CEOP-based Diagnosis of Prediction Skill of Four Operational GCMs and One Land Data Assimilation System

K. Yang, M. Rasmy, S. Rauniyar, T. Koike, K. Taniguchi, K. Tamagawa, P. Koudelova, M. Kitsuregawa, T. Nemoto, E. Ikoma, M. Yasukawa, M. Bosilovich, S. Williams

Abstract

Based on the platform of the Coordinated Enhanced Observing Period (CEOP) project, this study evaluated forecast skill of four operational GCMs (BMRC, JMA, NCEP, and UKMO) and NASA global land data assimilation system (GLDAS) through comparisons between in situ data and model output of CEOP/EOP 3 (2002/10/1~2003/9/30). This evaluation not only contributes to improving forecast skill but also provides guidance for data users to choose appropriate data from these model products for their applications.

Data: In situ and model output at 27 CEOP reference sites

Table 1 Geographic information and EOP3 data availability at CEOP reference sites (blank – no data; ●- data used; ◎- T_a used but q_a unavailable; O- data questionable or vegetation-type mismatched).

CSE	Reference Site Name	Code	Lat. (N)	Lon. (E)	Ele. (m)	Data availability: Ta&qa	Tsfc	Radiation	Flux	Rain	References/PI
BALTEX	Lindenberg	LIN	52.2	14.1	112	●	●	●	●	●	Beyrich & Adam(2004)
	Cabauw	CAB	52.0	4.9	-1	●	●	●	●	●	Isemer (2002)
	Sodankylä	SOD	67.4	26.7	179	●		●		O	Isemer (2002)
GAPP	ARM Southern Great Plains	SGP	36.6	-97.5	313	●		●	●	●	Raymond McCord
	Bondville	BON	40.0	-88.3	300	●	●	●	●	●	Tilden P. Meyers
	Fort Peck	FPE	48.3	-105.1	800	●	●	●	●	●	Tilden P. Meyers
	Oak Ridge	ORI	36.0	-84.3	275	●	●	●	●	●	Tilden P. Meyers
CAMP	Eastern Siberian Tundra	ES1	71.6	128.8	38	●	●	●			Ohata et al. (1999)
	Eastern Siberian Tiaga	ES2	62.3	129.6	220	●	●	●			Ohta et al. (2001)
	Mongolia	MON	45.7	106.3	1393	●	●			●	Kaihotsu et al. (2003)
	Tongyu	TON	44.4	122.9	184	●		●	●	●	Wenjie Dong & Huizhi Liu
	Tibet	TIB	31.4	91.9	4580	●	●	●		●	Ishikawa et al. (2001)
	West Tibet	GAI	32.5	84.1	4416	●	●	●			Ishikawa et al. (2001)
	Himalayas	HIM	28.0	86.8	5050	●		●		●	Bollasina et al. (2002)
	North South China Sea	NSC	25.0	121.2	8	O		O		●	Chen et al. (2004)
	Korean Haenam	KHA	34.6	126.6	14	◎		●			Kim et al. (2002)
	Korean Peninsula	KPE	37.4	127.9	330	●		●		●	Kim et al. (2002)
	Chao-Phraya River - Lampang	CPL	18.4	99.5	241	◎		●		●	Masatoshi AOKI
	North-East Thailand	NET	14.5	102.4	311	◎		●		●	Masatoshi AOKI
	Western Pacific Ocean	WPO	7.1	134.3	40	O		O		●	Kubota et el. (2002)
	Equatorial Island	EIS	-0.2	100.3	699	●				●	Mori et al. (2004)
LBA	Manaus	MAN	-2.6	-60.2	130	●	●	O	O	●	Marengo et al. (2003)
	Santarem	SAN	-3.0	-55.0	N/A	●		●	●	●	Marengo et al. (2003)
	Pantanal	PAN	-19.6	-57.0	N/A	●	●	●		●	Marengo et al. (2003)
ARM	ARM NSA-Barrow	NSA	71.3	-156.6	8	●	●	●		●	Tilden P. Meyers
	ARM TWP-Manus	MNS	-2.1	147.4	4	O		O		●	Raymond McCord
	ARM TWP-Darwin	DAR	-12.4	130.9	29.9	●	●	●		●	Raymond McCord

Table2 In situ and GCMs-used vegetation types at CEOP sites (UKMO column shows the type with the maximum area fraction. No explicit vegetation in BMRC model).

