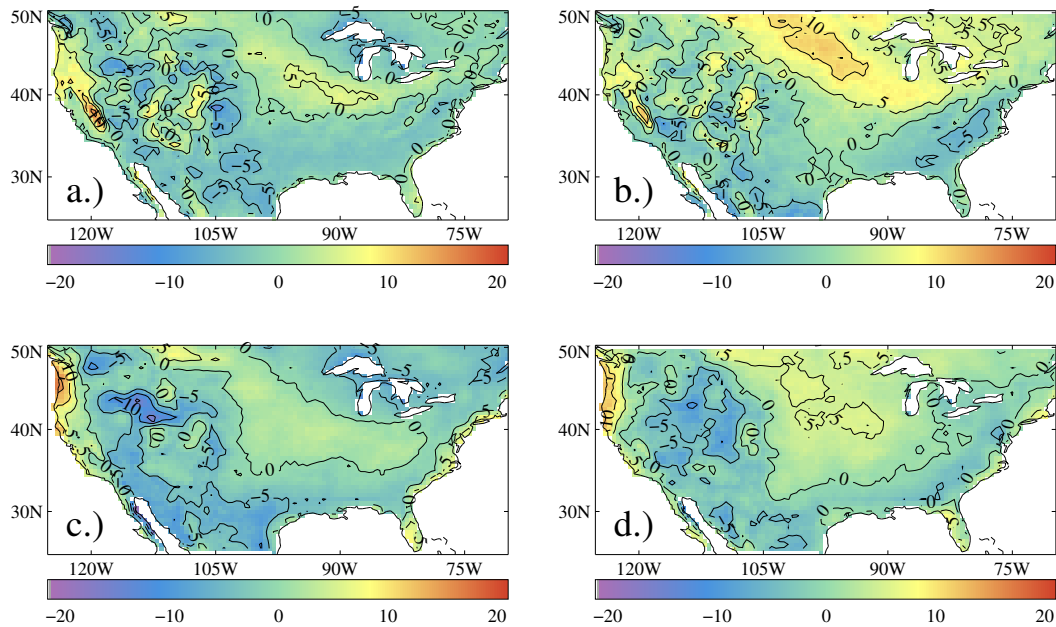
Fig. 3. Differences in maximum and minimum surface temperatures (K) (model-retrieval) under clear-skies in January. **(a)** Differences in maximum temperature (HadGEM2-A-retrieval). **(b)** Differences in maximum temperature (AM2D-retrieval). **(c)** Differences in minimum temperature (HadGEM2-A-retrieval). **(d)** Differences in minimum temperature (AM2D-retrieval).

a warm bias in the mean air temperature at 2 m in this region in the winter; by contrast HadGEM2-A has a small cold bias. However, since there may be significant temperature gradients close to the surface, the models' biases in the land surface and near-surface air temperatures will in general not necessarily be the same.

Figure 4 shows the retrieved and simulated diurnal ranges for July. Here, the models agree less well with the retrievals: both significantly overestimate the diurnal ranges across the country, giving ranges in excess of 45 K in the west and of around 20 K in the east. The differences between the modelled and retrieved diurnal maxima and minima in Fig. 5 show that the high diurnal ranges in the Mississippi valley in both models are due to moderate overestimates of the minimum temperature, accompanied by more pronounced overestimates of the maximum temperature. Towards the west, the signal is more complicated, but both models tend to overestimate the maxima and underestimate the minima. The spatial patterns of these errors are broadly similar to the errors in the near surface temperatures in both models, although the magnitudes of the errors in land surface temperature are typically larger than those in near-surface air temperature.

To provide an indication of the relationship between the biases in land surface and near-surface air temperatures, Fig. 6 shows the diurnal maxima and minima of land surface and near-surface air temperature for the eastern United States (here 100–85 W, 32–40 N), where topographic variations are

not so extreme as to the west, for four months through the annual cycle.

Grey diamonds show the extrema for HadGEM2-A under all-sky conditions and grey triangles the equivalent minima for AM2D under these same conditions. Lines connect corresponding surface and near-surface points, but these are for ease of interpretation and do not represent atmospheric profiles. As expected, the models show a larger diurnal range in surface temperature than in near-surface air temperature; in July the diurnal range of the surface temperature is greater by 50%. Grey crosses show the extrema of near-surface air temperature taken from the CRU climatology (New et al., 1999). (This paper should also be consulted for discussion the accuracy of the climatology and comparisons with the possible alternatives.) Relative to the observations, both models show restricted diurnal ranges of near-surface air temperature in January and April, with warm minima and cool maxima. In July both models show significant warm biases in the maximum and the minimum, while in October the models show a modest warm bias in the minimum, but simulate the maximum quite well.

