Tests with the ECOCLIMAP physiographic database in HIRLAM

T. Fernández and E. Rodríguez
Spanish Meteorological Institute (INM), P.O. Box 275, 28070 Madrid, Spain

1 Introduction

The state-of-art atmospheric models at all scales describe land surface processes with an increasing level of complexity. Richarson (1922), in his seminal book on NWP, already mentioned the importance of vegetation to model the exchange of water between land and atmosphere. Many studies have demonstrated that both soil water content and vegetation parameters have a strong impact on the heat fluxes coupling the atmosphere and the land surface (see, *e.g.*, (Rodriguez-Camino and Avissar, 1998) and the references therein contained). The number of vegetation parameters depends logically on the features of each Soil-Vegetation-Atmosphere Tranfer (SVAT) scheme considered. However, even the simplest SVAT schemes are based on at least the following vegetation parameters: vegetation cover, leaf area index (LAI), vegetation albedo, minimum stomatal resistance, root depth, vegetation roughness length, etc. The importance of the different parameters for the land-surface processes is highly variable. Of course, the vegetation cover will primarily determine which pathway will be predominantly used to evaporate, either direct evaporation from the uppermost surface soil layer or vegetative evapotranspiration mainly from deeper layers. The LAI parameter gives some sort of effective leaf size and, considering that the transpiration process is roughly proportional to the leaf surface involved, it is also an important parameter. Some of the other mentioned parameters either are difficult to estimate from global direct measurements or are not so crucial for the SVAT schemes.

Traditionally, two global datasets of surface parameters have been used for climate and NWP models: the one produced by Wilson and Henderson-Sellers (1985) (WHS) and the one from the International Satellite Land Surface Climatology Project (ISLSCP-2) initiative (Hall *et al.*, 2003). The WHS dataset has a resolution of 1^o and classifies 18 land use types. Vegetation parameters are then produced using lookup tables. The HIRLAM reference system physiography was originally based on WHS with some additions from national databases. The vegetation parameters computation in HIRLAM is still based on the WHS 18 land use categories. The ISLSCP-2 dataset based on satellite measurements has a much wider scope (see $http://islscp2.sesda.com$, for further details) in terms of surface variables and parameters. The available resolution is here 0.25^o, 0.5^o and 1^o.

Some efforts have been recently conducted to increase the resolution of vegetation datasets over Europe. Champeaux *et al.* (2000) produced a vegetation mapping over Western Europe using an automatic clustering of multi- temporal Normalized Difference Vegetation Index (NDVI) maximum values from the NOAA Advanced Very High Resolution Radiometer (AVHRR) satellite data. This method leads to 11 vegetation classes where the forest were readily identified from a thorough analysis of visible reflectances in early summer. This dataset is available at 2 km resolution. The usage of look-up tables allows for a final mapping of the primary parameters needed by the ISBA land surface scheme used operationally at Météo-France. Short-range forecasts produced by the ALADIN model has improved after the implementation of this new high resolution vegetation dataset.

Recently, a new global database of land surface parameters, called ECOCLIMAP (Masson *et al.*, 2003), has been put available to the atmospheric modelling community. This new database has some features which meet the requirements of the high resolution atmospheric models at their present level of development. First, it is a global dataset and consequently the problems of lack of homogeneity associated to the usage of different databases over big domains are circumvected. Second, it has 1-km resolution meeting the horizontal resolution needs of the current HIRLAM project. Third, it makes use of the very detailed information over the European region coming from the CORINE (Heymann *et al.*, 1993) and PELCOM (Mucher *et al.*, 2000) projects. Fourth, it makes use of full resolution maps of the vegetation index NDVI derived from the AVHRR visible and near-infrared spectral bands to provide the appropiate temporal and spatial scales. Fifth, it assigns 215 ecosystems, each of them represents a vegetation entity, which promotes a better assignment of vegetation parameter sets. Sixth, it includes aggregation rules to derive surface parameters at the desired model resolution. And seventh, the database allows the tiling approach, as used by the HIRLAM surface scheme.

The preliminary tests of the ECOCLIMAP database presented in this contribution address the very relevant problem of the big impact of the vegetation fields on heat fluxes and on screen level parameters. It was expected that a more realistic description of the vegetation features would be automatically translated on an improvement of the screen level parameters scores.