CSE	Reference Site Name	Vegitation Type: In-situ	JMA	NCEP	UKMO	GLDAS
BALTEX	Lindenberg	Grassland	Savana	Cultivations	Grass	Cropland
	Cabauw	Short grass	Savana	Groundcover only	Grass	Cropland
	Sodankylä	Needleleaf evergreen trees	Needleleaf evergreen trees	needleleaf evergreen trees	Needleleaf trees	Bareground
GAPP	ARM SGP	Grassland	Mixed forest	Cultivations	Grass	Cropland
	Bondville	Cropland	savanna	Cultivations	Grass	Cropland
	Fort Peck	Grassland	Grasslands	Groundcover only	Grass	Grassland
	Oak Ridge	Mixed forest	Mixed forest	Mixed forest	Broadleaf trees	Wooded Grassland
CAMP	Eastern Siberian Tundra	Open Shrubland	Ocean	Tundra	Bare Soil	Open Shrubland
	Eastern Siberian Tiaga	Needleleaf deciduous trees	Needleleaf deciduous trees	Needleleaf deciduous trees	Needleleaf trees	Wooded Grassland
	Mongolia	Grassland	Grasslands	Bare soil	Bare Soil	Grassland
	Tongyu	Cropland	Grasslands	Cultivations	Grass	Cropland
	Tibet	Grassland	Grasslands	Bare soil	Grass	Grassland
	west Tibet	Grassland	Bare soil	Bare soil	Shrubs	Bareground
	Himalayas	Grassland	Grasslands	Mixed forest	Grass	Bareground
	North South China Sea	N/A	Ocean	groundcover only	Grass	Shrubs
	Korean Haenam	groundcover only	Savanna	Groundcover only	Grass	Wooded Grassland
	Korean Peninsula	Broadleaf deciduous trees	Broadleaf deciduous trees	Cultivations	Grass	Cropland
	Chao-Phraya River - Lampang	Deciduous Forest	Savanna	Cultivations	Broadleaf trees	N/A
	North-East Thailand	Broadleaf decidous trees	Savanna	Broadleaf decidous trees	Broadleaf trees	Cropland
	western Pacific Ocean	N/A	Ocean	Ocean	Ocean	Ocean
	Equatorial Island	N/A	Tropical forest	Cultivations	Broadleaf trees	Grassland
LBA	Manaus	Tropical forest	Tropical forest	Tropical forest	Tropical forest	Tropical forest
	Santarem	Tropical forest	Tropical forest	Tropical forest	Topical forest	Tropical forest
	Pantanal	Savanna	Savanna	Savana	Grass	Wooded Grassland
ARM	ARM NSA-Barrow	Open Shrubland	Tundra	Tundra	Bare Soil	Open Shrubland
	ARM TWP-Manus	N/A	Ocean	Ocean	Ocean	Ocean
	ARM TWP Darwin	Savana	Savanna	Savana	Broadleaf trees	Wooded Grassland

Methodology

For appropriate comparisons between in situ observations and grid-based model output, this study follows three rules.

- First, we compare monthly-mean values or monthly-mean diurnal cycle instead of hourly or 3-hourly values, because spatial variability can be effectively smoothed through temporal averaging.
- Second, sites that have a very different land use between in situ and models are excluded from the comparisons of surface temperature and fluxes, because of their sensitivity to surface conditions. For example, observations of surface variables at small island sites can be very different from model output that actually represents the values on surrounding sea surface.
- Third, a systematic bias is suggested only if it is found for most models or most sites.

General evaluation of four GCMs (Fig.1)

a. Monthly-mean air temperature (Tair) and humidity (qair)

All the models yielded small MBEs for Tair and qair. **BMRC** gave a RMSE of 4.1 K for Tair and 2 g kg^{-1} for qair, but other models produced much smaller RMSEs (~2.5 K for Tair; ~1 g kg^{-1} for qair).

b. Monthly-mean surface temperature (Tsfc)

All the models produced negative bias of Tsfc. NCEP yielded the smallest scattering, and UKMO yielded the largest scattering. BMRC data are not available.

c. Monthly-mean surface radiation (SWD, LWD)

SWD was over-predicted by all the models. LWD was under-predicted by NCEP and JMA while its prediction by UKMO is much better. JMA gave the maximum biases for SWD and LWD (30 W m^{-2}), and UKMO gave the minimum biases. However, the errors in the two components counteract or compensate each other, which generally results in a small bias for the total downward radiation. As a result, both JMA and UKMO produced better total radiation than NCEP and BMRC.