Blue diamonds and triangles show the models' surface and near-surface air temperatures under clear-sky conditions. As would be expected, clear-sky minima are lower than the corresponding all-sky minima, while the clear-sky maxima are higher than the all-sky maxima. The clear-sky surface temperatures retrieved from the satellite data are shown as blue crosses. In January the models simulate the clear-sky
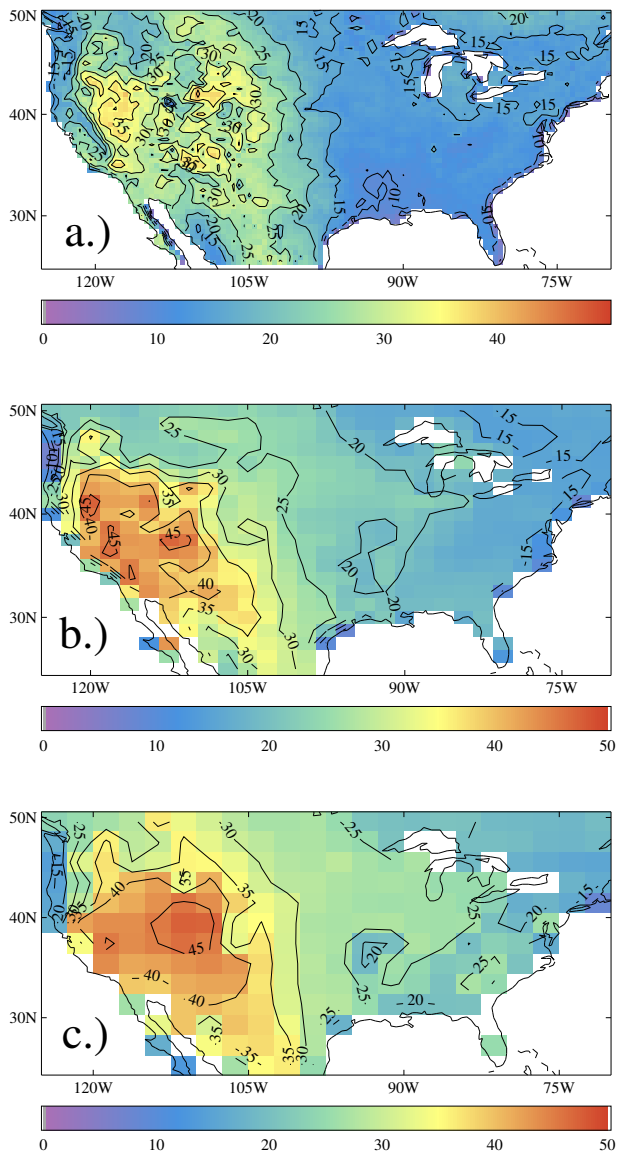
Fig. 4. Diurnal range of land surface temperature (K) under clear-skies in July. **(a)** retrieval. **(b)** HadGEM2-A. **(c)** AM2D.

surface temperatures well (c.f. Fig. 2), while in July they simulate significantly higher temperatures than the retrievals (c.f. Fig. 4). In the intervening months of April and October the surface temperatures are generally simulated well, although AM2D shows a warm maximum in October.

It is also of interest to compare the retrieved clear-sky surface temperatures with the observed all-sky near-surface air temperatures. The expectations that the clear-sky maxima at the surface should exceed the all-sky maxima in the near-surface air temperature and that the clear-sky minima at the surface should be lower than the all-sky minima in the near-surface air temperature are generally met. However, the clear-sky surface maximum in July is barely larger than the all-sky air temperature. Maps of the differences (not shown) indicate that over large areas of the south-eastern United States the clear-sky maximum land surface temperature in July is very slightly lower than the all-sky maximum air temperature, with the maximum deficit being 2 K. Whilst the reason for this is not fully understood, it may be due to the impact on the retrieved land surface temperature of the higher water vapour content of the atmosphere in this month (Sun et al., 2006a). The expectation also fails in winter over the Rocky Mountains and the northern States, where the clear-sky maximum land surface temperatures can be several degrees lower than the all-sky maximum temperature. This is believed to be an effect of snow cover, where, because of the high albedo, there is limited shortwave heating of the surface during the day in clear-sky conditions, while cloudy conditions are generally associated with warmer temperatures. But the areas where this effect occurs lie largely to the north and west of the region considered here so it is not apparent in Fig. 6.

4   The composite diurnal cycle

More insight into the models' performance can be gained by linking the simulation of land surface temperature to the surface flux budget and considering the diurnal cycle of temperatures and fluxes on a regional basis. Figure 7 shows the three regions selected for this analysis: these comprise the lower Mississippi valley (region M, 100–85 W, 32–38 N), the northern Great Plains (region P, 105–92.5 W, 42–48 N), and the south-west (region S, 117.5–102.5 W, 34–40 N). For each region monthly diurnal composites of the retrieved surface temperature under clear-skies are shown in conjunction with diurnal composites of the models' surface temperatures under clear-sky and all-sky conditions. Data for the months of January, April, July and October are presented here. Since the retrievals do not provide complete coverage of the diurnal cycle for each month, retrieved values are shown only where the minimum fraction of valid points for each 1/2-degree box within a region exceeds 10%. These composites are then related to diurnal composites of the surface fluxes and to composites of the monthly mean surface fluxes and the hydrological cycle through the year. Because of the way in which cloud and radiation diagnostics are derived within the models, clear-sky surface temperatures are available from AM2D only at 3-hourly intervals, so intermediate values are filled in by spline interpolation, except around the minimum at dawn, where linear interpolation is preferred. In the case of HadGEM2-A shortwave surface fluxes are available only as means over 3-h intervals.
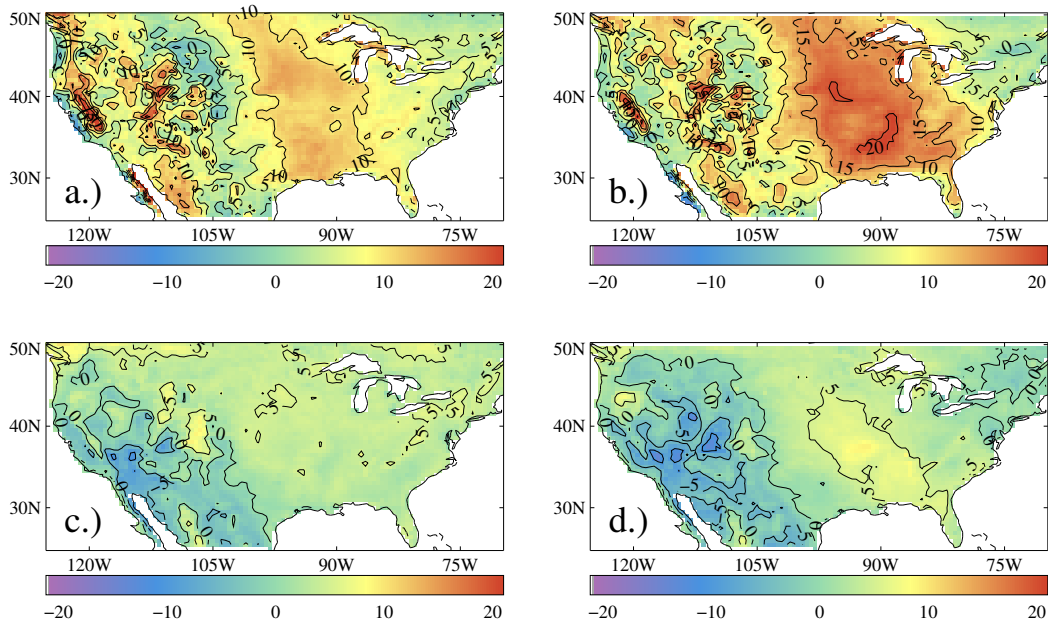
Fig. 5. Differences in maximum and minimum surface temperatures (K) (model-retrieval) under clear-skies in July. **(a)** Differences in maximum temperature (HadGEM2-A-retrieval). **(b)** Differences in maximum temperature (AM2D-retrieval). **(c)** Differences in minimum temperature (HadGEM2-A-retrieval). **(d)** Differences in minimum temperature (AM2D-retrieval).
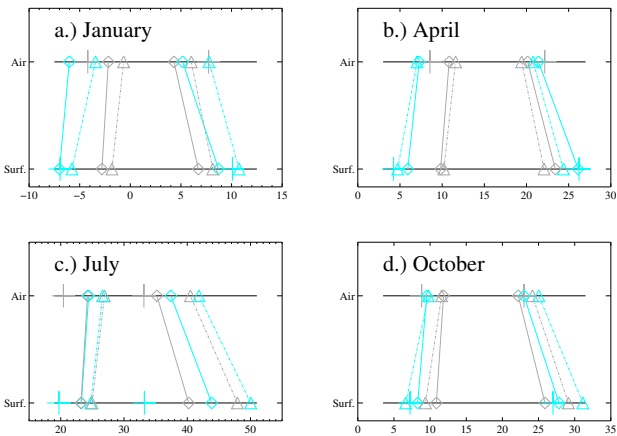

