ERA report series

11 Flux tower observations for the evaluation of land surface schemes: Application to ERA-Interim

Tomáš Král

ECMWF Research Department and CHMI, Prague

Series: ERA Report Series

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

Abstract

Recent increase in availability of flux tower measurements opened new possibilities for validation of land surface schemes. Unlike conventional measurements (e.g. 2 meter temperature, relative humidity), the surface fluxes provide better handle on the actual representation of physical processes in the model. In this document we describe the work that was carried out to create a verification dataset of in-situ observations from a number of flux tower networks (FLUXNET, CEOP, SMOSMANIA, BERMS) and demonstrate the potential use of this data in a comparison with ERA-Interim model outputs.

1   Introduction

The parametrization of land surface processes is an important aspect of numerical weather prediction (NWP) models. In a coupled mode, surface scheme provides fluxes of momentum, heat and moisture to the atmospheric model. Proper representation of those fluxes is a necessary precondition for good quality of forecasts. In the past years several validation studies were assessing quality of ERA-Interim and ERA-40 reanalysis (Simmons et al. 2004, van den Hurk et al. 2000, Betts et al. 2006). The studies were mostly employing the conventional meteorological observations (SYNOPs or TEMPs) but not many of them have been done using flux tower measurements. In this paper we focus on the analysis of the surface energy budget in ERA-Interim forecasts using the very recent data from the eddy-covariance flux tower network FLUXNET. This network covers a broad range of climate regimes and vegetation zones which is particularly important for validation of the land surface schemes in the context of global weather forecasting. The emphasis is put mainly on the accuracy of the representation of diurnal and annual cycles using selected observation sites. We further analyse and discuss the results and try to explain observed discrepancies between observations and the model.

2   Model configuration - ERA-Interim

ERA-Interim is a reanalysis project taking advantage of the latest state of the art assimilation system at ECMWF covering a data-rich period since 1989 until the present (Uppala et al. 2005). The ERA-Interim reanalysis system is based on cycle 31r2 of the ECMWFs Integrated Forecasting System (IFS) which is using 4-dimensional variational assimilation with 12h observation window. Apart from the analyses, ERA-Interim is complemented with 10 day forecasts that are provided with a spatial resolution of approx. 79 km for grid point fields (on reduced Gaussian grid) and a T255 spectral harmonics representation for dynamic spectral fields. The vertical resolution of the model is 60 levels and the time step of the forecasts is 30 minutes. The ERA-Interim output is available every 3h for gridded data, however, for a limited set of diagnostic points one can obtain data with 1 hourly steps which is more suitable for a diurnal cycle validation.

The ERA-Interim uses an external climate database derived from the Global Land Cover Characteristics (GLCC). Four parameters are derived for each grid-box: dominant vegetation type for low and high vegetation and their respective area fractions. There are 20 distinct vegetation types characterized by a set of parameters: minimum canopy resistance, leaf area index (LAI), vegetation coverage, a coefficient of dependence of the canopy resistance on water vapour pressure deficit and the root distribution over the soil layers. It is important to note that these parameters are fixed and do not change within the year which has some implications for the seasonal variability in the model as we will see later (Viterbo et al. 1995).

Proper initialization of land surface prognostic variables is of great importance not only for short-range forecasts but also at medium and seasonal timescales due to slow temporal evolution of soil wetness. Because soil moisture observations are not available on a global scale, analysis of soil moisture relies indirectly on screen level parameters and the underlying atmospheric model. Initialization of land surface moisture in ERA-Interim is based on the optimal interpolation method where analysis increments of 2m relative humidity and 2m temperature are used to apply corrections to moisture in each soil layer (Douville et al. 2000). The basic idea is that a too dry soil will lead to a too dry and too warm boundary layer in the first-guess forecasts compared to the analysis (observations). This method prevents soil moisture from drifting at longer timescales.

3 FLUXNET data

In this study we use the 2006 data from the FLUXNET LaThuile Synthesis dataset which is a collection of flux tower eddy-covariance measurements from a number of regional flux tower networks across the globe (Baldocchi et al. 2001). Carbon dioxide, moisture and heat fluxes in these sites are determined by measuring the covariance between fluctuations in a vertical velocity and a mixing ratio of particular trace gases. Temporal information provided by individual sites ranges from time scales of seconds up to seasons and years. Raw data are typically measured at frequencies around 10-20 Hz and further processed to generate 1/2-hourly flux data which are passed from the local sites to their regional networks and from there on to FLUXNET where they are quality assessed and gap filled (using methods described in Moffat et al. 2007). The location of the sites is usually selected with emphasis on achieving a reasonably good spatial representativeness of the flux footprint in the area. The spatial scale of observations typically ranges from 100 to 1000 meters which however, is still far less then the typical resolution of state-of-the-art global circulation models (GCMs).

4 Methodology

In this study we used the nearest grid point values rather then interpolation methods in order to preserve the exact surface characteristics of the model grid box. Daily 12-36 hours forecasts were used for comparison with the observation time series to stay reasonably close to the analysis but allowing for the initial spin-up of the model. Although all observations were available with 30 min. sampling, the values have been averaged over 1 hour intervals to match the output frequency of the model. The data comes with different quality and availability so we have selected only a subset of sites for this study (see Table 3). To assess the land surface energy budget we have looked at its components: net radiation (NR), latent heat flux (LH), sensible heat flux (SH) and ground heat flux and complemented it by hourly accumulated precipitation, relative humidity and temperature at 2m. We have also looked at evaporative fraction (EF) which was computed as an average between 10-14 hours of mean solar time when the latent heat and sensible heat fluxes reach the maximum values.

5 Results

Considering a generally large spatial variability of land surface characteristics (e.g. albedo, vegetation type) over short distances, it is not easy to make a straightforward comparison and the results must be treated with caution. Table 3 shows vegetation types as seen by the model and as seen at the actual

Site id	Site name	Site veg. type	Model veg. type
Au-Fog	Fogg Dam Australia	Permanent wetlands	Interrupted forests
Au-How	Howard Springs Australia	Woody savannas	Interrupted forests
Au-Tum	Tumbarumba Australia	Evergreen broadleaf forests	45% Crops, 55% Evergreen broadleaf forests
Au-Wac	Wallaby Creek Australia	Evergreen broadleaf forests	50% Crops, 50% Interrupted forests
Ca-Qcu	Quebec Boreal Cutover Site	Evergreen needleleaf forests	Evergreen needleleaf forests
Ca-Qfo	Quebec Mature Boreal Forest	Evergreen needleleaf forests	Evergreen needleleaf forests
Fi-Hyy	Hyytiala Finland	Evergreen needleleaf forests	Evergreen needleleaf forests
Fr-Hes	Hesse Forest France	Deciduous broadleaf forests	65% Crops, 35% Deciduous broadleaf forests
Fr-Lbr	Le Bray France	Evergreen needleleaf forests	Evergreen needleleaf forests
It-Amp	Amplero Italy	Grasslands	40% Crops, 60% Interrupted forests
It-Cpz	Castelporziano Italy	Evergreen broadleaf forests	55% Evergreen shrubs, 45% Interrupted forests
It-Mbo	Monte Bondone Italy	Grasslands	25% Short grass, 75% Interrupted forest
It-Ro1, It-Ro2	Roccarespampani Italy	Deciduous broadleaf forests	50% Evergreen shrubs, 50% Interrupted forest
Nl-Ca1	Cabauw Netherlands	Grasslands	Crops
Nl-Loo	Loobos Netherlands	Evergreen needleleaf forests	70% Crops, 30% Interrupted forests
Ru-Fyo	Fyodorovskoye Russia	Evergreen needleleaf forests	Interrupted forests
Se-Faj	Fajemyr Sweden	Evergreen needleleaf forests	Interrupted forests
Us-Arb, Us-Arc	ARM Lamont USA	Grasslands	75% Short grass, 30% Interrupted forests
Us-Arm	ARM Southern Great Plains	Croplands	Crops
Us-Fr2	Freeman Ranch USA	Woody savannas	80% Crops, 20% Mixed forests/woodlands
Us-Ha1	Harvard Forest USA	Deciduous broadleaf forests	Deciduous broadleaf forests
Us-Srm	Santa Rita Masquite USA	Woody savannas	Semidesert
Us-Syv	Sylvania Wilderness Area USA	Mixed forests	Mixed forests/woodlands
Us-Ton	Tonzi Ranch USA	Woody savannas	30% Tall grass, 70% Interrupted forest
Us-Wcr	Willow Creek USA	Deciduous broadleaf forests	Mixed forests/woodlands

Table 1: List of FLUXNET sites used in the comparison and their vegetation types compared to the vegetation type of the model grid.

observation site. However, the vegetation classifications used in FLUXNET and ERA-Interim are different and thus the comparison provided in the table has only qualitative value. The vegetation type of FLUXNET sites is based on MODIS International Geosphere-Biosphere Programme (IGBP) land cover classification which includes 16 land cover classes. As we can see, the conformance of vegetation types varies significantly from site to site. At some sites (e.g. Au-Fog, Us-Ton and Us-Fr2) we can observe a considerable mismatch in vegetation type.

5.1   Seasonal cycle

In Table 2 we show the energy budget statistics for individual FLUXNET sites. As we can see, on daily time scale, the correlations of NR and LH fluxes are fairly good with average values around 0.8 and higher while correlations of SH fluxes reach 0.73 on average. We can also observe that at the majority of the sites we have a positive bias in LH flux and negative bias in SH flux. Bias of NR varies from site to site.

On the hourly time scale (see Table 3), in terms of biases we observe more or less the same pattern. The average correlation on hourly time scales for NR is 0.93 and for LH and SH fluxes about 0.84 which are fairly high values. A comprehensive seasonal cycle comparison for all FLUXNET sites can be found in Appendix A.

In a further evaluation of the energy budget we have looked at the partitioning between LH and SH fluxes which can be expressed in terms of evaporative fraction. We have observed that particularly for broadleaf forest sites (Fr-Hes, It-Ro1, Us-Ha1, Us-Wcr) the model fails to resolve seasonal variability of EF. These sites are characteristic with significantly varying LAI over the year which is reflected in a strong annual cycle in observed EF. However this seasonal variability of LAI is not accounted for in the ERA-Iterim land surface scheme which results in an unrealistically flat seasonal cycle of modelled EF as can be seen on Figure 1. We assume that this deficiency of the model can be attributed to the constant climatology of LAI used in ERA-Interim.

The lack of seasonal variability of LAI in the model also quantitatively affects a response of the surface evaporation to precipitation. Precipitation is either intercepted by the vegetation or is infiltrated in the soil. Moisture from the interception reservoir can evaporate more effectively (with less resistance) than the soil moisture. As the size of the interception reservoir is linearly proportional to the LAI, there is more intercepted water available for evaporation for higher LAI values.

In the Figure 2 and 3 we can observe that precipitation related LH flux seems to exceed observed values more frequently in the spring and autumn season when the LAI tends to be overestimated in the model.

5.2   Diurnal cycle

Systematic errors in the diurnal cycle can be substantial and behave quite differently at different sites. The examples shown in Figs 4 and 5 show that the surface radiative forcing is too strong at day and night (leading to some compensation in the diurnal averages). What we can see in general is that that positive or negative biases in net radiation (if present) automatically lead to errors in latent heat and sensible heat fluxes. However, in cases that net radiation is reasonably good, we still often get a positive bias in latent heat.
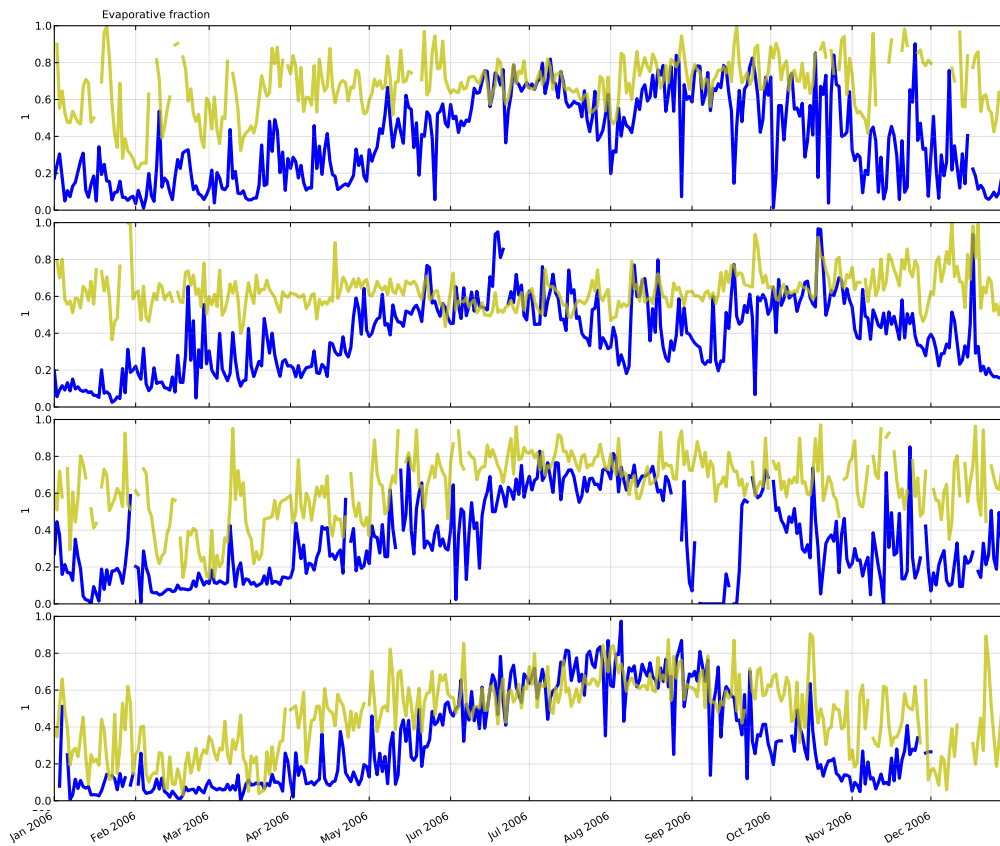
Figure 1: Seasonal cycle of evaporative fraction for four deciduous broadleaf forest sites Fr-Hes, It-Ro1, Us-Ha1, Us-Wcr in 2006. Blue lines indicate observations and yellow lines the model.
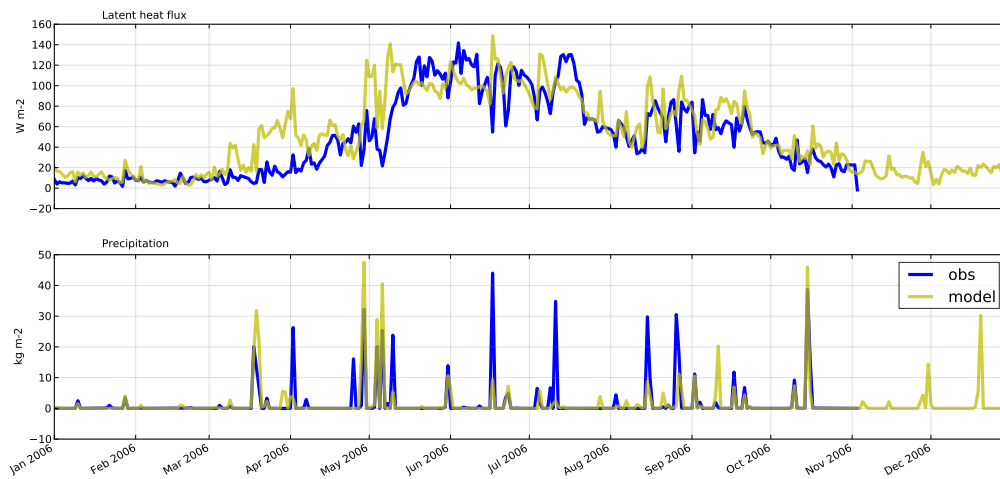

Figure 2: Response of latent heat flux to precipitation shown on 2006 daily averages for FLUXNET site Us-Arb.