d. Monthly-mean surface heat fluxes (H, lE)

RMSEs in H of all the models (> 20 W m^{-2}) are comparable to the order of the observed one (17 W m^{-2}). JMA produced the smallest MBE and RMSE for H and lE and seems to be the best one. NCEP much over-predicted latent heat fluxes (or evapo-transpiration) while under-predicted sensible heat fluxes. BMRC produced the maximum RMSE value and a moderate MBE value.

e. Monthly-mean precipitation (P)

NCEP and JMA over-predicted precipitation while BMRC and UKMO under-predicted it. The difference between NCEP MBE and BMRC MBE is even comparable to the magnitude of the observed precipitation (~2.5 mm d^{-1}).

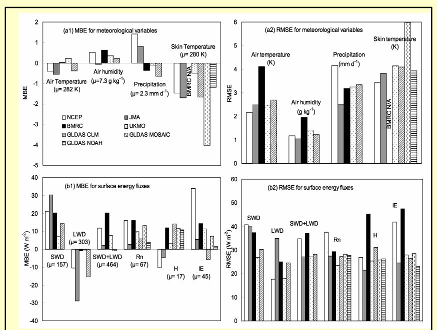

Figure 1 The Mean Bias Error (MBE) and Root Mean Square Error (RMSE) of EOP3 at all the sites.

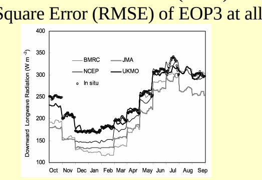

Figure 2 Monthly-mean diurnal variation of downward longwave radiation at Eastern Siberian Tundra site.

Evaluation of representations of physical processes

1. Radiation scheme

All these schemes systematically over-predicted downward shortwave radiation at almost all the sites. This over-estimation by GCMs was also found by many early studies (e.g. Cess et al. 1995). Ramanathan et al. (1995) suggested that the cloud absorption in the model might be less than observed. Recent evaluation by Wild et al. (2006) shows that atmospheric clear-sky absorption is under-estimated 10 W m^{-2} in many GCMs. If this is true for the CEOP participating GCMs, then about half of the errors in Fig. 1(b1) could be explained by the clear-sky absorption.

General under-prediction of downward longwave radiation by NCEP and JMA is also found for many other models (e.g. Garratt and Prata 1996; Wild et al. 2001). An example is shown in Fig. 2 for East Siberia Tundra site, where the longwave radiation was well predicted by UKMO whereas poorly predicted by other models. Therefore, we tend to believe longwave schemes rather than external factors (biomass burning, scale-mismatch etc.) should play the major role for the under-estimation. The RRTM, which is supposed to be the best for clear-sky radiation computation (Morcrette, 2002), indeed show small errors in NCEP model (see Fig. 1), but the Slingo and Wilderspin (1986) scheme used in UKMO shows even higher accuracy for all-sky conditions.

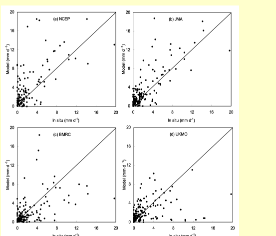

Figure 3 Monthly mean precipitation of EOP3 at all the sites.

2. Cloud scheme

Figure 1 shows that JMA and NCEP over-predicted precipitation and BMRC under-predicted it. UKMO gave the smallest MBE. Fig. 3 shows errors in forecasts of monthly-mean heavy precipitation are a major contributor to the total model biases. JMA and NCEP predicted more heavy-precipitation months while BMRC predicted fewer heavy-precipitation months. This suggests that Arakawa-Schubert scheme that is used in JMA and NCEP tend to predict more precipitation.

Table 3 shows the total precipitation amount in the rainy season at individual sites for EOP3 (the rainy season is defined in Table 3). All the models show systematical and significant over-estimates of precipitation in some regions (HIM and CAB) or under-estimates in some other regions (DAR, MNS and NSA), suggesting that precipitation schemes in these GCMs also share some model deficiencies.