Fig. 6. Comparison of the maximum and minimum surface and near-surface air temperatures (Celsius) across the eastern United States for the months of **(a)** January, **(b)** April, **(c)** July and **(d)** October, from the two models and from observations. Crosses represent observations ("Air" from ground-based observations and "Surf." from satellite observations respectively), diamonds represent HadGEM2-A and triangles AM2D. Clear-sky quantities are shown in blue and all-sky quantities in grey. Lines connect corresponding model data for ease of interpretation, but do not represent profiles.
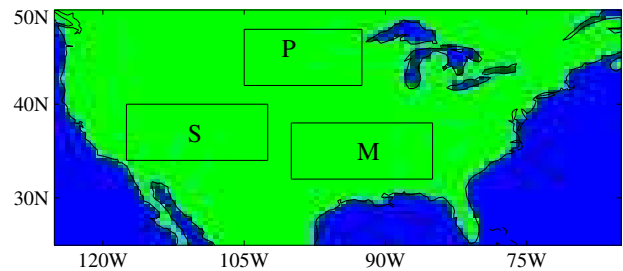

Fig. 7. The regions selected for composite analysis.

4.1 The lower Mississippi valley

Figure 8a shows the diurnal composites of the surface temperature over the lower Mississippi valley (region M) in January. For clarity of display of the full diurnal cycle, the data for the initial 6 h of the day are repeated on the right of each panel. Under clear-sky conditions the models agree well with the retrievals during the earlier part of the night, although HadGEM2-A has become slightly cold by dawn. Agreement during the day is less good, with both models underestimating the maximum surface temperature. Under all-sky conditions, the models are in broad agreement, though AM2D is slightly warmer than HadGEM2-A. Figure 8b shows the corresponding components of the surface flux budget. For consistency here we shall adopt the convention that upward fluxes are positive, but for clarity of display in the diurnal
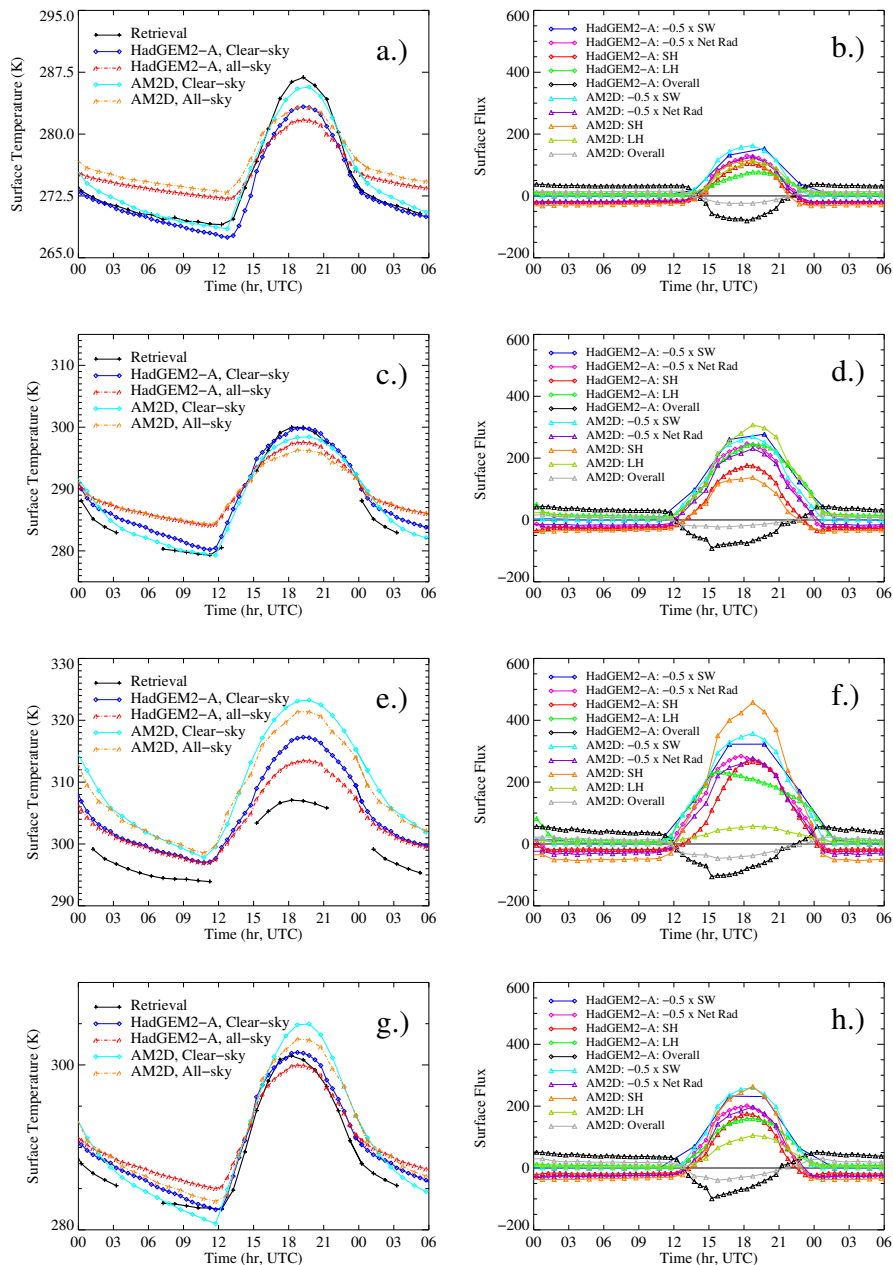
Fig. 8. The composite diurnal cycles of surface temperature over the lower Mississippi valley (region M) from retrievals and from HadGEM2-A and AM2D for the months of **(a)** January, **(c)** April, **(e)** July and **(g)** October. The right column shows the corresponding diurnal cycle of the surface fluxes for the months of **(b)** January, **(d)** April, **(f)** July and **(h)** October.