6 Conclusion

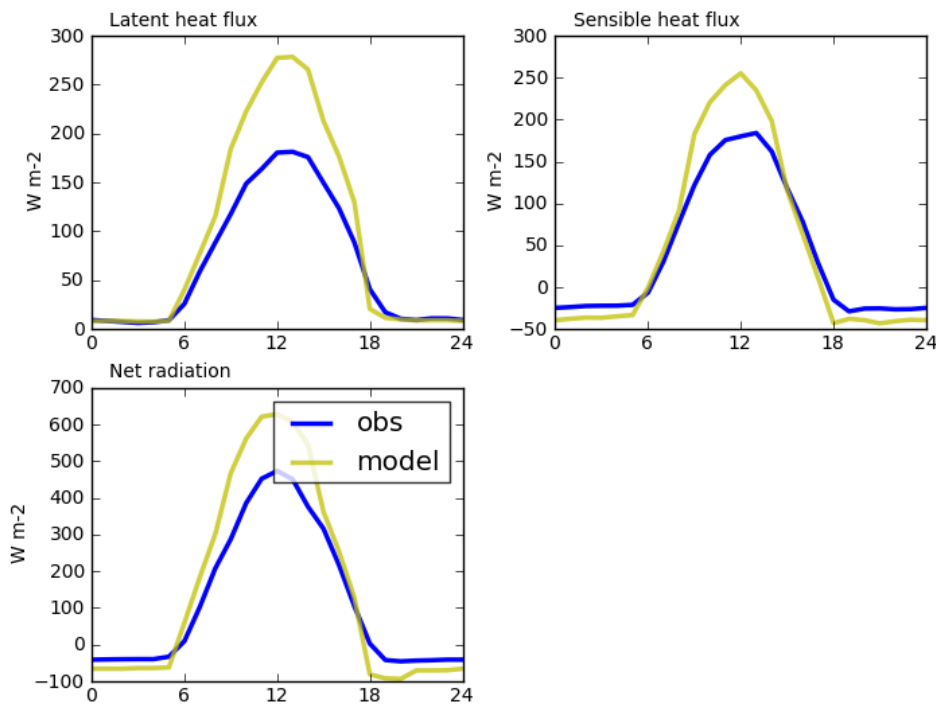Surface characteristics play an important role in the partitioning of the surface energy fluxes available for land surface processes. In state-of-the-art GCMs it is still not feasible to have these parameters correctly
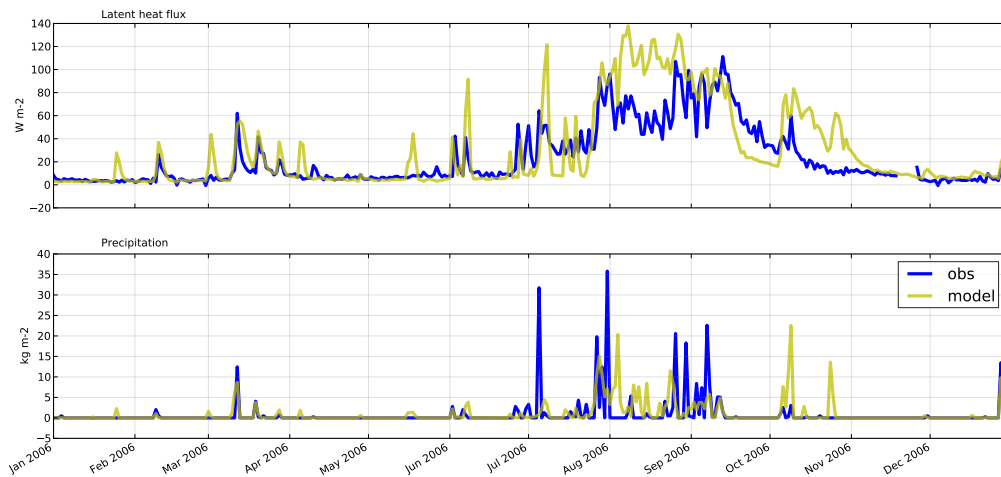
Figure 3: *Response of latent heat flux to precipitation shown on 2006 daily averages for FLUXNET site Us-Srm.*

Figure 4: *Averaged diurnal cycles of latent heat, sensible heat and net radiation fluxes for the ARM Southern Great Plains site (Us-Arm, 36.61ºN/262.51ºE) in April-May-June. The horizontal axis is in solar time.*

specified at the scales of flux tower measurements especially in heterogeneous locations. However, we have seen that at least for a half of the sites the representativity is fairly good.

We have also seen that in ERA-Interim a constant climatology of land surface parameters (specifically LAI) have detrimental effects on the resolution of seasonal cycle of EF (partitioning between LH and SH fluxes). This was evident in all deciduous broad leaf forest sites with strong seasonal LAI variation. The impact of constant LAI was also been observed in the excessive LH flux response on precipitation.

	LH			SH			NR		
	bias	**rmse**	**corr**	**bias**	**rmse**	**corr**	**bias**	**rmse**	**corr**
Au-Fog	-39.9	62.10	0.17	26.1	35.20	0.77	-7.4	47.20	0.54
Au-How	6.4	24.70	0.77	-0.5	16.30	0.65	5.4	30.00	0.76
Au-Tum	-22.8	32.80	0.82	9.0	22.70	0.73	0.14	33.10	0.93
Au-Wac	2.2	14.10	0.86	11.6	27.40	0.84	21.9	38.20	0.89
Ca-Qcu	8.3	15.20	0.91	5.7	28.20	0.75	1.3	32.10	0.90
Ca-Qfo	13.14	19.60	0.91	-9.11	23.70	0.86	-8.64	27.50	0.93
Fi-Hyy	7.70	16.20	0.88	-15.88	30.60	0.82	-8.66	21.30	0.96
Fr-Hes	32.95	38.60	0.86	-1.40	17.50	0.82	-13.77	25.90	0.95
Fr-Lbr	21.31	34.20	0.65	-9.35	28.10	0.86	-2.53	25.80	0.95
It-Amp	6.96	20.60	0.90	5.19	19.60	0.66	-10.32	30.40	0.89
It-Cpz	32.46	41.80	0.66	-17.08	32.80	0.90	-35.30	58.30	0.78
Nl-Ca1	8.08	17.56	0.91	5.18	25.20	0.89	0.88	27.60	0.94
Nl-Loo	12.19	24.40	0.82	-10.65	30.40	0.81	-25.72	38.90	0.93
Ru-Fyo	11.53	19.00	0.90	-9.85	25.03	0.84	-17.05	30.30	0.94
Se-Faj	26.52	39.20	0.83	-7.39	22.80	0.84	10.32	37.40	0.90
It-Mbo	15.42	24.50	0.88	15.60	24.10	0.63	23.89	42.30	0.87
It-Ro1	24.10	30.00	0.87	-5.19	31.20	0.62	-12.11	38.10	0.90
It-Ro2	20.14	30.46	0.82	5.15	44.31	0.20	-12.20	36.60	0.91
Us-Arb	7.84	22.00	0.85	-6.81	24.30	0.65	13.63	26.20	0.92
Us-Arc	3.28	22.70	0.84	-4.94	25.20	0.60	7.39	24.80	0.91
Us-Arm	24.38	38.10	0.66	-2.44	30.60	0.59	30.37	40.30	0.89
Us-Fr2	3.91	19.90	0.83	11.39	30.10	0.59	-8.70	38.50	0.79
Us-Ha1	38.86	48.80	0.78	-14.82	31.80	0.75	-13.19	39.60	0.87
Us-Srm	6.56	21.90	0.81	-4.61	20.60	0.82	-6.60	20.30	0.85
Us-Syv	37.39	46.00	0.84	3.44	23.90	0.77	27.81	48.70	0.85
Us-Ton	25.49	29.80	0.83	-7.53	19.30	0.91	1.73	16.90	0.97
Us-Wcr	24.20	33.20	0.85	5.64	30.10	0.66	10.96	37.00	0.89
average	13.28	29.16	0.80	-0.87	26.71	0.73	-0.98	33.83	0.88

Table 2: Bias, root mean square error (rmse) and correlation coefficient (corr) of surface fluxes for 2006 based on daily averages.

|	LH			SH			NR		
	bias	**rmse**	**corr**	**bias**	**rmse**	**corr**	**bias**	**rmse**	**corr**
Au-Fog	-39.49	109.9	0.74	26.54	86.58	0.84	-6.09	76.28	0.96
Au-How	6.17	62.04	0.89	-0.60	42.93	0.9	6.41	75.45	0.96
Au-Tum	-22.82	63.13	0.89	9.29	51.94	0.92	0.32	83.99	0.95
Au-Wac	2.19	38.3	0.86	11.60	55.26	0.87	21.91	94.4	0.92
Ca-Qcu	8.34	29.43	0.86	5.76	66.17	0.78	1.30	82.05	0.89
Ca-Qfo	13.15	31.2	0.87	-9.07	49.99	0.87	-8.62	63.51	0.94
Fi-Hyy	7.73	26.74	0.85	-15.86	50.47	0.83	-8.66	49.78	0.94
Fr-Hes	32.95	59.02	0.84	-1.40	38.18	0.81	-13.76	60.18	0.94
Fr-Lbr	21.31	59.03	0.74	-9.35	56.46	0.85	-2.52	67.23	0.95
It-Amp	6.49	44.83	0.91	4.93	36.87	0.79	-10.33	76.82	0.93
It-Cpz	32.20	69.59	0.76	-17.27	61.73	0.92	-35.66	0.34	0.94
Nl-Ca1	8.09	34.17	0.9	5.17	46.24	0.87	0.88	66.38	0.94
Nl-Loo	12.20	45.94	0.82	-10.65	55.58	0.81	-25.72	70.14	0.93
Ru-Fyo	11.53	31.37	0.88	-9.85	45.43	0.85	-17.09	60.44	0.93
Se-Faj	25.82	61.06	0.82	-7.35	45.87	0.84	9.54	96.63	0.89
It-Mbo	15.42	47.57	0.85	15.60	40.71	0.71	23.90	89.7	0.87
It-Ro1	24.10	56.24	0.85	-5.20	60.27	0.79	-12.11	84.36	0.94
It-Ro2	20.16	56.01	0.83	5.16	73.81	0.62	-12.16	87.71	0.94
Us-Arb	7.83	41.51	0.88	-6.93	54.97	0.89	13.60	84.6	0.96
Us-Arc	3.30	41.95	0.88	-4.99	47.71	0.91	7.35	79.21	0.95
Us-Arm	24.39	63.32	0.75	-2.40	62.42	0.83	29.83	111.43	0.92
Us-Fr2	3.78	40.07	0.87	10.97	59.63	0.89	-8.12	87.43	0.94
Us-Ha1	38.90	74.45	0.8	-14.78	54.74	0.86	-13.22	82.07	0.92
Us-Srm	6.62	43.81	0.76	-4.25	50.47	0.91	-5.90	71.37	0.95
Us-Syv	37.59	71.7	0.81	3.61	59.27	0.86	27.81	106.81	0.91
Us-Ton	25.51	51.54	0.83	-7.55	42.51	0.92	1.72	56.4	0.97
Us-Wcr	24.42	53.26	0.83	5.78	61.75	0.84	11.10	79.02	0.94
average	13.25	52.12	0.84	-0.86	54.00	0.84	-0.90	75.69	0.93

Table 3: Bias, root mean square error (rmse) and correlation coefficient (corr) of surface fluxes for 2006 based on hourly averages.
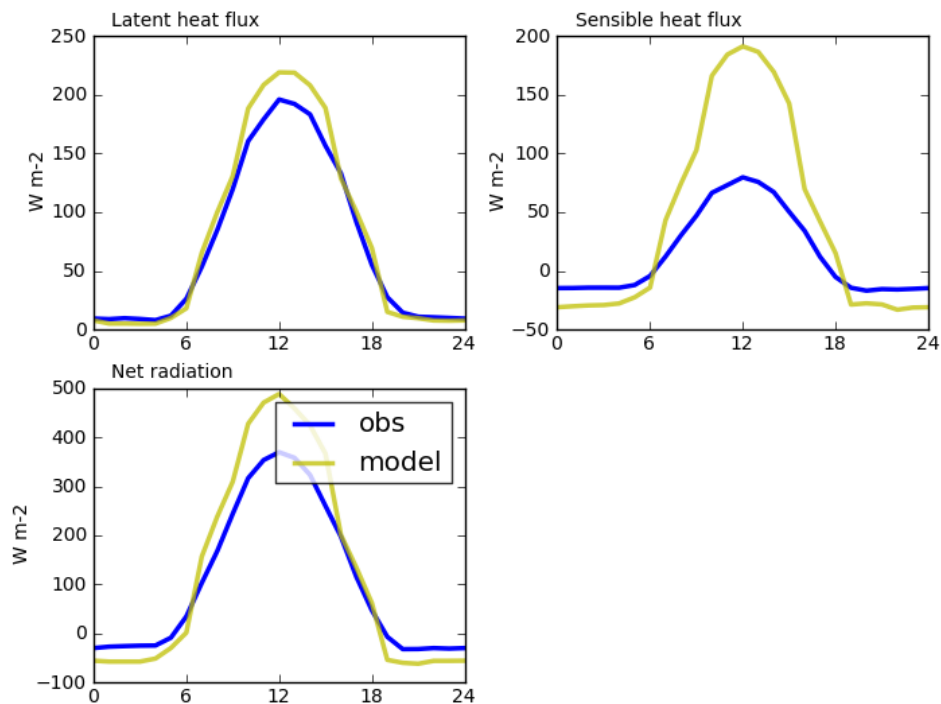
Figure 5: Averaged diurnal cycles of latent heat, sensible heat and net radiation fluxes for the Cabauw site (Nl-Ca, 51.97ºN/5.00ºE) in April-May-June. The horizontal axis is in solar time.

References

Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, R. Evans, and Others, (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide. *Bull. Amer. Meteorol. Soc.* **82**, 2415-2434.

Betts, A.K., J.H. Ball, A.G. Barr, T.A. Black, J.H. McCaughey, P. Viterbo, (2006): Assessing land-surface-atmosphere coupling in the ERA-40 reanalysis with boreal forest data. *Agric. Forest Meteorol.*, **140**, 365-382.