CSE	Site code	Months	Lat (N)	Lon (E)	OBS.	BMRC F	BMRC A	UKMO F	UKMO A	JMA	NCEP
BALTEX	LIN	JJA	52.2	14.1	126	31	9	146	159	158	191
	CAB	**JJA**	**52**	**4.9**	**21**	**31**	**26**	**116**	**150**	**159**	**164**
GAPP	SGP	JJA	36.6	-97.5	175	90	89	419	400	400	385
	BON	JJA	40	-88.3	244	95	57	375	360	464	553
CAMP	MON	JJA	46.3	107.3	140	42	42	90	85	184	50
	TON	JJA	44.4	122.9	235	84	87	410	366	406	192
	TIB	JJA	32	91.9	363	131	198	350	403	407	99
	HIM	**JJA**	**28**	**86.8**	**352**	**1157**	**769**	**648**	**569**	**1334**	**1950**
	NSC	JJA	25	121.2	449	166	211	539	755	381	368
	WPO	MAM	7.1	134.3	770	571	522	981	977	998	1487
	EIS	MAM	-0.2	100.3	877	395	421	371	402	1121	1007
LBA	MAN	MAM	-2.6	-60.2	786	311	397	655	777	606	1268
	SAN	MAM	-3	-55	636	507	440	467	437	710	885
ARM	*NSA*	*JJA*	*71.3*	*-156.6*	*102*	*32*	*23*	*86*	*80*	*52*	*92*
	MNS	*MAM*	*-2.1*	*147.4*	*1268*	*536*	*586*	*838*	*893*	*952*	*903*
	DAR	*DJF*	*-12.4*	*130.9*	*1565*	*537*	*522*	*815*	*831*	*1210*	*926*

Table 3 Total precipitation in rainy seasons of EOP3 at 16 sites where data are available (**Bold** row:- all the models overestimate; *Italic* row:- all the models underestimate).

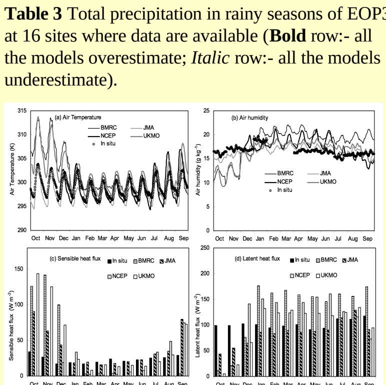

Figure 4 Monthly mean values of EOP3at LBA-Santarem site

3. Land surface scheme

Fig. 4 shows BMRC much over-predicted Tair and much under-predicted qair in the summer season (Oct-Dec) at LBA/Santarem. Similar biases are found at other LBA/sites (Manus and Pentanal) and BALTEX sites (Linderberg and Cabauw) in the summer season. By contrast, BMRC under-predicted air temperature in some cold regions (e.g. Tundra and NSA-Barrow). UKMO also shows similar large over-prediction of air temperature for the tropical regions (e.g. Amazon, Thailand, and Australian Darwin) in the summer and under-prediction for West Tibet. These large biases were caused by incorrect surface energy partition between sensible heat and latent heat. An example for Santarem is shown in Figs. 4(c) and 4(d). This explained why the RMSE of Tsfc, H, Tair and qair predicted by BMRC is larger than other models (see Fig.1). This result is not surprising because BMRC uses a simple bucket hydrological model that does not explicitly account for vegetation processes.

Figure 5 shows the relationship between errors in precipitation and errors in evapo-transpiration of NCEP and JMA at eight flux sites. NCEP gave a positive correlation between errors in precipitation and in evapo-transpiration, suggesting that the over-prediction of precipitation by NCEP results in its over-prediction of evapo-transpiration and under-prediction of sensible heat fluxes (See Fig. 1). On the other hand, JMA gave a quite weak correlation, though both JMA and NCEP use the SiB2 land model.

Figure 6 shows that all the models under-predicted monthly-mean diurnal range (maximum minus minimum temperature) of surface skin temperature at West Tibet and Mongolia. Similar bias is also found at other sites (East Tibet and Fort Peck) in arid and semi-arid regions. Therefore, under-prediction of Tsfc diurnal range is a common phenomenon for arid and semi-arid regions and should be related to a model deficiency. This under-prediction implies that current land surface models may under-predict aerodynamic resistance or over-predict turbulent heat transfer capability for bare soil and sparse vegetation surfaces. In other words, current land models can well describe dense canopy processes but might be questionable for describing bare soil and sparse canopy processes. An example that improves representations of bare soil and sparse canopy processes was presented by Yang et al. (2006).