composites, the net shortwave (SW) and total net radiative fluxes are shown scaled by -0.5: this is indicated in the legend. The overall (atmospheric) flux, F_{atm}, is then the sum of the net radiative, latent and sensible fluxes just above the canopy. Explicitly,

$$F_{atm} = SW^{\uparrow} - SW^{\downarrow} + LW^{\uparrow} - LW^{\downarrow} + F_{sens} + F_{lat}, \quad (1)$$

where the terms on the right represent respectively the upward and downward shortwave fluxes, the upward and downward longwave fluxes and the sensible and latent heat fluxes. The most significant differences between the two models are that AM2D shows noticeably higher latent heat fluxes by day and slightly more negative sensible heat fluxes by night, thus mitigating nocturnal surface cooling due to longwave radiation. Overall, the amplitude of the diurnal cycle in

HadGEM2-A is more than twice that in AM2D, despite the similarity of the diurnal range of surface temperatures.

Moving on to April (Fig. 8c), the clear-sky maximum surface temperature from HadGEM2-A agrees closely with the retrieved value, while AM2D is slightly colder. However, HadGEM2-A is rather warm during the evening transition and the earlier part of the night, when AM2D agrees more closely with the retrieval. Under all-sky conditions the two models are generally similar, except that HadGEM2-A has a higher day-time maximum. Consistent with this difference in surface temperatures by day, the sensible heat flux is higher in HadGEM2-A, while the latent heat flux is higher in AM2D (Fig. 8d). Again, the amplitude of the diurnal cycle of the overall flux is significantly larger in HadGEM2-A, despite the fairly similar diurnal ranges of surface temperature.

Differences between the models and with the observations are more significant in July (Fig. 8e). Under clear skies the models are significantly warmer than the retrievals throughout the day and especially around local noon (compare Fig. 5). Although the modelled minimum temperatures under both clear-sky and all-sky conditions are not too dissimilar, AM2D is significantly warmer than HadGEM2-A by day. The latent heat fluxes are significantly different between the two models (Fig. 8f), but the position is reversed relative to that seen in April with the latent heat flux in AM2D now being much lower than that in HadGEM2-A. Correspondingly, AM2D shows much larger sensible heat loses than HadGEM2-A. It is interesting to note the close similarity of the net shortwave fluxes, showing that differences in the models are rather due to their treatment of the surface turbulent fluxes than to differences in the radiation budget due to clouds. Again, the overall surface flux shows a higher amplitude in HadGEM2-A, despite the smaller diurnal temperature range of this model.

In October (Fig. 8g) HadGEM2-A simulates the clear-sky maximum and minimum temperatures well, but does not cool sufficiently rapidly through the evening transition, resulting in a warm bias in the earlier part of the night. AM2D cools more rapidly through the evening transition, in better agreement with the retrievals at this time, but overestimates the amplitude of the diurnal cycle. Similarly, under all-sky conditions AM2D also shows a higher diurnal range of temperature. Once again, the amplitude of the diurnal cycle of the overall surface flux budget is significantly greater in HadGEM2-A (Fig. 8h).