Douville, H.P. Viterbo, J.-F. Mahfouf, A. Beljaars, (2000): Evaluation of the optimal interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Wea. Rev.*, **128**, 1733-1756.

Moffat A., et al., (2007) Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric. Forest Meteorol.*, **147**, 209-232.

Simmons, A.J., P.D. Jones, V. da Costa Bechtold, A.C.M. Beljaars, P.W. Kallberg, S. Saarinen, S.M. Uppala, P. Viterbo, and N. Wedi, (2004): Comparison of trends and low frequency variability in CRU, ERA-40, and CEP/NCAR analyses of surface air temperature. *J. Geophys. Res.*, **109**, D24115

Uppala, S.M., P.W. Kallberg, A.J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P. McNally, J-F. Mahfouf, J-J. Morcrette, N.A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo and J. Woollen, 2005: The ERA-40

re-analysis. *Quart. J. Roy. Meteorol. Soc.*, **131**, 2961-3012.

Van den Hurk, B.J.J.M. and Viterbo, P. and Beljaars, A.C.M. and Betts, A.K., (2000): Offline validation of the ERA40 surface scheme. *ECMWF Technical Memorandum*, **295**.

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

Boussetta S., et al., 2011: Impact of a satellite-derived Leaf Area Index monthly climatology in a global Numerical Weather Prediction model. *ECMWF Technical Memorandum*, **640**.

Acknowledgements

Appendices

A   ERA-Interim - FLUXNET comparison

In this appendix we provide time series plots of daily averages for all FLUXNET sites used in the comparison with ERA-Interim outputs for year 2006. The compared variables are latent heat flux, sensible heat flux, net radiation, evaporative fraction, ground heat flux, 2m temperature, 2m relative humidity and precipitation.

Additionaly, comparison with Lindenberg, Chestnut Ridge and Quebec Boreal Forest sites, using WebgraF utility provided by Ulf Andrae, can be found in `ec:/pa8/WebgraF.tar.gz`.