2 Derivation of surface parameters using the ECOCLIMAP database

The ECOCLIMAP database classifies the globe into 215 ecosystems with a 1-km resolution. In addition, soil texture is taken from the FAO data at 10-km resolution (FAO, 1988). The assignation of the vegetation and soil parameters to each of the considered ecosystems takes into consideration the existing climatic regions. The use of a climatologically based classification adds crucial information needed for the assignation of vegetation parameters. For instance, a deciduous forest can show completely different features in the Scandinavian and Mediterranean regions, as a consequence of different environmental factors affecting vegetation characteristics such as density of the canopy, sparseness of trees in the forest, yearly cycle, etc. In order to correctly allocate the surface parameters to each ecosystem, the ecosystem is splitted in turn into 12 primary vegetation types having different weight for each ecosystem. Aggregation rules are used both to derive surface parameters of a ecosystem compose of several primary vegetation types and to derive surface parameters for the usually coarser model resolution (see (Masson *et al.*, 2003), for further details). The code accompaning the ECOCLIMAP database supplies all vegetation and soil fields needed by the ISBA surface scheme. The computation is based on a comprehensive classifications of terrestrial ecosystems and on the extensive use of the aggregation rules proposed by Noilhan and Lacarrére (1995). Vegetation and soil parameters are obtainable either with monthly or decennial (10 days) frequency.

3 How to assess a physiographic database?

The question of the most appropiate method to evaluate a physiographic database is by no means a trivial one. The most straightforward way is to compare against point measurements or estimations of vegetation and soil features. This approach, however, has two serious drawbacks: first, the comparison is restricted only to certain landuses and climate conditions and second, the representativeness problem. A gridded database assigns one value to each gridpoint which is representative of the grid square around the gridpoint. Frequently, the observation of the vegetation features are not averaged over the surrounding area in order to account for the usually high heterogeneity of the vegetation and soil characteristics.

The second way of evaluating a physiographic database is to introduce it in a forecasting model and to compare the forecasted relevant parameters against the corresponding observations using the standard scores. The regularly observed parameters most affected by the vegetation features are the 2-metre temperature and relative humidity. This method has the obvious advantage of its globality. It allows the assessment of the physiographic database over the entire model domain, provided that there is a sufficient number of observations covering the whole integration area. The main drawback, however, is that models are usually tuned to their climatic fields and vegetation properties are part of those climatic fields. That means that a new physiographic database would be usually in clear disadvantage with respect to the previous physiographic description, which is expected to be well tuned to the rest of the model.

Another way to estimate the quality of a vegetation database is by comparing with other databases and by looking at the raw data (either using satellite information, terrain inspection or both) used to arrive to the vegetation maps. Also the comparison of the algorithms used to classify ecosystems and the aggregation rules, to upscale from the original database resolution to the resolution used by the forecast model, can shed some light on the quality of the database.

When a weather forecasting model is used to evaluate a vegetation database, it must be born in mind that models have usually compensation mechanisms to minimize errors usually coming either from observations or from other parts of the model. This is the case of the assimilation of soil water content based on the optimal interpolation method used by the HIRLAM model (Rodriguez *et al.*, 2003). The soil water content is corrected at every assimilation cycle to minimize errors of the 6 hours forecasted 2-metre temperature and relative humidity. Douville *et al.* (2000) explored, in the context of the FIFE data, the impact of the misspecification of the vegetation cover. They arrived to the conclusion that the assimilation of soil water content based on optimal interpolation, as proposed originally by Mahfouf (1991), was able to correctly simulate latent heat flux during a 1D 4 months simulation, when the vegetation cover was drastically reduced (from 87% to 8.7%). The soil water correction was able to compensate the low evaporative fraction induced by the low vegetation cover (bare soil tends to be less efficient in evaporating soil moisture than vegetated terrain). This was achieved by adding a lot of water into the soil. Their conclusion was that the soil water correction was a very robust method with respect to latent heat flux, although the method is not able to improve simultaneously turbulent fluxes and soil moisture if the vegetation parameters were poorly specified.

4 Experimentation with parallel runs

The preliminary evaluation presented here is based on the impact of the ECOCLIMAP database on the forecasts conducted with version 6.2 of the HIRLAM model. Both 'reference' and 'modified' experiments have the common following features:

- Domain: Area corresponding to the INM operational suite with non-rotated grid (65.0N, 30.0E, 15.5N, 66.5E) and 0.5° horizontal resolution.
- 194 * 100 grid points; 31 levels in the vertical.
- Each suite with its own data assimilation (3DVAR, 6 h cycling).
- Lateral boundary conditions: ECMWF analyses.
- 48 h forecasts from all analyses (00, 06, 12 and 18 UTC).
- Period: 1-15 July 1995.