It is useful to extend the comparison to monthly mean composites through the annual cycle. Figure 9a shows the monthly mean values of the components of the surface flux budget for region M. The most prominent differences occur during the summer months, where the latent heat flux in AM2D is much lower than that in HadGEM2-A, while the sensible heat flux is much larger. Figure 9b shows the hydrological budget for the region. The green line represents the monthly mean precipitation from the GPCP climatology (Huffman et al., 1997), with the precipitation from the two
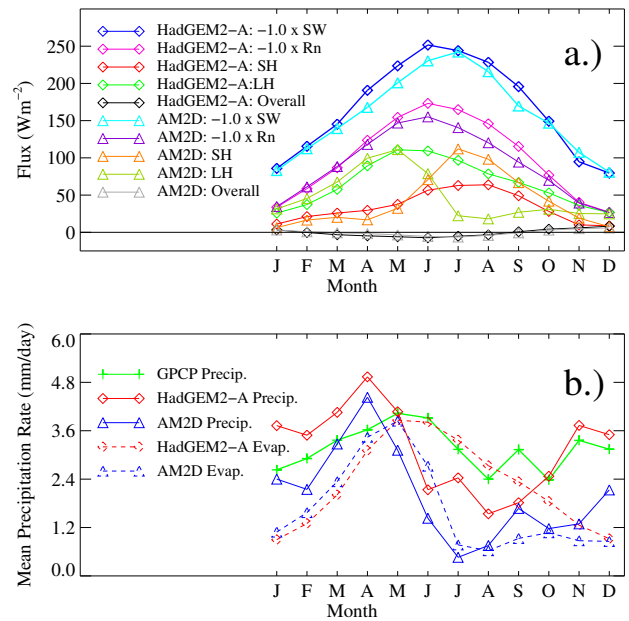

Fig. 9. The annual cycle of the monthly mean surface fluxes **(a)** and hydrological budget **(b)** over the lower Mississippi valley. In the legend Rn represents the net upward radiative flux.

models being shown as solid lines. Both models show a significant lack of precipitation during the summer and a tendency toward excessive precipitation in the spring, which is more pronounced in HadGEM2-A. (It may also be noted that comparison with the alternative CMAP climatology of precipitation (Xie and Arkin, 1997) shows the same features and that the two climatologies agree well over the US (Yin et al., 2004).) The dashed lines show the models' rates of evaporation. As would be expected, these lag the precipitation, but an interesting difference between the models is that the lag in AM2D is shorter than that in HadGEM2-A.

These results suggest that differences in the land surface schemes of the two models probably play an important role in explaining the differences in behaviour between them. AM2D receives less precipitation in the winter and spring and is more realistic here, but it re-evaporates that water more readily (explaining the higher latent heat fluxes during winter and spring, despite the lower precipitation). By summer, AM2D seems to have lost a significant amount of soil moisture causing its surface temperature to rise above that of HadGEM2-A. In addition, the smaller amplitude of the diurnal cycle of the surface flux budget in AM2D, despite broadly similar diurnal temperature ranges in the two models, suggests that the thermal inertia of the soil is smaller than that of HadGEM2-A. In large part, the smaller amplitude of the diurnal cycle of the overall surface flux budget in AM2D results from larger turbulent loses from the surface by day and increased turbulent warming by night (more
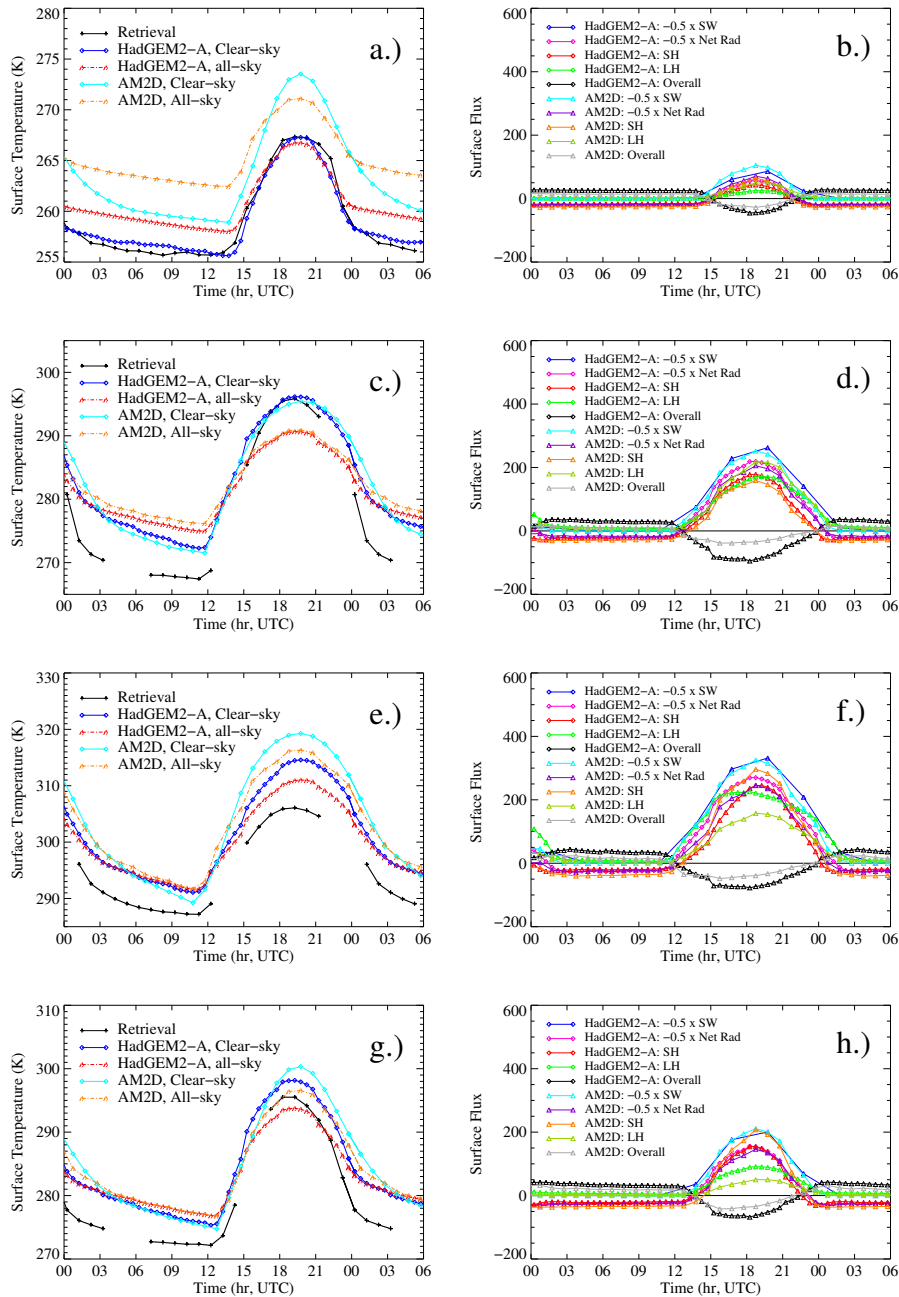
Fig. 10. As Fig. 8, but for the northern Great Plains (region P).