Fogg Dam Australia -12.54N 131.31E (model -12.28N 131.48E)
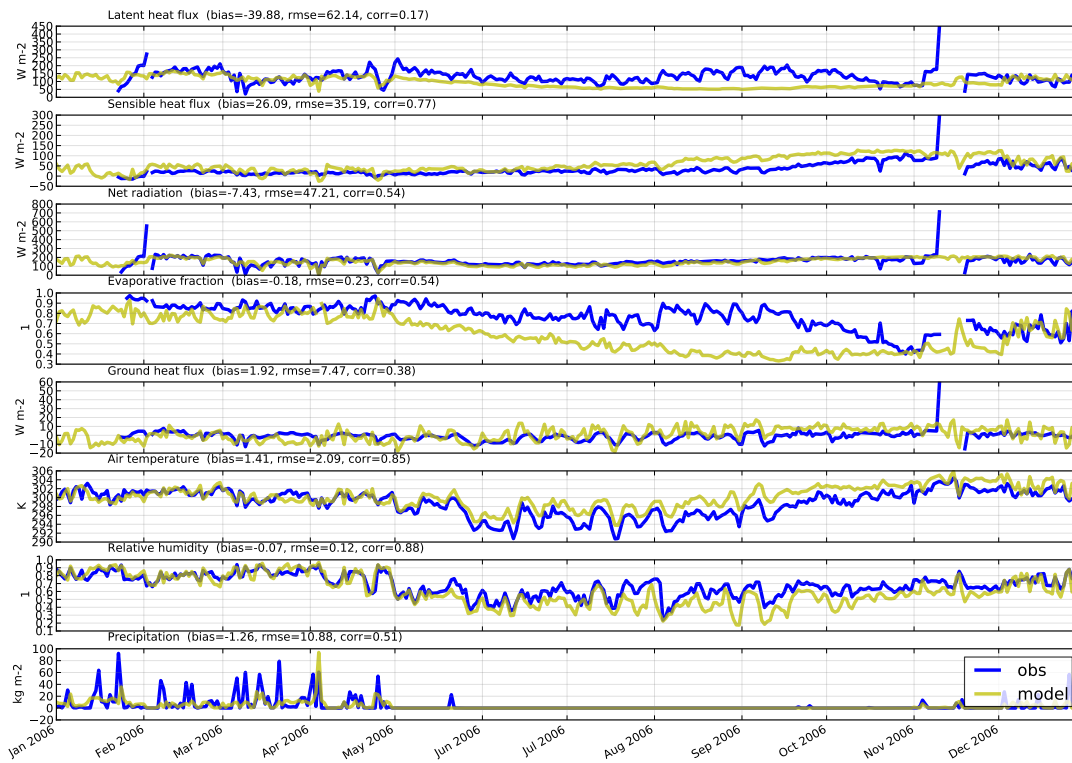
Howard Springs Australia -12.49N 131.15E (model -12.28N 131.48E)
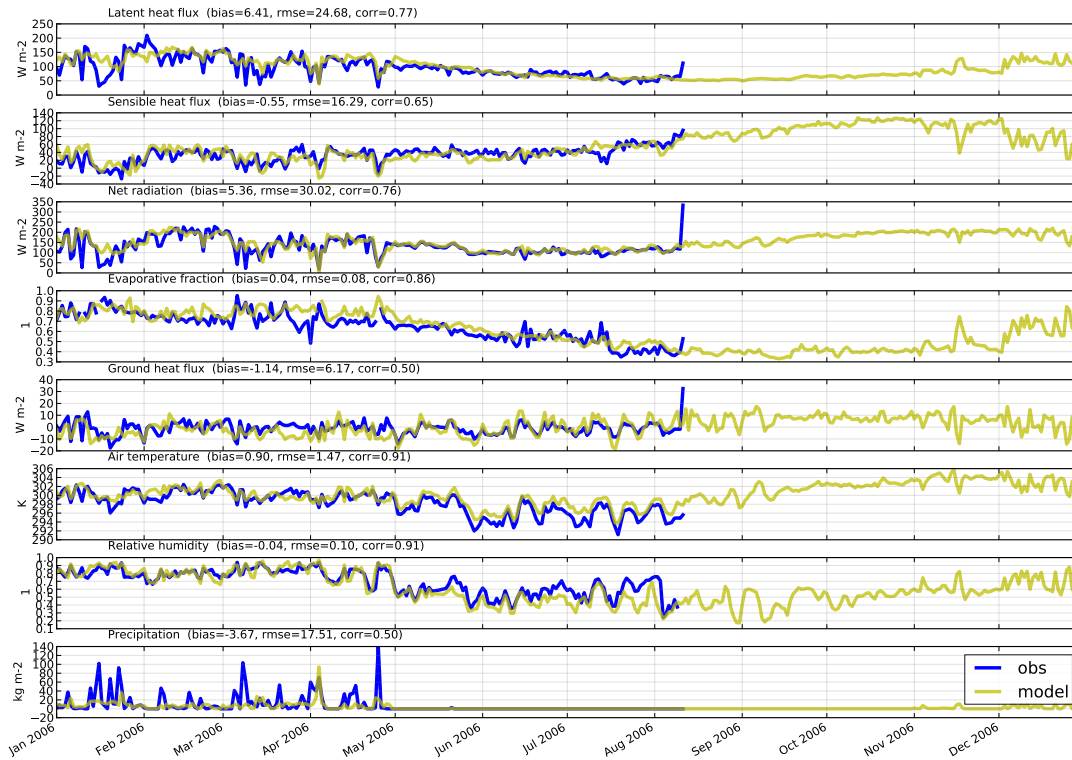
Tumbarumba Australia -35.66N 148.15E (model -35.44N 148.00E)
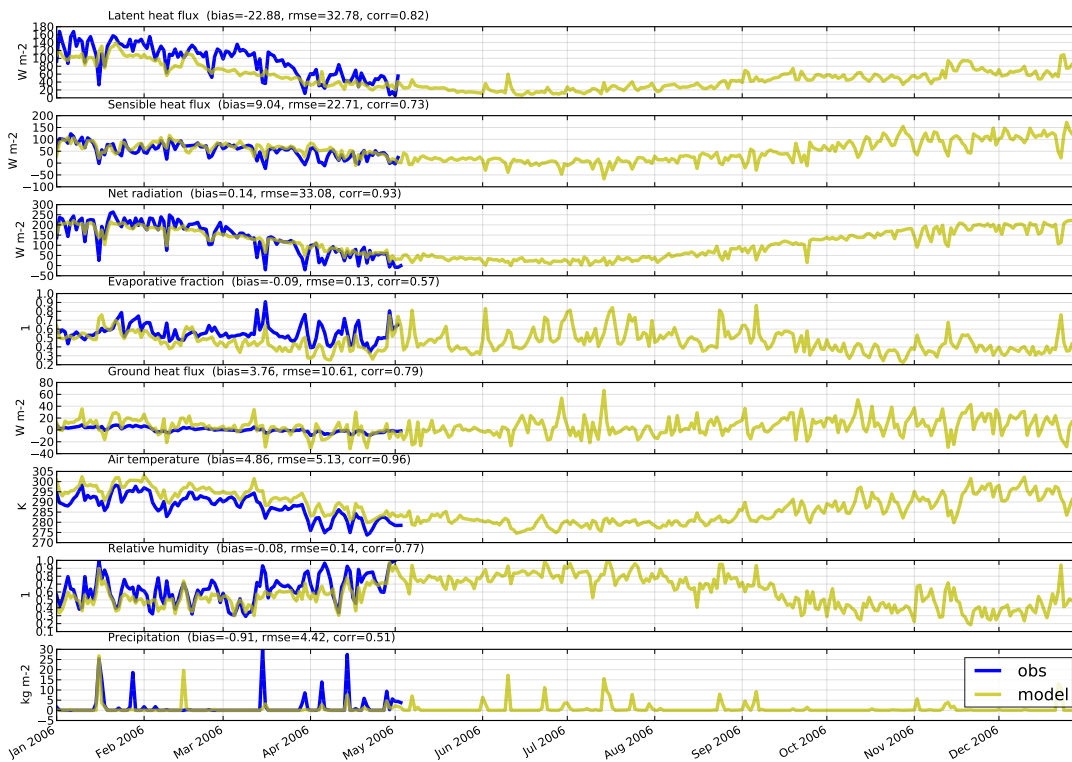
Wallaby Creek Australia -37.43N 145.19E (model -37.54N 145.00E)
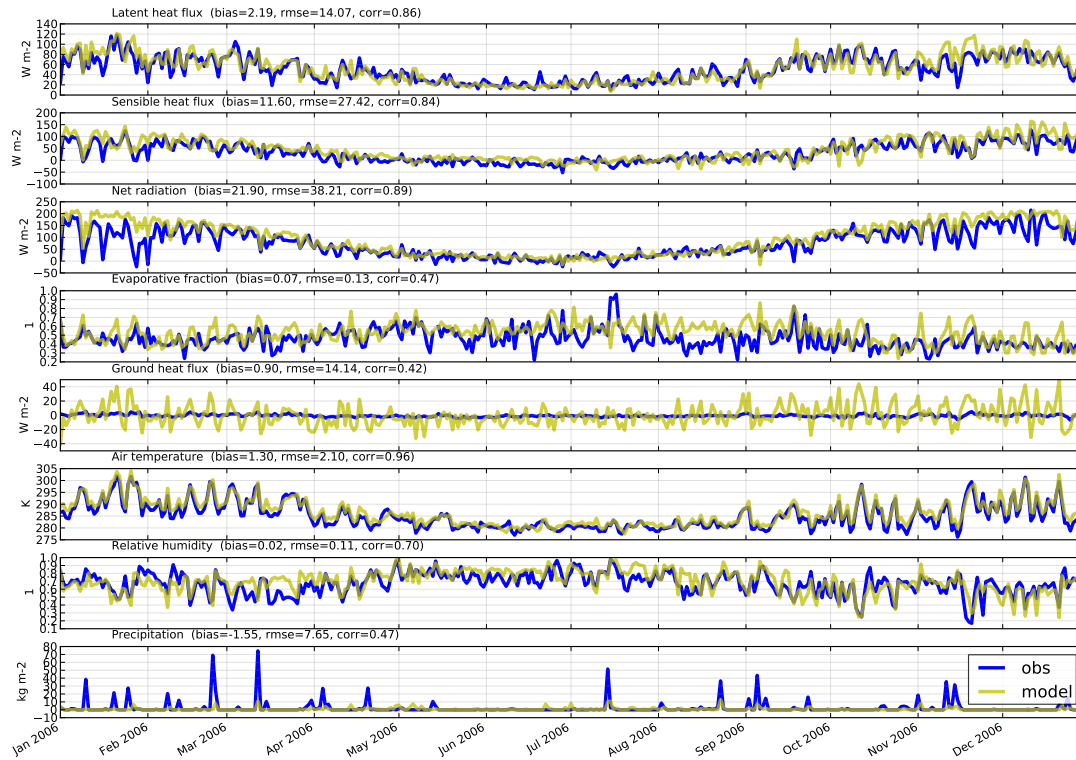
Quebec Boreal Cutover Site Canada 49.27N 285.96E (model 49.47N 286.00E)
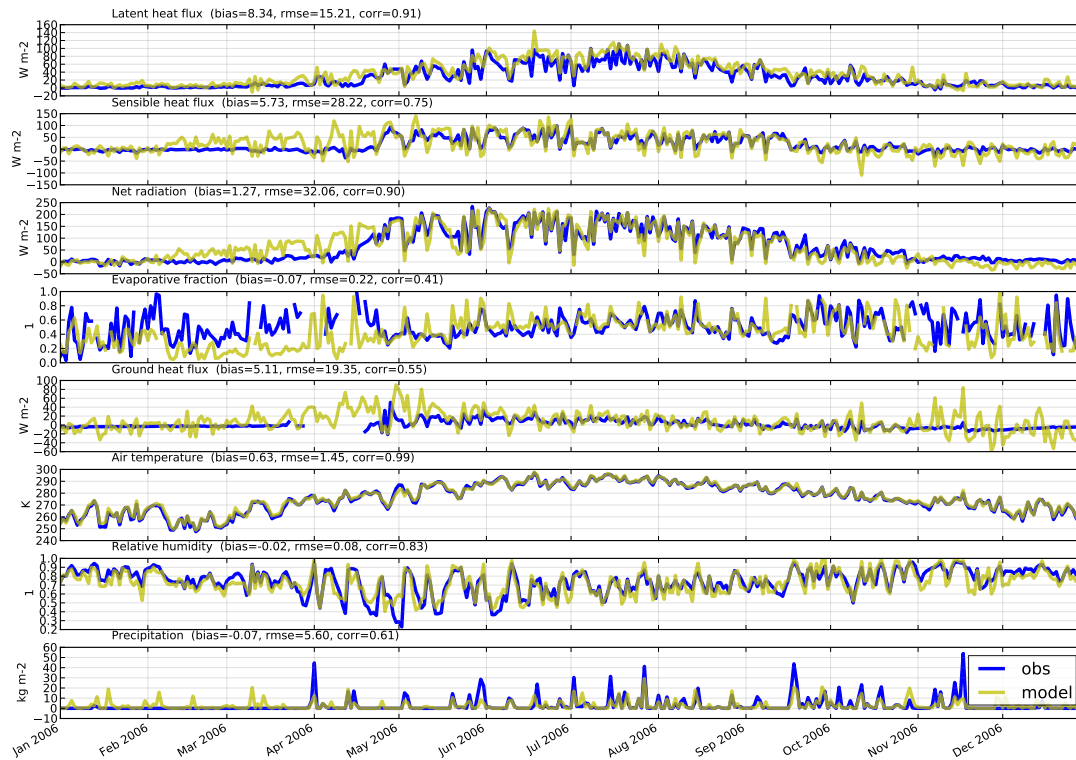
Quebec Mature Boreal Forest Site Canada 49.69N 285.66E (model 49.47N 286.00E)
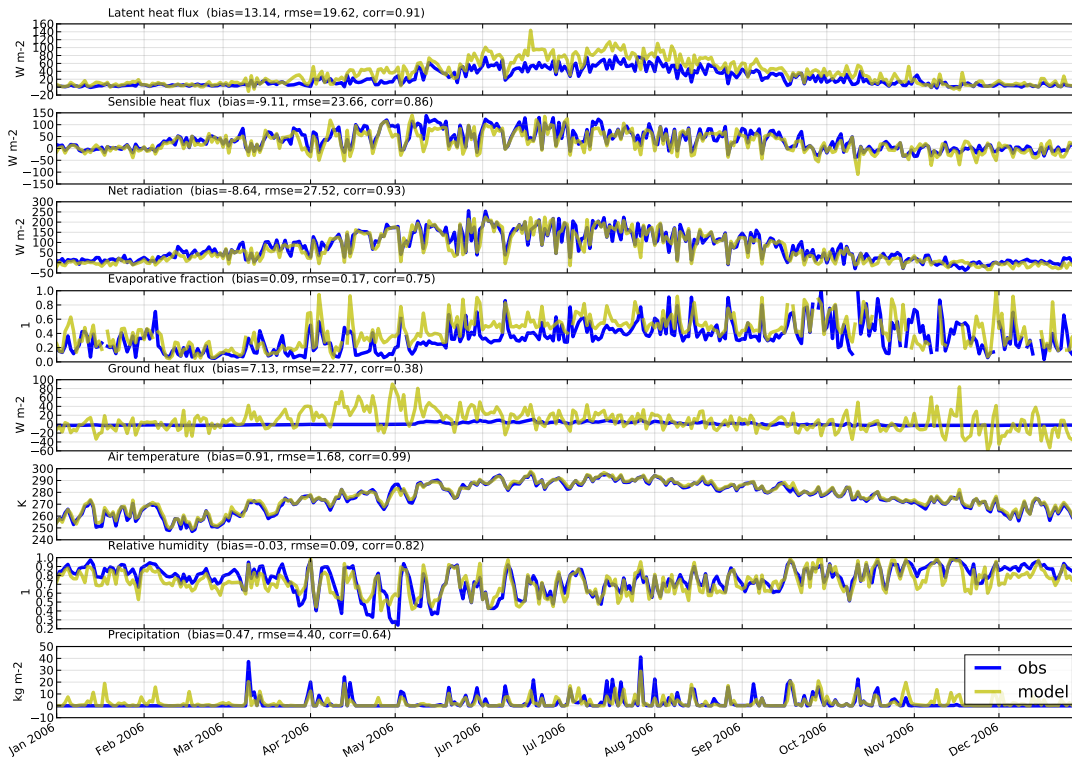
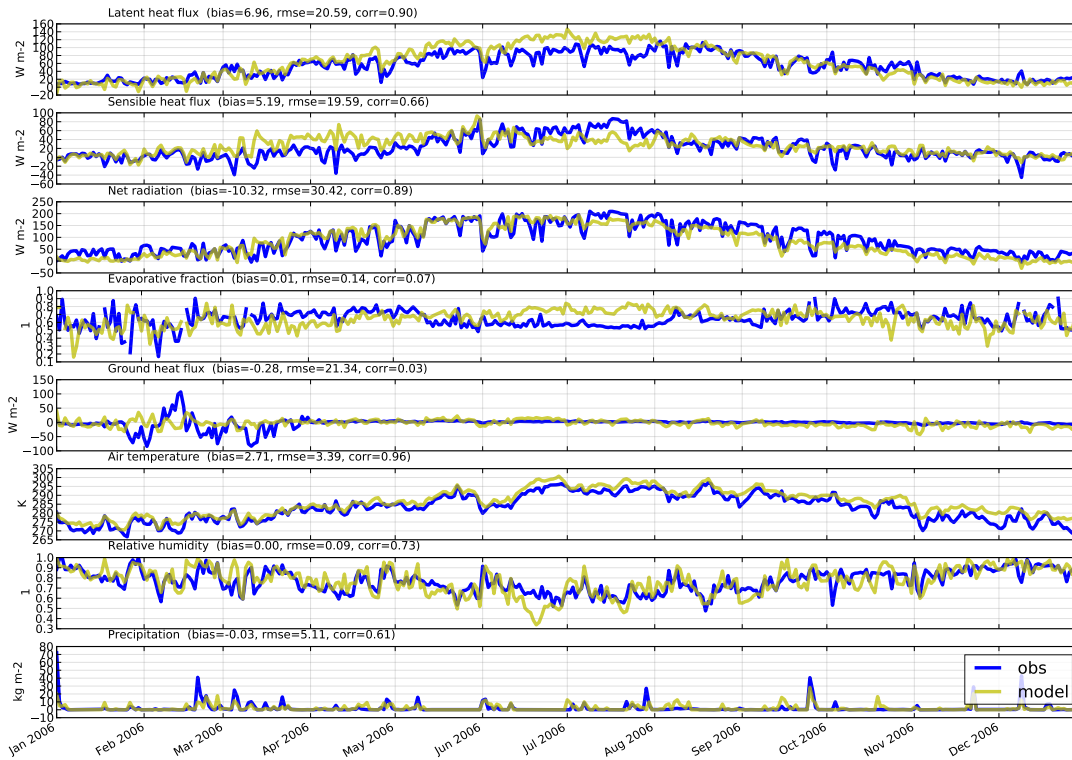Hyytiala Finland 61.85N 24.29E (model 62.11N 23.91E)
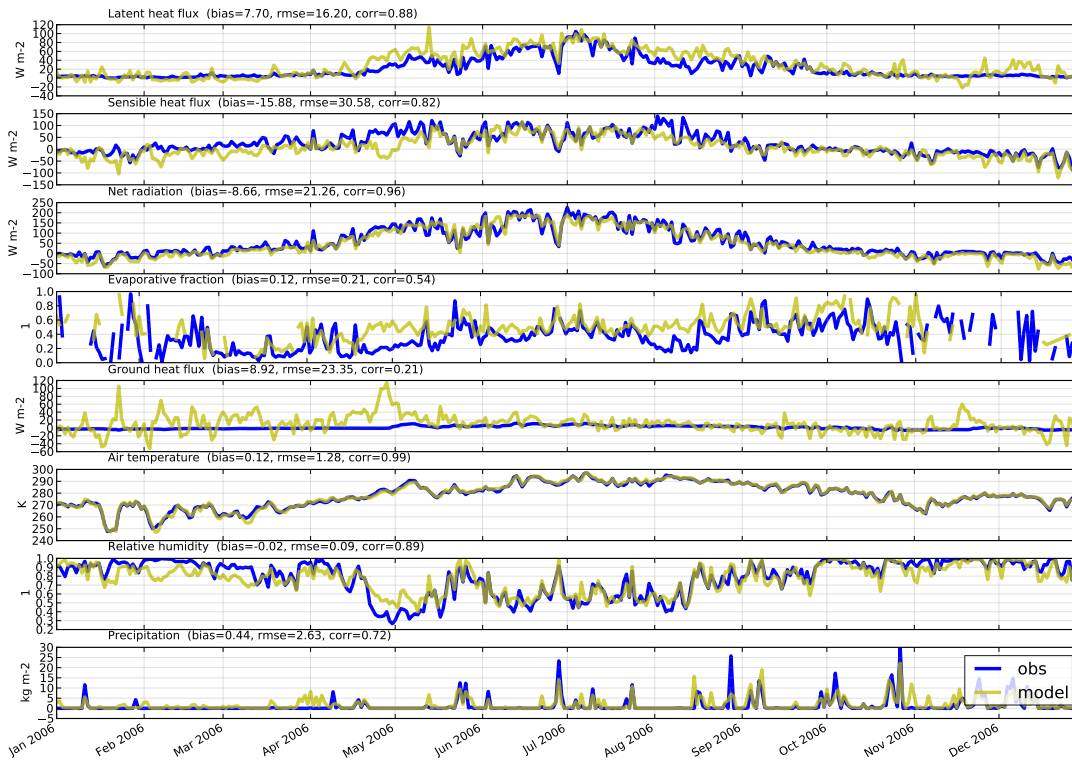
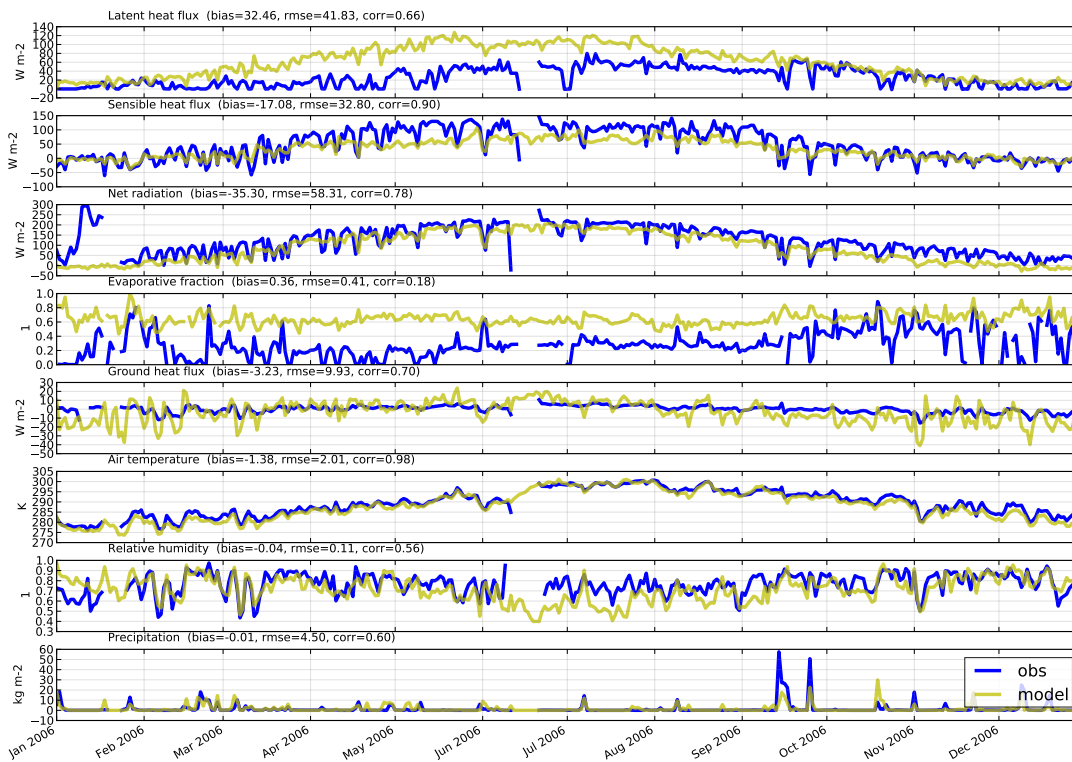Hesse Forest- Sarrebourg France 48.67N 7.06E (model 48.77N 7.00E)
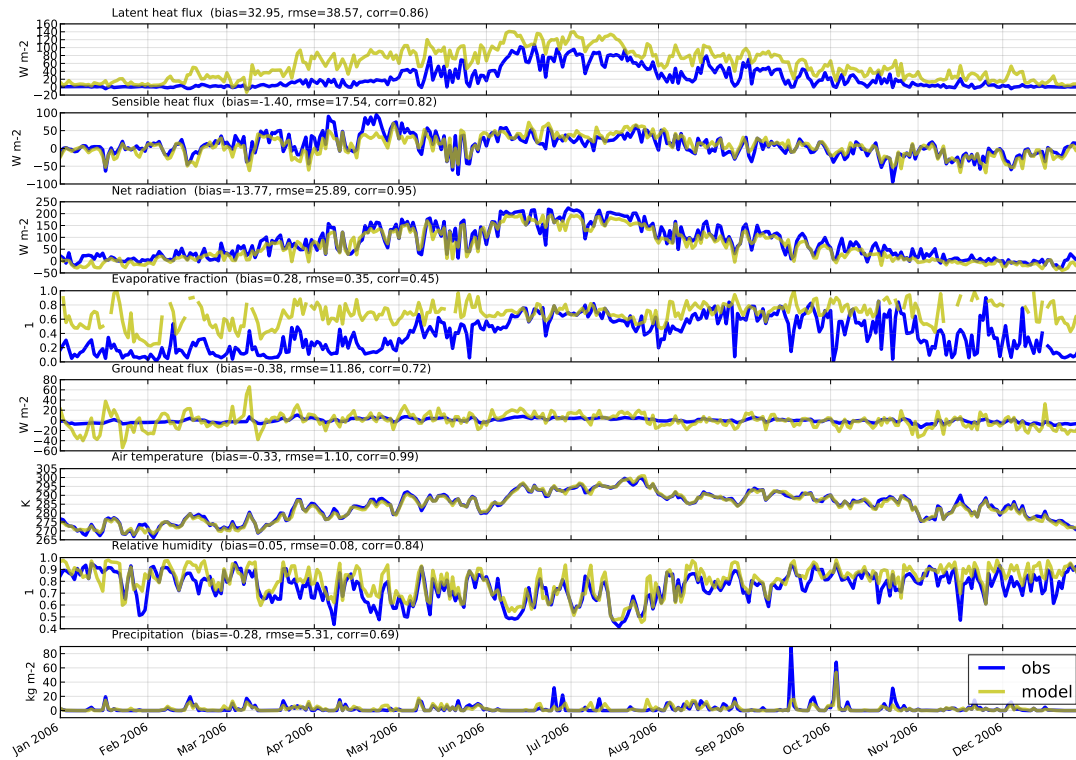
Le Bray France 44.72N 359.23E (model 44.56N 359.06E)
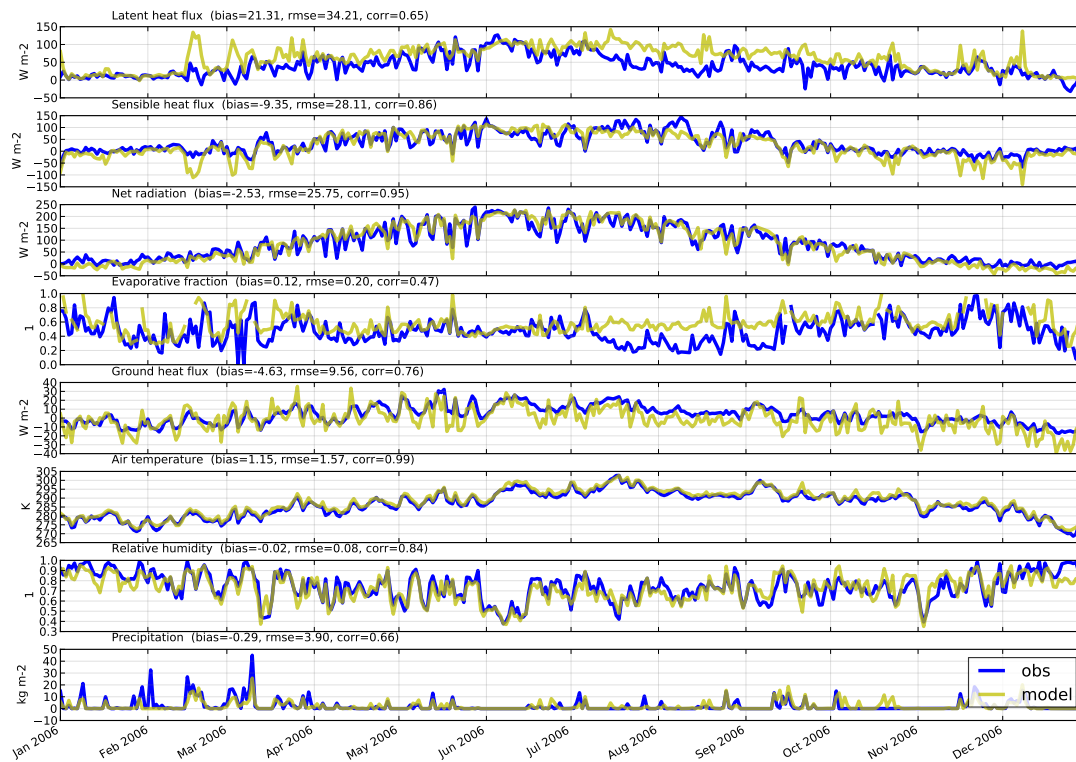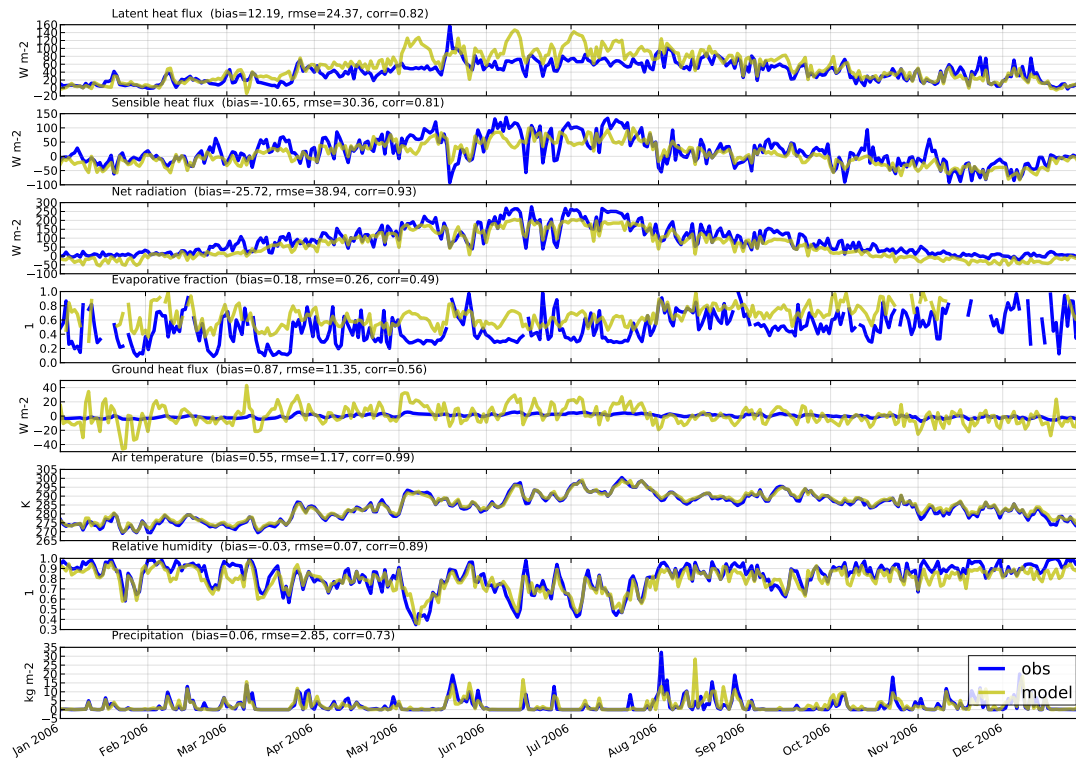
Amplero Italy 41.90N 13.61E (model 41.75N 13.50E)