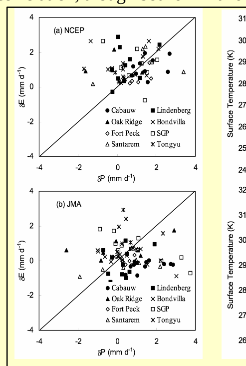

Figure 5 Comparison of model errors between GCM monthly-mean rainfall and evaporation

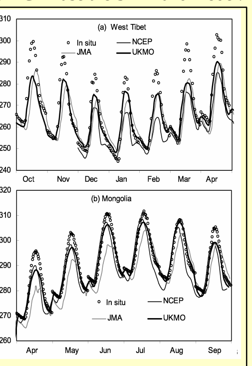

Figure 6 Monthly-mean diurnal variation of Tsfc in arid and semi-arid regions

Evaluation of precipitation diurnal cycle

1. Results from in situ data were comparable with those in the literature (not shown)
2. Fig. 7 shows the composite diurnal cycle of precipitation intensity and frequency pattern. It is shown that the peak of precipitation intensity occurs at 14 LST and 2 hours earlier than that of precipitation frequency.
3. Fig.7 shows a local minimum of precipitation intensity at 18 LST data, which has not been reported in the literature
4. Fig.8 shows an afternoon peak and a nighttime peak. At the tropical sites (8a), the afternoon peak is much stronger than the nighttime one.At higher latitudes (8b), the afternoon peak is comparable to the nighttime. The nighttime peak time varies from site to site, and thus it is not clear in the composite diurnal pattern.
5. Figure 7 shows that all the models produced the afternoon peak, but no model is able to produce the 18 LST local minimum and the nighttime peak. For the afternoon peak, the frequency patterns of JMA and NCEP are most close to the observed one, but UKMO predicted 1-2 hours earlier and BMRC 4-5 hours earlier than the observed one. There is no remarkable difference in precipitation intensity and frequency between forecast and analysis output for both UKMO and MBRC. This implies the diurnal change is mainly determined by model's nature rather than initial conditions.

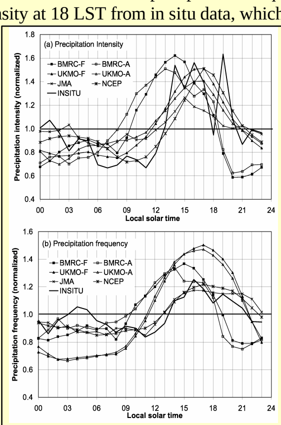

Figure 7 Composite diurnal cycle of precipitation derived from in situ data and GCM output in rainy seasons.

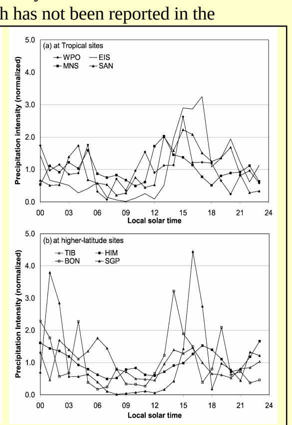

Figure 8 Diurnal cycle of precipitation derived from in situ data at tropical sites and at four higher-latitude sites

Evaluation of GLDAS

Fig. 9 shows four examples of comparisons of monthly-mean diurnal variation of Tsfc between in situ data and three GLDAS products. At West Tibet, GLDAS does not improve skin temperature simulation compared to the GCMs (see Fig.6(a)). Nevertheless, GLDAS/CLM and Noah clearly reproduced the observed surface temperature at other three sites. This is not surprising because GLDAS assimilates remotely sensed surface skin temperature. GLDAS/Mosaic produced very clear cold biases at nighttime, which should be attributed to its model deficiency since all the forcing data and assimilation data are identical in the three land models.

However, the three land models produced quite different fluxes, as shown in Fig. 10 for Bondville and Cabauw sites. Although both Noah and CLM well produced surface skin temperature at the two sites (See Fig.9), they yielded very different surface energy partition. GLDAS/Noah generally yielded better energy partition than CLM and Mosaic. For the summer season, not only all the modeled energy partitions deviate far from the observations, but also the differences among the models are comparable to the differences between models and measurements. These results indicate that (1) uncertainties in this data assimilation system are quite large and representations of key land processes and model parameters play a more important role than data assimilation technique, and (2) it is still difficult for a LDAS to generate reliable surface energy partition by only assimilating surface temperature. It is expected that these uncertainties would decrease after the system assimilates soil moisture-relevant satellite data, such as low frequency data of the Advanced Microwave Scanning Radiometer (AMSR) instrument carried by NASA Earth-observation satellite Aqua, as demonstrated by Yang et al. (2006).