negative sensible heat fluxes) relative to HadGEM2-A. For each of the months shown in Fig. 8, except April, the difference in the sensible heat fluxes during the later part of the night (05:00 UTC–10:00 UTC) between the two models is larger in absolute terms than the difference in the net radiative fluxes.

4.2 The northern Great Plains

Figure 10 shows corresponding diurnal composites of surface temperatures and fluxes for the northern Great Plains (region P). Many features of this figure are similar to those seen in Fig. 8, particularly in regard to the warm bias in the surface temperatures of both models in the summer, the greater amplitude of the diurnal cycle of the overall surface flux in
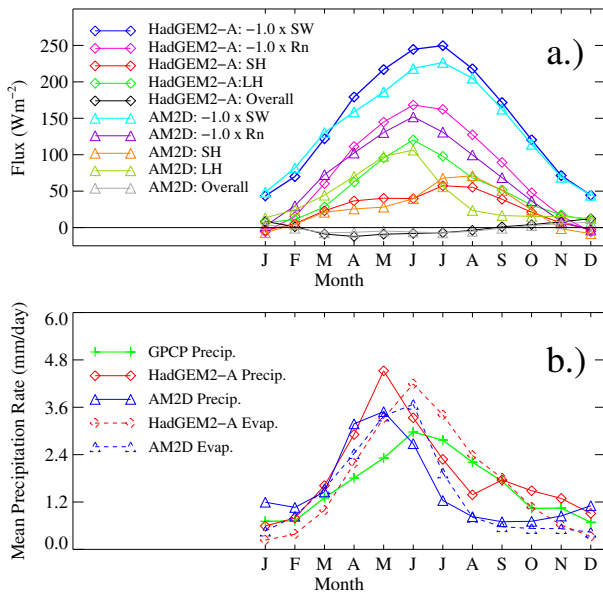
Fig. 11. As Fig. 9, but for the northern Great Plains.

HadGEM2-A throughout the year, and the tendency of the latent heat flux in AM2D to peak earlier in the annual cycle. AM2D is significantly warmer than HadGEM2-A in January, both under clear-sky and all-sky conditions, as expected from Fig. 3. As in the Mississippi valley the sensible heat fluxes during the later part of the night are more different in absolute terms than the net radiative fluxes.

The annual cycle of the surface fluxes and the hydrological budget is shown in Fig. 11. The annual variation of the hydrological cycle is similar to that seen for the Mississippi valley. Both models overestimate precipitation during the winter and spring, but underestimate it during the summer. The evaporation in AM2D lags the precipitation more closely than in HadGEM2-A, again as seen in the case of the Mississippi valley.

4.3 The south-west

Figure 12 shows diurnal composites of the surface temperature and fluxes for the south-west (region S). In this region the models generally agree more closely than in the other two regions under clear-sky conditions and whilst there is still a warm bias relative to the retrievals in July, this is less pronounced than elsewhere. Again, AM2D shows higher latent heat fluxes than HadGEM2-A during the day in April and lower latent heat fluxes in July, together with a smaller amplitude of the diurnal cycle of the overall surface flux. Here the nocturnal differences between the sensible heat fluxes are more pronounced than those in the net radiative fluxes in all months.

The annual cycles of the surface flux budget and the hydrological cycle are shown in Fig. 13. The differences between the two models are qualitatively similar to those seen in the other two regions, although they are not so pronounced because the latent heat fluxes are lower. Both models show excessive precipitation in the winter and the early spring, while AM2D shows a lack of precipitation in the summer. Once more it seems that in AM2D the evaporation lags the precipitation more closely than in HadGEM2-A, but this is less pronounced than in the other two regions.

5 Conclusions

The land surface temperatures simulated in two climate models have been compared with satellite retrievals. In broad terms, the models show similar features relative to the retrievals, with realistic, though sometimes excessive diurnal ranges. Maximum and minimum temperatures, however, are more subject to biases, particularly in the summer, where both models show significant warm biases over the lower Mississippi valley. Given its significantly larger diurnal range, it could be argued that the simulation of land surface temperature provides a more stringent test of a model than the simulation of near-surface air temperature, while surface temperatures have the advantage of being more directly linked to the surface flux budget.

The link between surface temperatures and the surface flux budget has been explored by considering the diurnal cycle over three regions. These show that although AM2D and HadGEM2-A yield fairly similar diurnal ranges of surface temperature, the amplitude of the diurnal cycle of the overall surface flux is more different, with that of HadGEM2-A being roughly twice as large as that of AM2D: this behaviour is consistent across the annual cycle. It is in the simulation of the turbulent fluxes that the models differ most significantly, with AM2D generally showing higher turbulent losses from the surface during the day and more turbulent warming of the surface by night.