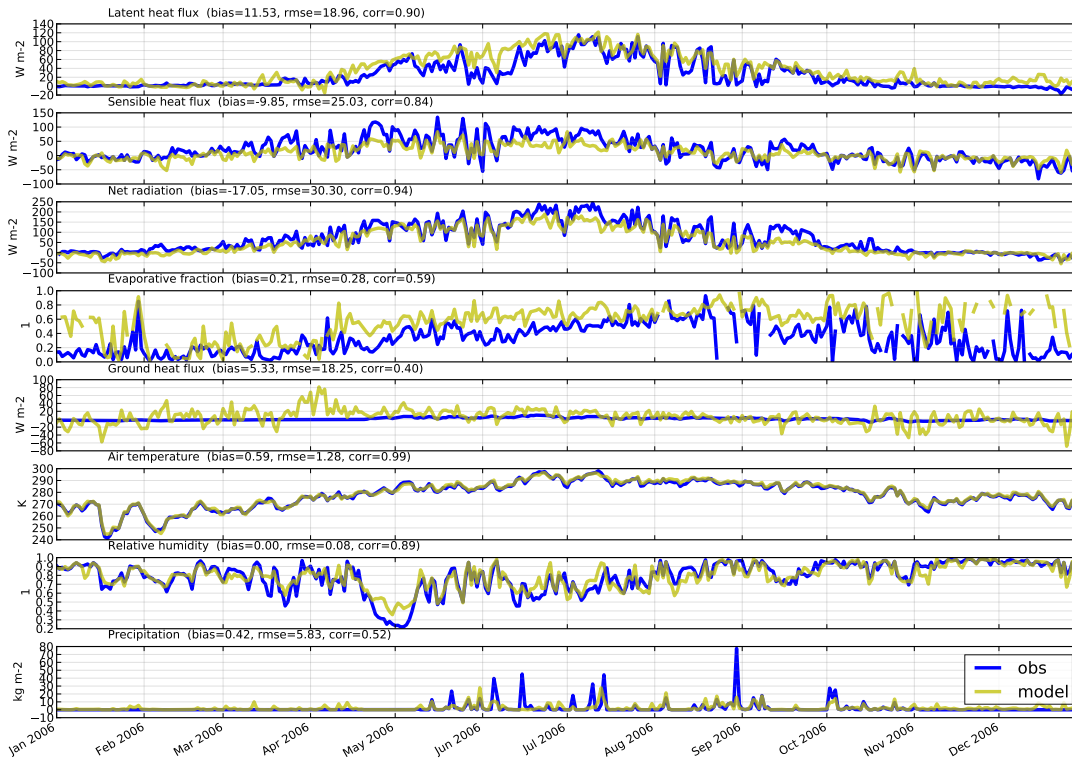Castelporziano Italy 41.71N 12.38E (model 41.75N 12.60E)

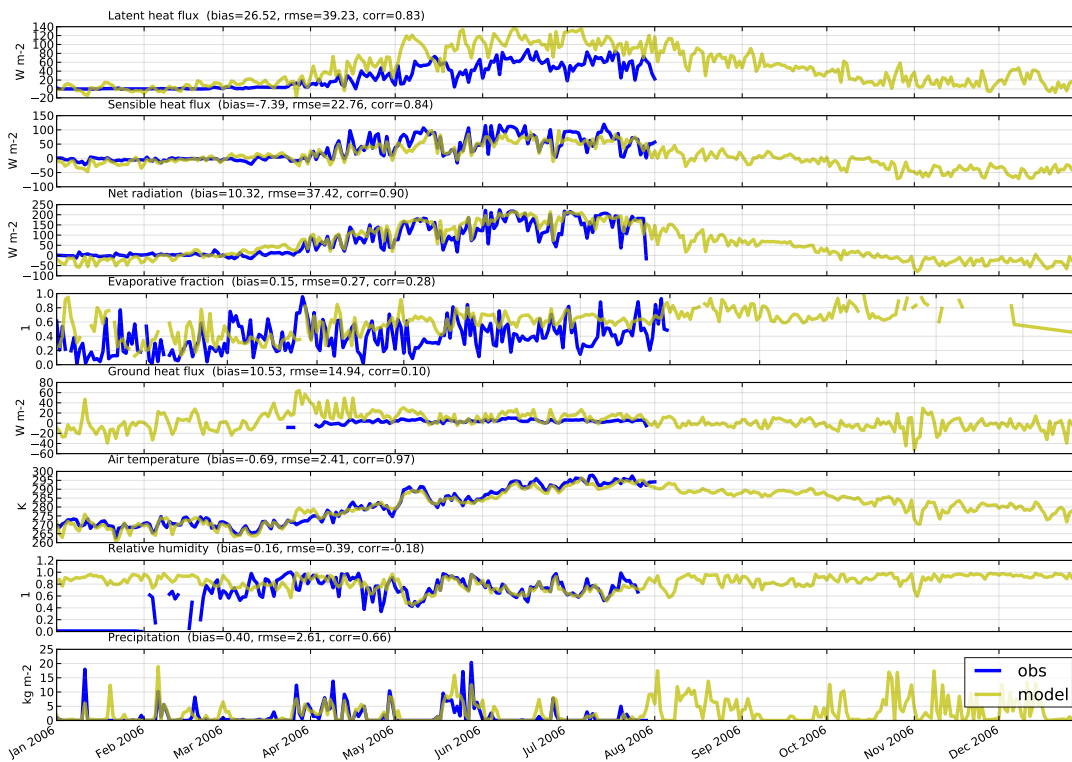Cabauw Netherlands 51.97N 4.93E (model 52.28N 5.00E)
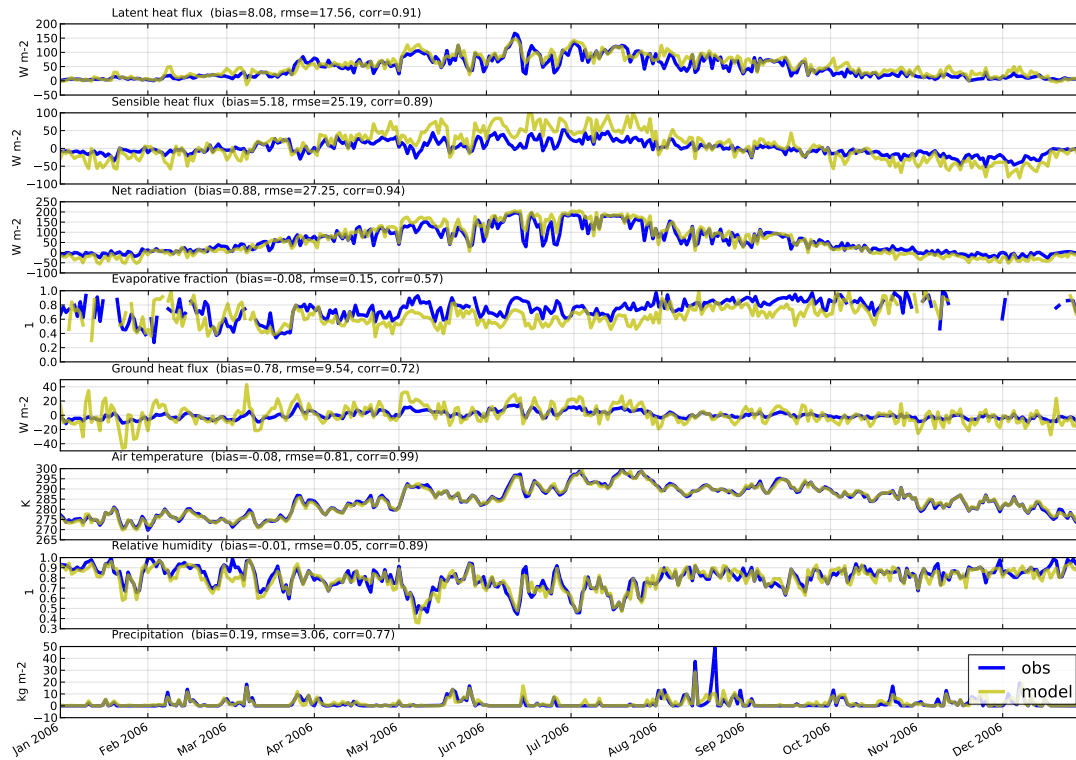
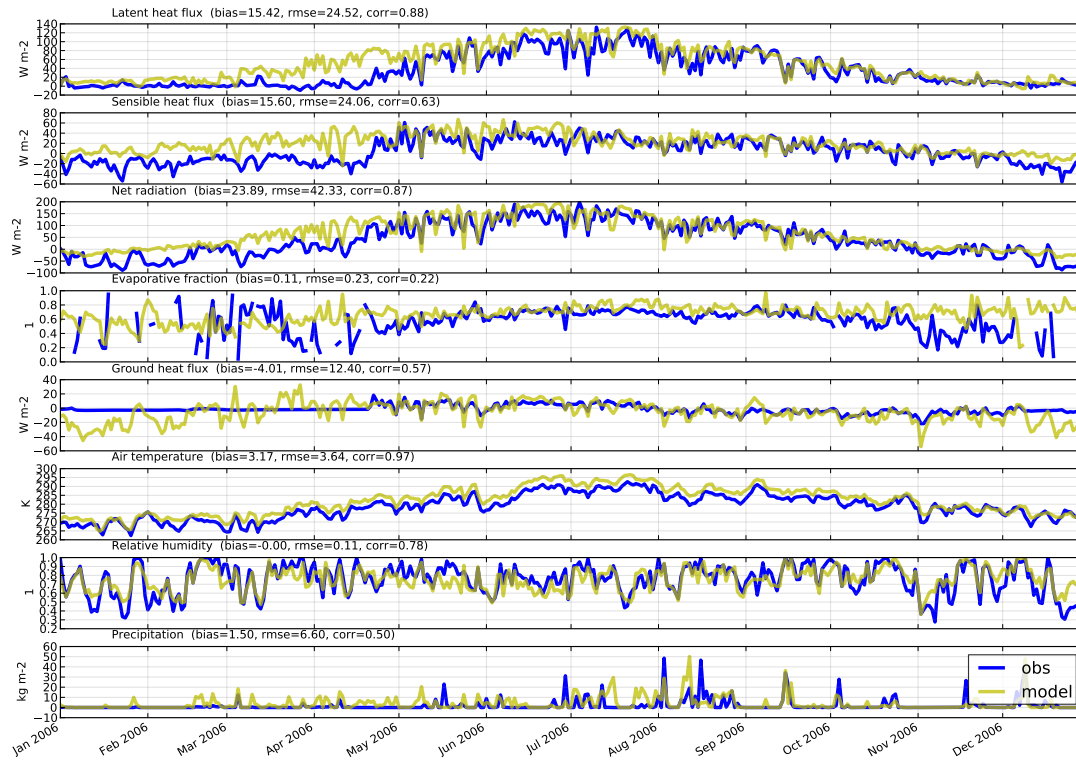Loobos Netherlands 52.17N 5.74E (model 52.28N 5.00E)