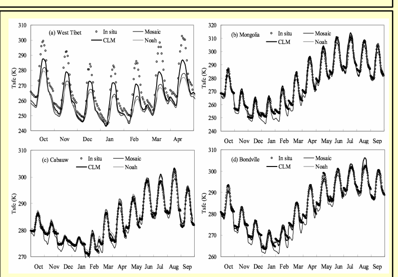

Figure 9 Tsfc monthly-mean diurnal variation

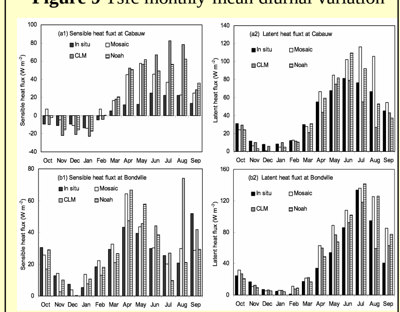

Figure 10 H & lE monthly-mean diurnal variation

Conclusions

- **Air temperature and humidity**. JMA and NCEP usually have good skill in estimating air temperature, humidity, but BMRC and UKMO have much higher biases in some regions in specific periods. In detail, BMRC over-predicted air temperature and under-predicted humidity in Amazon and Baltic regions during their summer season, and under-predicted air temperature in the Polar region. UKMO over-predicted air temperature for the tropical regions (e.g. Amazon, Thailand, and Australian Darwin) in the summer and under-predicted it for West Tibet in the winter. These biases were caused by incorrect surface energy budget in land surface models.
- **Surface temperature**. All the GCMs under-predicted the diurnal range of surface skin temperature in arid and semi-arid regions. As the diurnal range is strongly determined by surface resistances for heat transfer, this under-prediction in turn implies that the land surface models may under-predict aerodynamic resistance for heat transfer over bare soil and sparse vegetation surfaces.
- **Radiation**. Downward shortwave radiation was over-predicted in all the GCMs, perhaps due to under-estimation of clear-sky and/or cloud absorption. Downward longwave radiation was under-predicted by NCEP and JMA while its prediction by UKMO is much better. We tend to believe that it is the longwave scheme of the GCMs rather than some external factors that results in this under-prediction. JMA gave the poorest estimation for both downward radiation components, and UKMO gave the best. It is interesting that JMA and UKMO produced total downward radiation equally well, because the errors in the two downward components counteract or compensate each other.
- **Surface energy budget**. In general, the surface energy budget was not well predicted by all the models. JMA shows better skill to estimate surface energy budget. NCEP tends to over-predict latent heat fluxes, which is associated with its over-prediction of precipitation. BMRC uses a simple bucket hydrological model without explicit vegetation, which yielded unrealistic surface energy budget at some sites in specific seasons.
- **Precipitation**. Diurnal cycle of precipitation based on in situ data is comparable to that derived from dense observations or satellite data. Composite diurnal cycle shows an afternoon peak and a nighttime peak of precipitation intensity in rainy seasons. A low intensity around 18 LST was also observed at many sites. In the tropical regions, the afternoon peak is stronger than the nighttime peak. In other regions, both peaks are strong, but the onset time of the nighttime rainfall is more variable. JMA and NCEP tend to over-predict precipitation amount while BMRC tends to under-predict it. Particularly, there are more months with heavy precipitation in JMA and NCEP and fewer in BMRC. All the models produced an afternoon peak. JMA and NCEP predicted it well, but UKMO predicted it 1-2 hours earlier and BMRC predicted it 4-5 hours earlier. Reanalysis products show similar results, implying the modeled diurnal cycle is mainly determined by model's nature rather than initial conditions. No model reproduced the nighttime peak and the low intensity at 18 LST.
- **GLDAS**. Compared with the GCMs, GLDAS/CLM and Noah can improve surface skin temperature simulations except in dry regions, but there are noticeable differences in the surface energy partition among land models and also between the models and observations. Model uncertainties (model minus model) are comparable to model errors (model minus observation), indicating that representations of land processes and model parameters are of primary importance for simulations of surface energy budget, and data assimilation technique only plays a secondary role in GLDAS. To improve surface energy budget modeling, it is essential to improve parameterizations of land processes and to calibrate land models. Also, We suggested the necessity of assimilating not only surface skin temperature but also soil moisture-relevant satellite data into a land data assimilation system.