To complement the diurnal composite flux budgets, annual composite flux budgets have also been constructed for these regions. In both models there is a tendency towards excessive annual variation in the precipitation with excessive rainfall during the winter and early spring and lower rainfall during the summer. The latent heat fluxes in AM2D lag the precipitation more closely than those in HadGEM2-A, so that during April, latent heat fluxes tend to be higher in AM2D, particularly over the Mississippi valley, leading to lower surface temperatures, while by July the latent heat fluxes in AM2D have declined substantially, leading to a compensating increase in the surface temperature and consequently to higher sensible heat fluxes, so that maximum surface temperatures in AM2D are generally higher than in HadGEM2-A in this month.
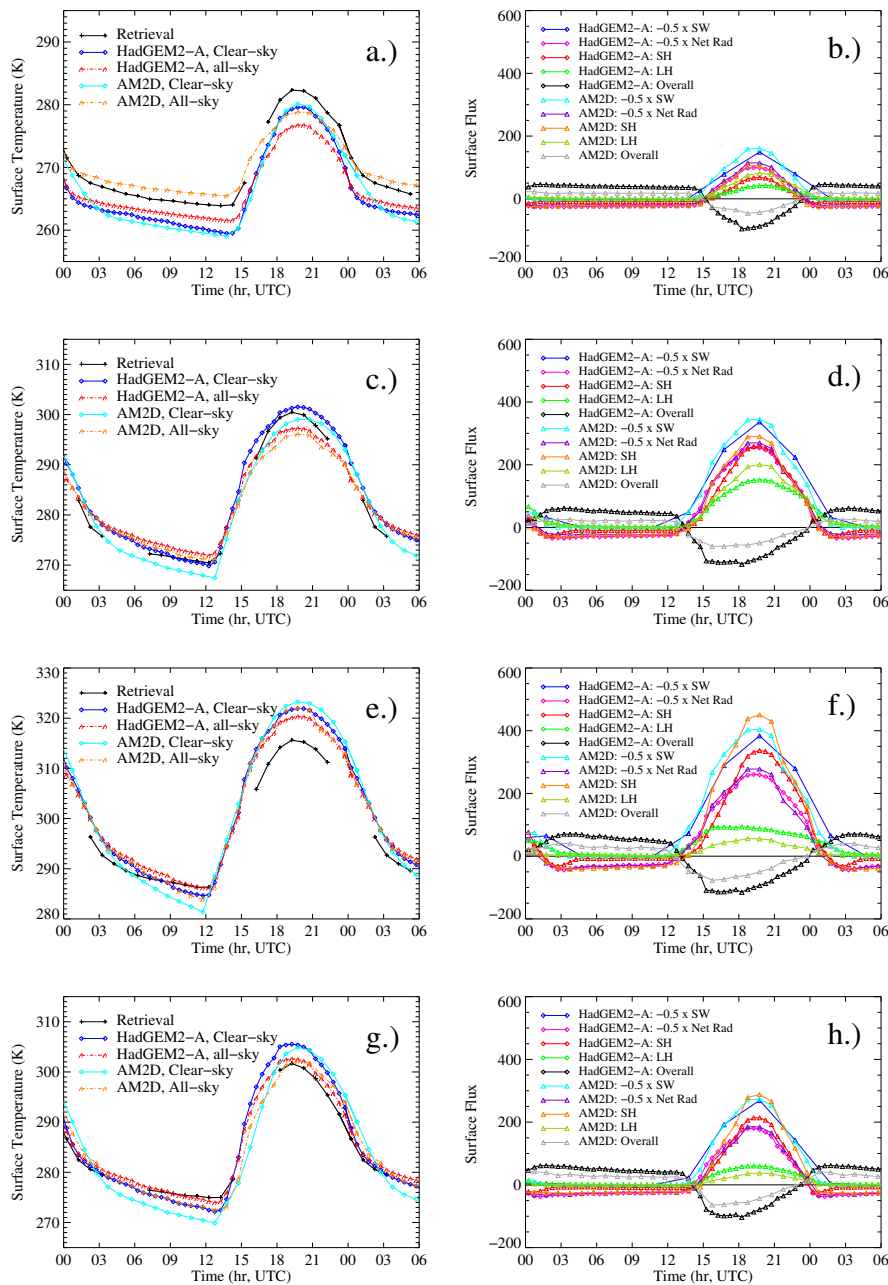
Fig. 12. As Fig. 8, but for the south-west (region S).

These comparisons demonstrate that the high temporal and spatial resolution of satellite data is useful in assessing the models' performance and suggest that comparison over wider domains would be instructive, if data were available. This study has focused on models running in climate mode, but for more precise comparisons, more attention must be given to the issue of cloud clearing. Consequently, there is an important role for complementary studies of surface temperature in forecasting models, where judicious selection of

cases would enable issues of cloud clearing to be circumvented. In this context data with high temporal resolution might be very instructive in establishing a more detailed picture of the diurnal cycle, particularly though the morning and evening transitions, especially if combined with ground measurements. The work of Jimenez et al. (2008) provides one example of what may be accomplished in this area.
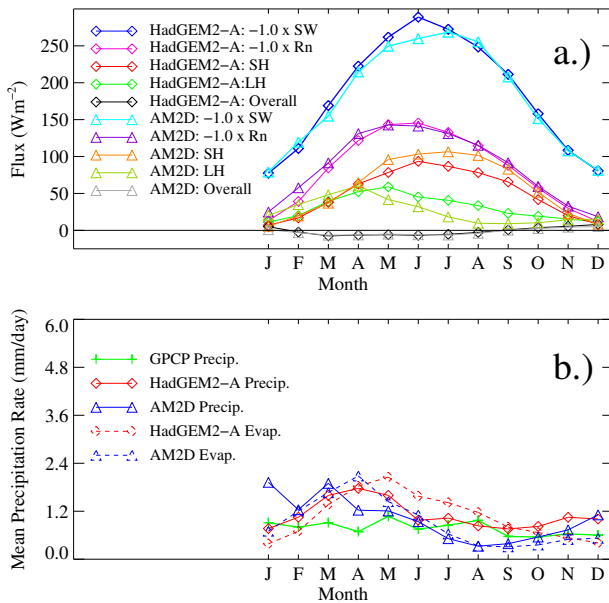
Fig. 13. As Fig. 9, but for the south-west.