Fyodorovskoye wet spruce stand Russia 56.46N 32.92E (model 56.49N 32.62E)
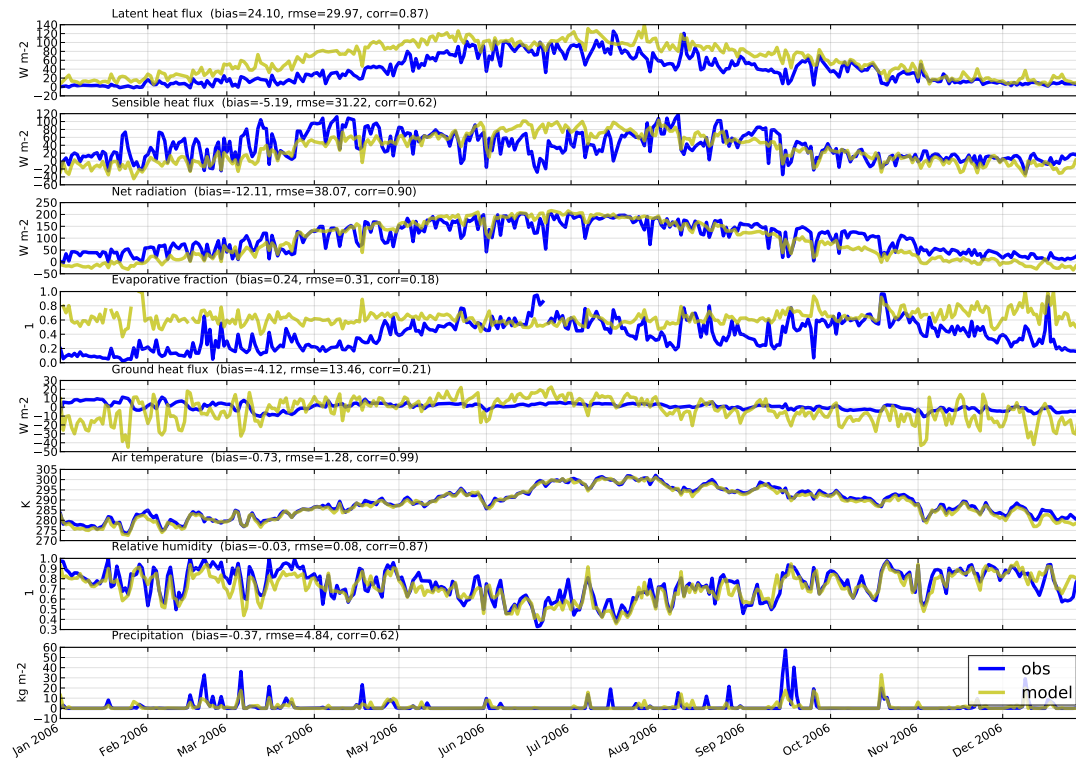

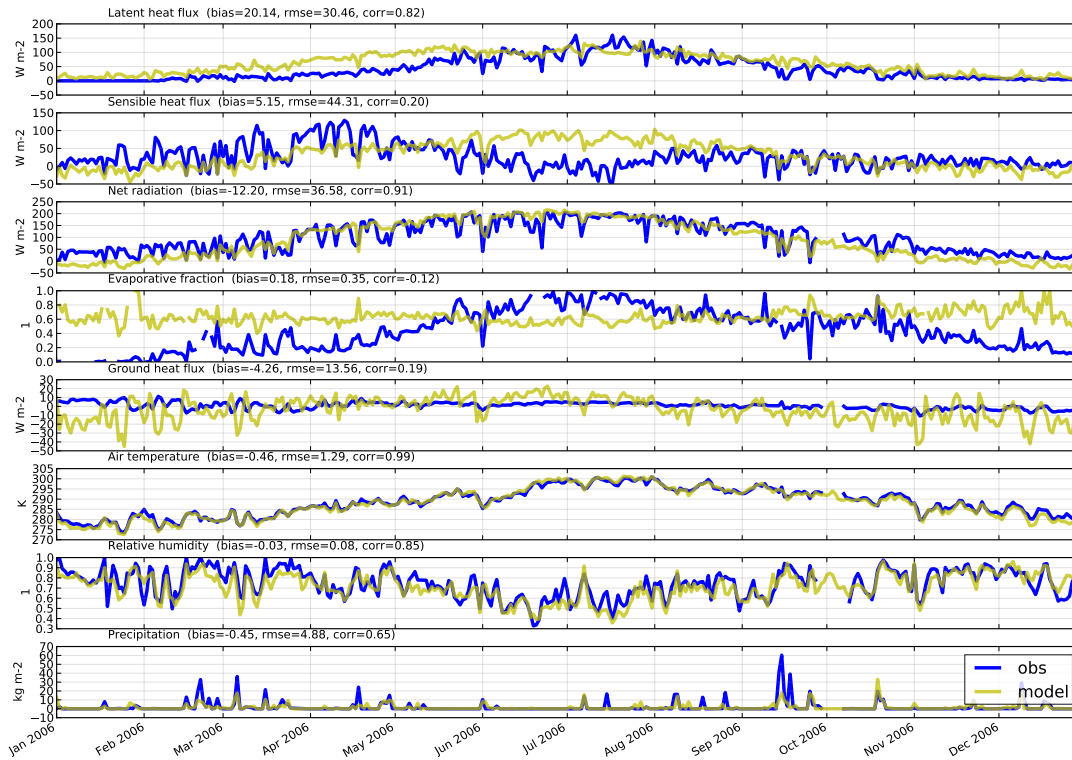Fajemyr (NECC) Sweden 56.27N 13.55E (model 56.49N 13.50E)

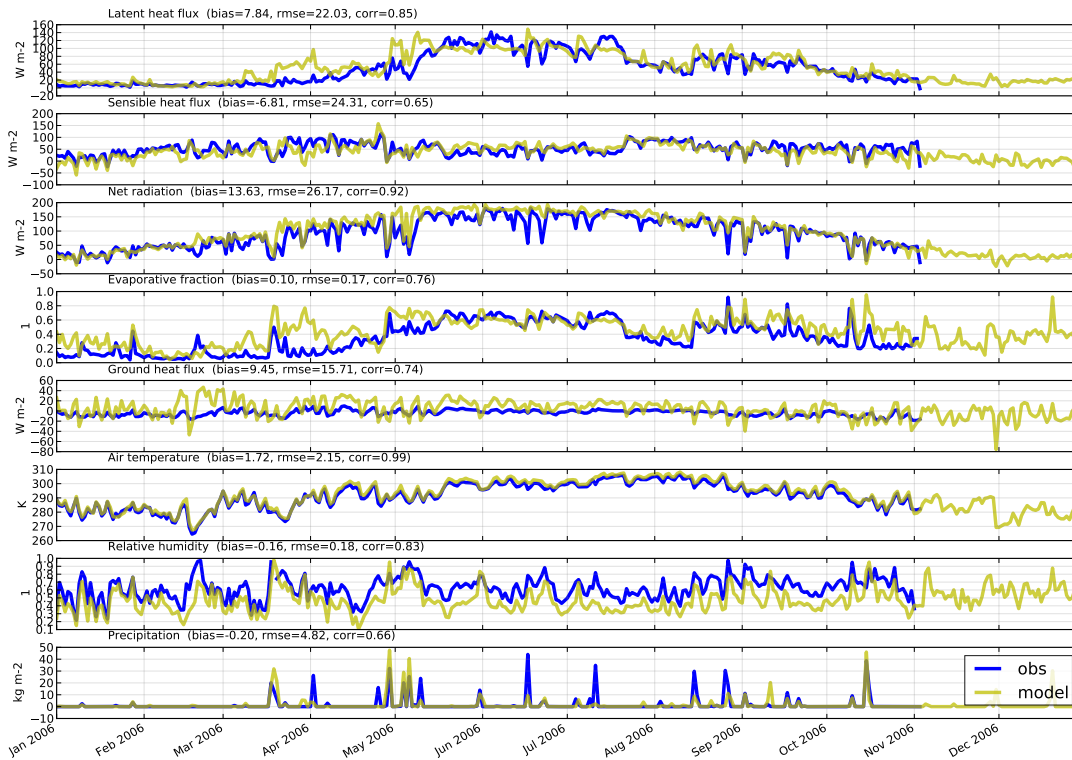Monte Bondone Italy 46.02N 11.05E (model 45.96N 11.52E)

Roccarespampani 1 Italy 42.41N 11.93E (model 42.46N 11.70E)
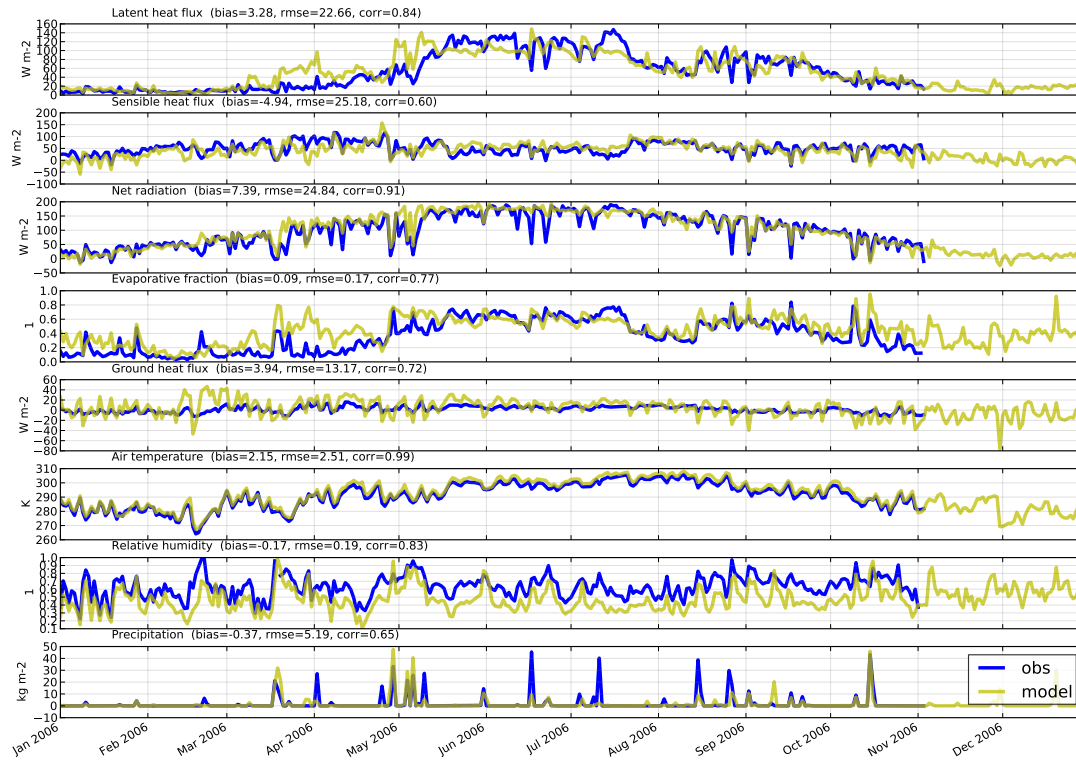
Roccarespampani 2 Italy 42.39N 11.92E (model 42.46N 11.70E)

OK - ARM Southern Great Plains burn site- Lamont USA 35.55N 261.96E (model 35.44N 261.60E)

OK - ARM Southern Great Plains control site- Lamont USA 35.55N 261.96E (model 35.44N 261.60E)