Edited by: P. Jöckel

References

Coulter, R. L. and Doran, J. C.: Spatial and Temporal Occurrences of Intermittent Turbulence during CASES-99, Bound.-Lay. Meteorol., 105, 329–349, 2002.

Cuxart, J., Holtslag, A. A. M., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., Conangla, L., Ek, M., Freedman, F., Hamdi, R., Kerstein, A., Kitagawa, H., Lenderink, G., Lewellen, D., Mailhot, J., Mauritsen, T., Perov, V., Schayes, G., Steeneveld, G. J., Svensson, G., Taylor, P., Weng, W., Wunsch, S., and Xu, K. M.: Single-column model intercomparison for a stably stratified atmospheric boundary layer, Bound.-Lay. Meteorol., 118, 273–303, 2006.

Dempsey, C. L., Howard, K. W., Maddox, R. A., and Phillips, D. H.: Developing Advanced Weather Technologies for the Power Industry, B. Am. Meteorol. Soc., 79, 1019–1035, 1998.

Donner, L. J.: A cumulus parameterization including mass fluxes, vertical momentum dynamics and mesoscale effects, J. Atmos. Sci., 50, 889–906, 1993.

Global Atmospheric Model Development Team: The New GFDL Global Atmosphere and Land Model AM2/LM2: Evaluation with Prescribed SST Simulations, J. Climate, 17, 4641–4673, 2004.

Hansen, J., Ruedy, R., Sato, M., Imhoff, M., Lawrence, W., Easterling, D., Peterson, T., and Karl, T.: 2001: A closer look at United States and global surface temperature change, J. Geophys. Res., 106, 23947–23963, doi:10.1029/2001JD000354, 2001.

Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolph, B., and Schneider, U.: The global precipitation climatology project (GPCP) combined precipitation dataset, B. Am. Meteorol. Soc., 78, 5–20, 1997.

Inamdar, A. K., French, A., Hook, S., Vaughan, G., and Luckett, W.: Land surface temperature retrieval at high spatial and temporal resolutions over the southwestern United States, J. Geophys. Res., 113, D07107, doi:10.1029/2007JD009048, 2008.

Karl, T. R., Jones, P. D.., Knight, R. W., Kukula, G., Plummer, N., Razuvayev, V., Gallo, K. P., Lindeseay, J., Charlson, R. J. and Peterson, T. C.: Asymmetric trends of diurnal maximum and minimum temperature, B. Am. Meteorol. Soc., 74, 1007–1023, 1993.

Jimenez, M. A., Mira, A., Cuxart, J., Luque, A., Alonso, S. and Guijarro, J. A.: Verification of a clear-sky mesoscale simulation using satellite-derived surface temperatures, Mon. Weather Rev., 136, 5148–5161, 2008.

Mahrt, L. and Vickers, D.: Bulk formulation of the surface heat flux, Bound.-Lay. Meteorol., 110, 357–379, 2004.

Martin, G. M., Ringer, M. A., Pope, V. D., Jones, A., Dearden, C., and Hinton, T. J.: The physical properties of the atmosphere in the new Hadley Centre Global Environmental Model, HadGEM1. Part 1: Model description and global climatology, J. Climate, 19, 1274–1301, 2006.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology, J. Climate, 12, 829–856, 1999.

Smith, R. N. B.: A scheme for predicting layer clouds and their water contents in a general circulation model, Q. J. Roy. Meteorol. Soc., 116, 435–460, 1990.

Solomon, S., Qin, D., Manning, M., Marquis, M., Averyt, K., Tignor, M. M. B., Miller Jr., H. L., and Chen, Z. (Eds.): Climate Change 2007. The Physical Science Basis, Cambridge University Press, 2007.

Sun, D. and Pinker, R. T.: Estimation of land surface temperature from a geostationary operational environmental satellite (GOES-8), J. Geophys. Res., 108(D11), 4326, doi:10.1029/2002JD002422, 2003.

Sun, D., Kafatos, M., Pinker, R. T., and Easterling, D. R.: Seasonal variations in diurnal temperature range from satellite and surface observations, IEEE Trans. Geosci. Rem. Sens., 44, 2779-2785, 2006a.

Sun, D., Pinker, R. T., and Kafatos, M.: Diurnal temperature range over the United States: a satellite view, Geophys. Res. Lett., 33, L05705, doi:10.1029/2005GL024780, 2006b.

Sun, J.: Diurnal variations of thermal roughness height over grassland, Bound.-Lay. Meteorol., 92, 407–427, 1999.

Tiedtke, M.: Representation of clouds in large scale models, Mon. Weather Rev., 121, 3040–3061, 1993.

Wilcox, E. M. and Donner, L. J.: The Frequency of Extreme Rain Events in Satellite Rain-Rate Estimates and an Atmospheric General Circulation Model, J. Climate, 20, 53–69, 2007.

Xie, P. P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs, B. Am. Meteorol. Soc., 78, 2539–2558, 1997.

Yin, X., Gruber, A. and Arkin, P.: Comparison of the GPCP and CMAP merged gauge-satellite monthly precipitation products for the period 1979–2001, J. Hydromet., 5, 1207–1222, 2004.

Zilitinkevich, S. S., Hunt, J. C. R., Esau, I. N., Grachev, A. A., Lalas, D. P., Akylas, E., Tombrou, M., Fairall, C. W., Fernando, H. J. S., Baklanov, A. A., and Joffre, S. M.: The influence of large convective eddies on surface layer turbulence, Q. J. Roy. Meteorol. Soc., 132, 1423–1456, 2006.