OK - ARM Southern Great Plains site- Lamont USA 36.61N 262.51E (model 36.84N 262.50E)
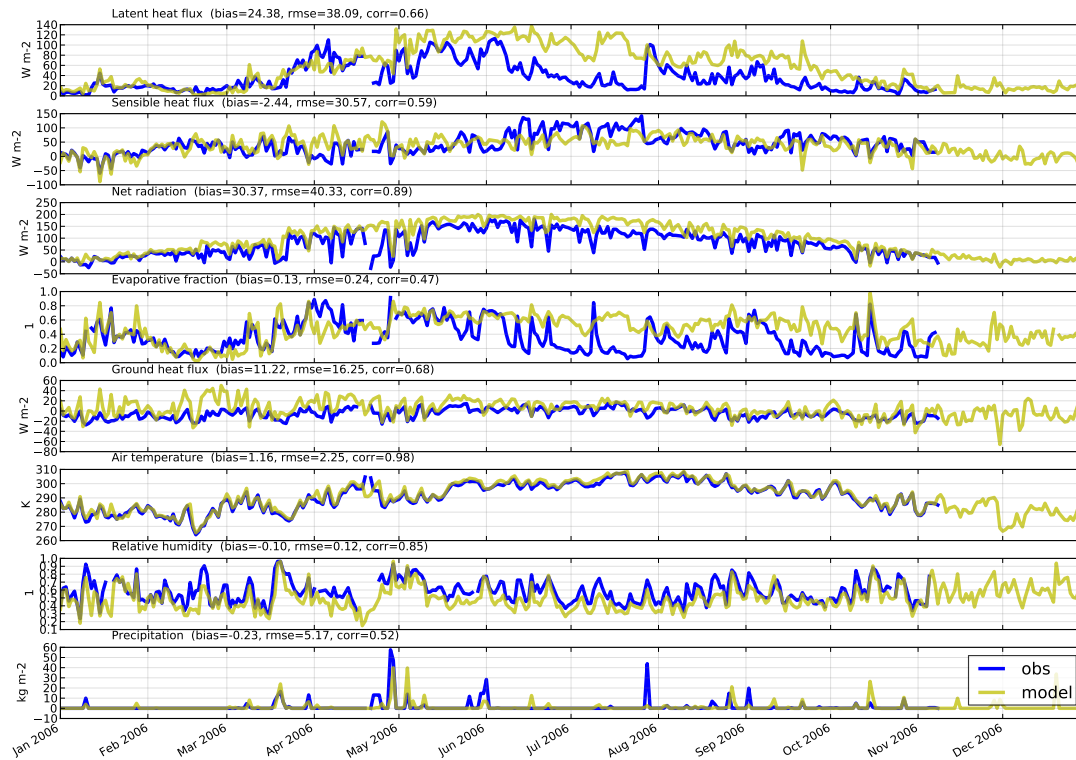
TX - Freeman Ranch- Mesquite Juniper USA 29.95N 262.00E (model 29.82N 261.75E)
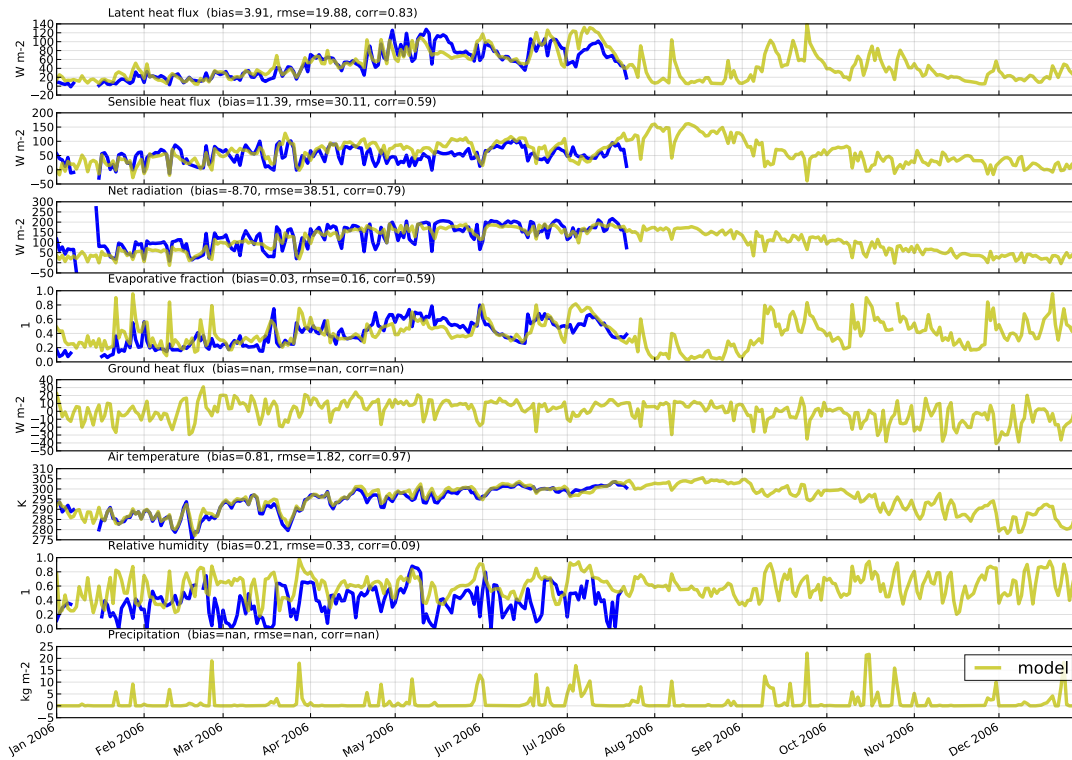
MA - Harvard Forest EMS Tower (HFR1) USA 42.54N 287.83E (model 42.46N 288.00E)
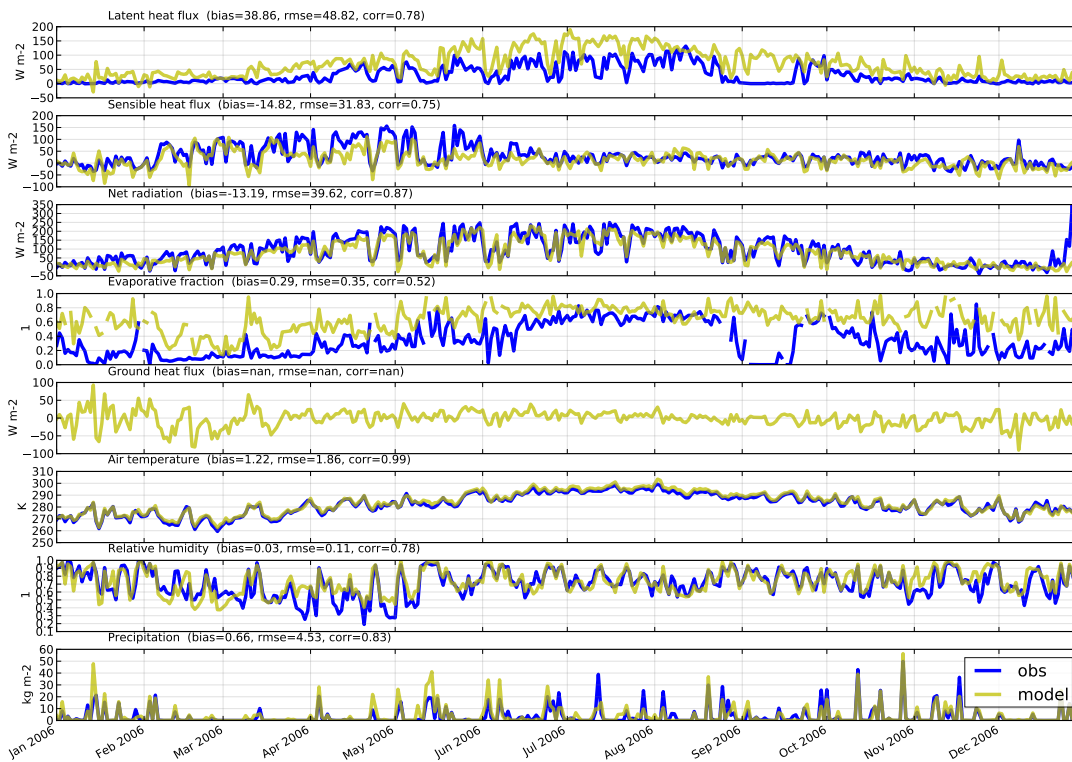
AZ - Santa Rita Mesquite USA 31.82N 249.13E (model 31.93N 249.00E)

MI - Sylvania Wilderness Area USA 46.24N 270.65E (model 45.96N 270.72E)

CA - Tonzi Ranch USA 38.43N 239.03E (model 38.25N 239.17E)

WI - Willow Creek USA 45.81N 269.92E (model 45.96N 269.76E)

B CEOP

The main objective of Coordinated Energy and water cycle Observations Project (CEOP) is to produce consistent research quality data for model development and evaluation. The various national agencies contribute to this project to develop archives of in-situ observations, model outputs and satellite remote sensing data. Detailed information ab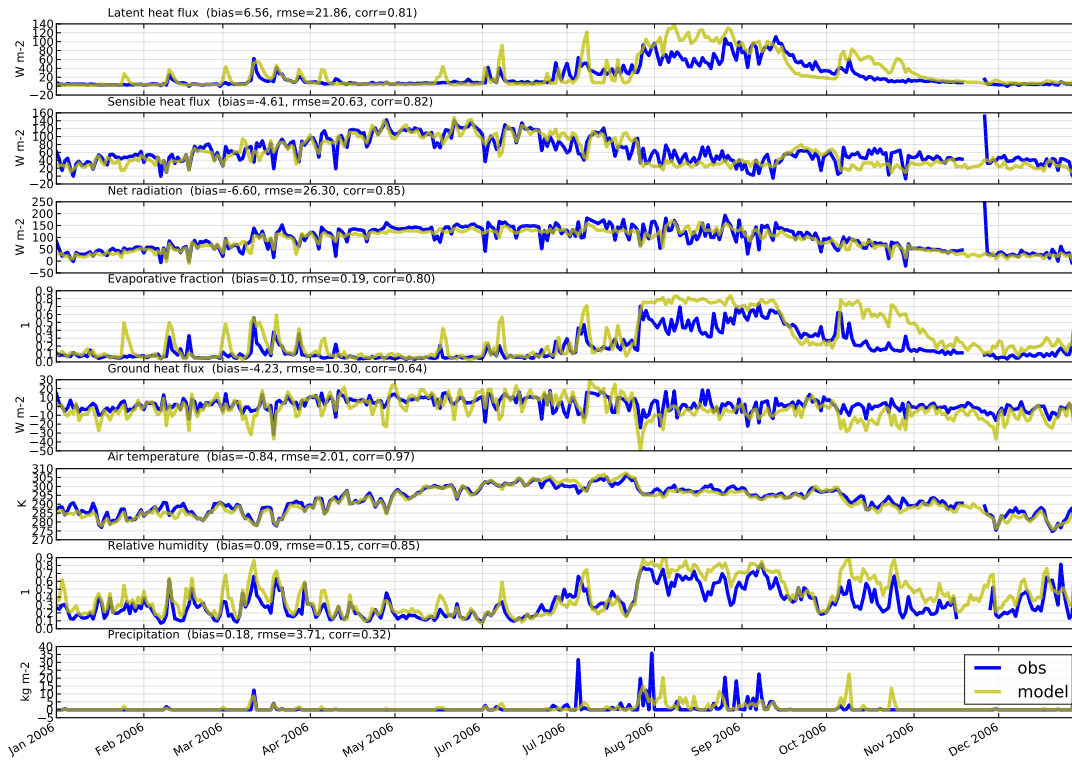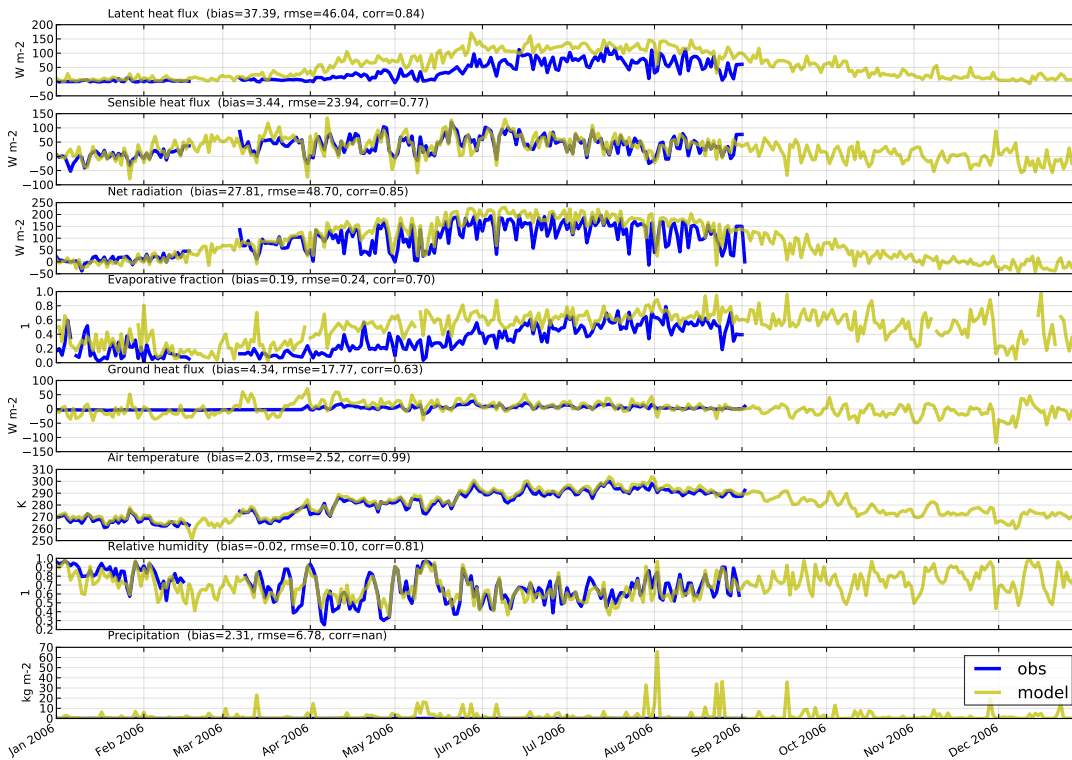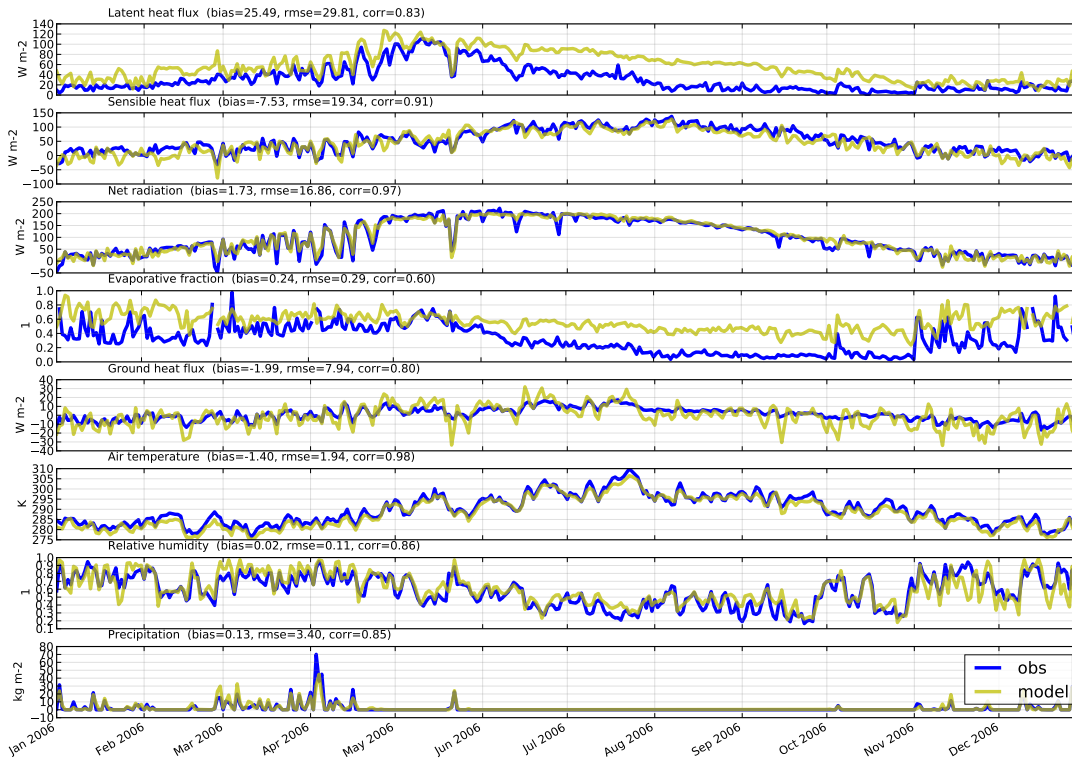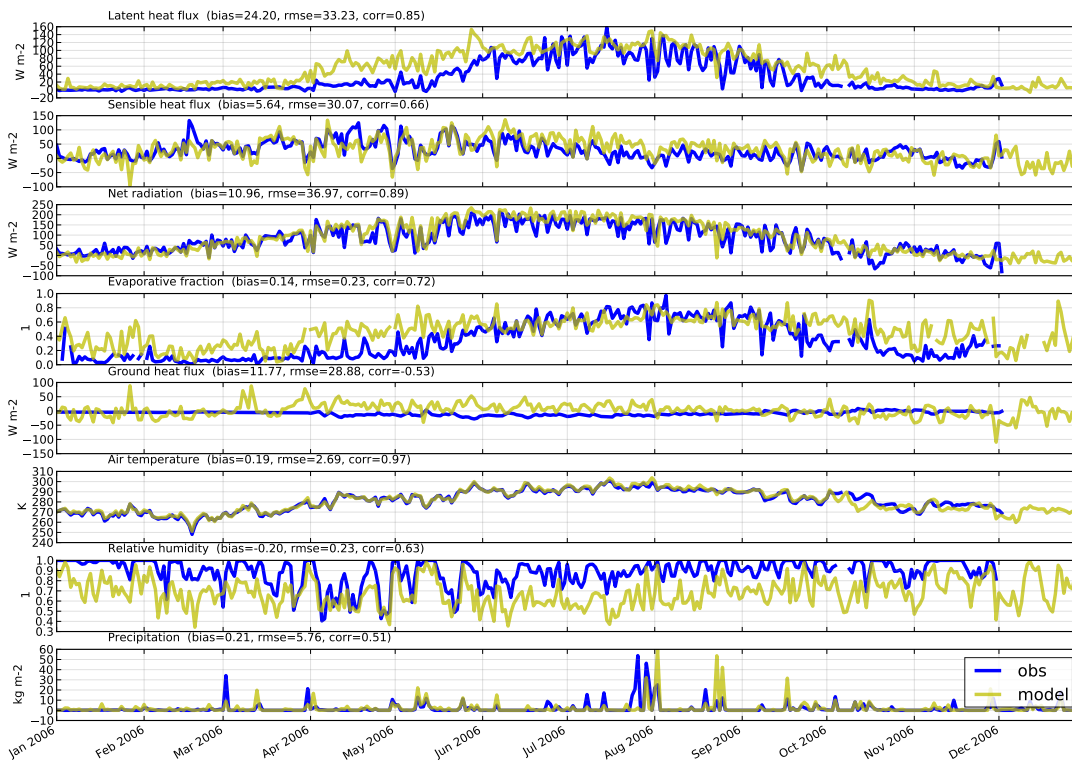out CEOP can be found at `www.ceop.net`. ECMWF is traditionally providing analysis and weather forecast data from the operational IFS (Integrated Forecasting System) model to the CEOP's model output archive. In the former Phase-I of the CEOP project this was done in a form of in-house ASCII data format. However, in the Phase-II, the decision has been made by CEOP to homogenize the model output archive using NetCDF format.

The following conventions have been agreed upon:

- there is only one file per station, per initialization time

- each file contains hourly time series of all variables at all levels

- forecast steps are until t+36h for the main initialization time 12UTC and until t+6 for minor initialization times 00, 06, 18UTC (note that since ECMWF does not provide forecasts from 06UTC and 18UTC, it was agreed that provided forecast steps will be extended up to t+12 for the minor initialization time of 00UTC)

- variable names comply with CF conventions

- naming convention for the output NetCDF files:

  ```
  <institute>_<model>_<site>_<init_time>.nc
  ```

The conversion procedure has been set up combining two scripts (first for retrieval and export of binary DDH files into ASCII format and second for conversion from ASCII files into NetCDF). The scripts and README file with usage instructions can be found in `/home/rd/pa8/ceop` directory. Up to now, the IFS DDH outputs from the years 2007, 2008 and 2009 have been successfully uploaded to the CEOP's model output archive. These data can be also find locally in `ec:/pa8/ceop`.

C In-situ land surface observation dataset

C.1 Introduction

In the past, the land surface modelling community was in lack of in-situ observations for model evaluation. However, more observation data became available recently via projects such as FLUXNET or CEOP that provide quality assessed flux tower measurements from different geographic regions and climate zones. The objective of this work was to collect data from different sources and to compile them into an unified observation dataset. Thus collected and homogenized observations can be used for a routine 'benchmarking' of the land surface parametrization schemes.

C.2   Specification

For this purpose, NetCDF was chosen as the most suitable format for the following reasons: a) it is widely used in research community for data exchange; b) there are many utilities to manipulate NetCDF files; c) the content of NetCDF files is self-documenting. The primary standard used for description of variables in the NetCDF files is the climate and forecast (CF) metadata convention (http://cf-pcmdi.llnl.gov).

Observations from different sources were aggregated into 1 hourly time series (in UTC time). The following is an example header of a NetCDF file:

```
netcdf fluxnet.nl-ca1.2004 {
dimensions:
    time = UNLIMITED ; // (8784 currently)
    lat = 1 ;
    lon = 1 ;
variables:
    double time(time) ;
        time:standard_name = "time" ;
        time:long_name = "Time" ;
        time:units = "hours since 2004-01-01 00:00:00" ;
    float lat(lat) ;
        lat:standard_name = "latitude" ;
        lat:long_name = "Latitude" ;
        lat:units = "degrees_north" ;
    float lon(lon) ;
        lon:standard_name = "longitude" ;
        lon:long_name = "Longitude" ;
        lon:units = "degrees_east" ;
    float LH(time, lat, lon) ;
        LH:standard_name = "surface_upward_latent_heat_flux" ;
        LH:long_name = "Latent heat flux" ;
        LH:units = "W m-2" ;
        LH:_FillValue = -9999.f ;
        LH:cell_methods = "time: mean" ;
    float SH(time, lat, lon) ;
        SH:standard_name = "surface_upward_sensible_heat_flux" ;
        SH:long_name = "Sensible heat flux" ;
        SH:units = "W m-2" ;
        SH:_FillValue = -9999.f ;
        SH:cell_methods = "time: mean" ;
    float NR(time, lat, lon) ;
        NR:standard_name = "surface_net_downward_radiative_flux" ;
        NR:long_name = "Net radiation" ;
        NR:units = "W m-2" ;
        NR:_FillValue = -9999.f ;
        NR:cell_methods = "time: mean" ;
    float NEE(time, lat, lon) ;
```

```
        NEE:standard_name = "net_ecosystem_exchange_of_carbon" ;
        NEE:long_name = "Net ecosystem exchange" ;
        NEE:units = "kg m-2 s-1" ;
        NEE:_FillValue = -9999.f ;
        NEE:cell_methods = "time: mean" ;

// global attributes:
    :site_name = "Cabauw Netherlands" ;
    :latitude = 51.97 ;
    :longitude = 4.93 ;
    :elevation = 0.7 ;
    :vegetation_type = "Grasslands" ;
    :history = "2010-10-07 19:10 GMT created by fluxnet2netcdf.py" ;
    :Conventions = "CF-1.0" ;
```

Each file contains one site-year of data, which means that there is only one file per site per year. The site ancillary data are included in the global attributes: `site_name`, `latitude`, `longitude`, `elevation` (not available for all sites) and `vegetation_type` (vegetation type according to IGBP classification, only available for FLUXNET sites).

The variables that are collected in the dataset include soil moisture, soil temperature, surface energy fluxes, precipitation, snow depth, surface carbon fluxes and near surface meteorological parameters. However, the number of available variables varies from station to station. The complete list of all variables can be found in the Table 4.

C.3   Dataset availability

The observation dataset is kept in the ECFS. The original files, as they were obtained (usually in ASCII format), can be found in:

`ec:/pa8/observations/original/<source>`

and files converted into NetCDF format in:

`ec:/pa8/observations/netcdf/<year>`

File naming convention was chosen to encompass the most relevant information:

`<source>.<site_id>.<year>.nc`

where:

source the source of data, this is a name of observation network or name of a project

site_id the site identifier, we keep the same identifier as defined in the original project for easier reference to the original data

year year of available data

examples:

`fluxnet.nl-ca1.2008.nc, smosmania.cdm.2009.nc, berms.sk-oa.2006.nc`

Var. name	Standard name (CF convention)	Units	Long name
NEE	net_ecosystem_exchange_of_carbon	kg m^{-2} s^{-1}	Net ecosystem exchange
GPP	gross_primary_productivity_of_carbon	kg m^{-2} s^{-1}	Gross primary production
Reco	ecosystem_respiration_carbon_flux	kg m^{-2} s^{-1}	Ecosystem respiration
Epot	water_potential_evaporation_flux	kg m^{-2} s^{-1}	Potential evapotranspiration
LH	surface_upward_latent_heat_flux	W m^{-2}	Latent heat flux
SH	surface_upward_sensible_heat_flux	W m^{-2}	Sensible heat flux
G	downward_heat_flux_at_ground_level_in_soil	W m^{-2}	Ground heat flux
VPD	water_vapor_saturation_deficit_in_air	Pa	Vapor pressure deficit
T	air_temperature	K	Air temperature
SKT	surface_brigtness_temperature	K	Skin temperature
DPT	dew_point_temperature	K	Dew point temperature
P	precipitation_amount	kg m^{-2}	Precipitation
WS	wind_speed	m s^{-1}	Wind speed
WD	wind_from_direction	degree	Wind direction
NR	surface_net_downward_radiative_flux	W m^{-2}	Net radiation
DLWR	surface_downwelling_longwave_flux_in_air	W m^{-2}	Downward longwave radiation
ULWR	surface_upwelling_longwave_flux_in_air	W m^{-2}	Upward longwave radiation
DSWR	surface_downwelling_shortwave_flux_in_air	W m^{-2}	Downward shortwave radiation
USWR	surface_upwelling_shortwave_flux_in_air	W m^{-2}	Upward shortwave radiation
R	relative_humidity	1	Relative humidity
Q	specific_humidity	1	Specific humidity
DR	subsurface_runoff_amount	kg m^{-2}	Drainage
CO2	mole_fraction_of_carbon_dioxide_in_air	1	Carbon dioxide concentration
H2O	mole_fraction_of_water_vapor_in_air	1	Water vapor concentration
SD	surface_snow_thickness	m	Snow depth
SMVF	volume_fraction_of_condensed_water_in_soil	1	Soil moisture volume fraction
SP	surface_air_pressure	Pa	Surface pressure
WSG	wind_speed_of_gust	m s^{-1}	Wind gust
SDWE	lwe_thickness_of_surface_snow_amount	m	Snow depth of water equivalent
ST	soil_temperature	K	Soil moisture
DPAR	surface_downwelling_photosynthetic_photon_flux_in_air	mol m^{-2} s^{-1}	Incoming PAR
UPAR	surface_upwelling_photosynthetic_photon_flux_in_air	mol m^{-2} s^{-1}	Outgoing PAR

Table 4: List of variables in the observation dataset including their standard names, units and long names.

C.4 FLUXNET

The FLUXNET project standardizes and harmonizes observations from eddy covariance flux tower stations from regional networks from around the world: CarboeuropeIP, AmeriFlux, Fluxnet-Canada, LBA, Asiaflux, Chinaflux, USCCC, Ozflux, Carboafrica, Koflux, NECC, TCOS-Siberia and Afriflux. The FLUXNET LaThuile dataset consist of quality assessed and gap-filled records of carbon dioxide, water vapor and energy fluxes.
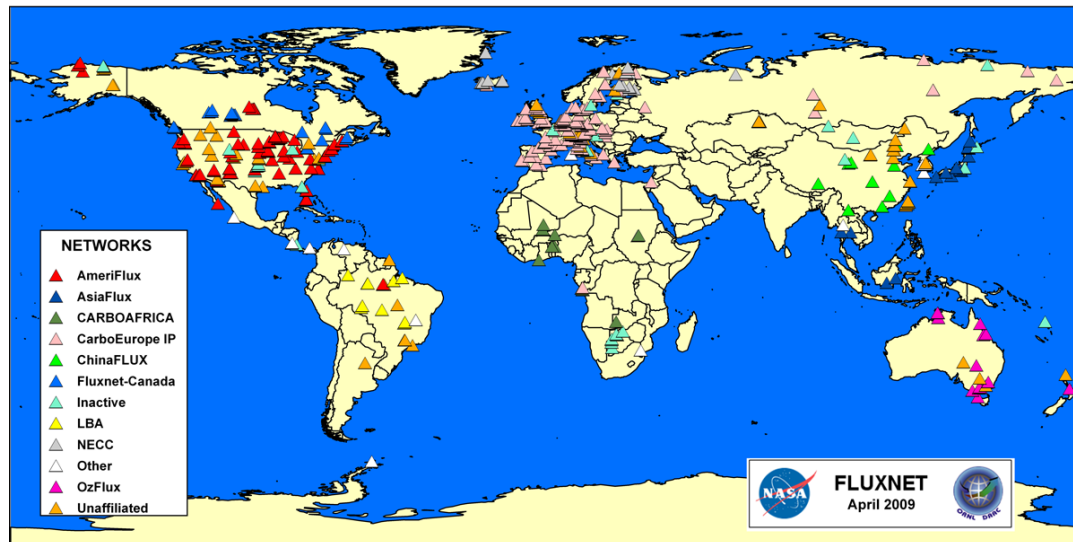

Figure 6: Distribution of tower sites within the regional networks.

The LaThuile FLUXNET dataset comes with 288 site-years of data in 61 stations from Austria, Australia, Botswana, Canada, Finland, France, Indonesia, Israel, Italy, Netherlands, Russia, Sweden, Switzerland, United States, Vanautu and South Africa.

The original LaThuile data are sampled with 30min intervals thus the data were averaged over one hour. The data are also gapfilled with four quality levels (0: data are original, 1: gap filled high quality, 2: gapfilled medium quality, 3: gapfilled low quality). In the conversion process we ignore all values with low quality flag and encode them as missing values.

The information on the height of installed sensors is not available (also the height of sensors is not standardised due to different height of canopy at the measurement sites)

C.5 SMOSMANIA

The aim of SMOSMANIA project (Soil Moisture Observing System - Meteorological Automatic Network Integrated Application) is to create network of automated soil moisture stations in South-Western France. The network consists of 12 stations creating Atlantic-Mediterranean gradient (see Figure 7). Soil moisture profiles are measured at 4 depths (-5, -10, -20, -30 cm). The acquisition of data started in 2007 and continues until present.
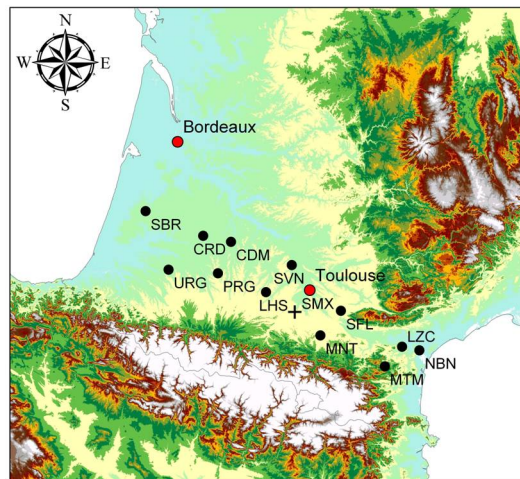
Figure 7: SMOSMANIA soil moisture measurements Network in the South West of France (Albergel et al., Hydrol. Earth Syst. Sci. Discuss., 5, 2008, http://www.hydrol-earth-syst-sci-discuss.net/5/1603/2008/.; Calvet et al., IGARSS, Barcelona, Spain, 23-28 July 2007, 1196-1199, doi:10.1109/IGARSS.2007.4423019, 2007.).