Several parallel experiments, differing only in the number of ECOCLIMAP fields used to substitute the reference system assignations, were carried out:

- REF: Reference system (HIRLAM 6.2)
- ECO: The same as REF but veg, LAI, land fraction distribution, vegetation albedo, $R_{s_{min}}$ and root depth given by ECOCLIMAP
- EC1: The same as REF but land fraction distribution given by ECOCLIMAP
- EC2: The same as REF but veg, LAI given by ECOCLIMAP
- EC3: The same as REF but land fraction distribution, veg, LAI given by ECOCLIMAP
- EC4: The same as REF but land fraction distribution, veg, LAI, $R_{s_{min}}$ and root depth given by ECOCLIMAP

As we have mainly focused on the ECOCLIMAP fields which are important for the surface heat fluxes (veg, LAI, $R_{s_{min}}$, vegetation albedo and root depth), we discuss the impact of different alternatives on 2-metre temperature and 2-metre relative humidity which are directly affected by surface heat fluxes.

Among the various satellite derived vegetation indices, the most frequently used is the NDVI $= (\rho_1 - \rho_2)/(\rho_1 + \rho_2)$, where ρ_1 and ρ_2 are reflectance measurements in AVHRR channels 1 ($0.63 \mu m$) and 2 ($0.85 \mu m$). It has been shown (see *e.g.*, (Gutman and Ignatov, 1998)) that one NDVI measurement does not allow the simultaneous derivation of vegetation cover (veg) and leaf area index (LAI). Both parameters give some idea of the vegetation density: veg in the horizontal dimension and LAI in the vertical dimension. As low values of veg can be compensated by high values of LAI, we have preferred to test simultaneously the fields veg and LAI provided by the ECOCLIMAP database.

5 Impact of different physiography on screen variables

The bias and rms error of 6 h forecasts computed against EWGLAM stations for the 1-15 July 1995 period are shown in Fig. 1 for 2-metre temperature (top) and 2-metre relative humidity (bottom), respectively. The first noticeable feature is the big similarity of the scores produced by the different experiments. In fact after a few days of the assimilation cycle the scores both of T_{2m} and RH_{2m} tend to converge. The biggest difference appears for RH_{2m} during the first 4-5 days of assimilation. The observation verification scores for different regions tend also to be very similar (not shown here) with the exception of the African region ($30^0 N$, $30^0 E$, $20^0 N$, $10^0 W$) (see Fig. 2), where 2-metre temperatures are clearly higher for the ECO experiment, which makes use of veg, LAI, $R_{s_{min}}$, vegetation albedo and root depth provided by the ECOCLIMAP database. The vegetation albedo is the parameter responsible for the improvement in the ECO experiment. Fig. 3 shows the big difference between REF and ECO effective albedos, mainly due to the unrealistic predominance of the bare land fraction in the reference system. It is important to stress here that the reference system has a vegetation albedo which does not change with the yearly cycle in contrast with the seasonal dependance of the ECOCLIMAP albedo, which is obviously more realistic.

Some differences, although not so exaggerated, were also noticed over other regions and they will be object of a separate contribution. Also the impact of the substitution of each individual parameter (experiments EC1, EC2, EC3 and EC4) will be described in detail in the mentioned contribution.

6 Impact of different physiography on soil water content

If 2-metre temperature and relative humidity fields are not substantially affected by changes in the vegetation parameters, the next question arising from this behaviour is, which is the responsible mechanism of such reduced sensitivity? We have already commented in Sec. 3 that the soil moisture assimilation algorithm could compensate deficiencies in the estimation of vegetation parameters by adding/removing soil water. Figures 4 and 6 show the evolution of soil water content, 2-metre temperature and 2-metre relative humidity for two tipical grid points with coordinates $(46.5^0 N, 9.0^0 E)$ and and $(42.5^0 N, 2.0^0 W)$, respectively. The soil water content is represented in form of soil wetness index (SWI). SWI is defined by $(w - w_{wilt})/(w_{fc} - w_{wilt})$, where w_{fc}, w_{wilt} and w are the field capacity, the wilting point and actual soil water content, respectively. It takes the value 1 for soil water content at the field capacity, 0 for soil water content at the wilting point, negative values for soil water content below the wilting point and values bigger than one for soil water content between the field capacity and the saturation point. Both the 6 h forecasted values and the analysis increments (vertical lines) are represented for SWI, 2-metre temperatures and 2-metre relative humidity . It is clearly seen that the forecasted 2-metre temperatures are roughly the same for experiments REF and ECO, having a different set of vegetation parameters. Tables 1 and 2 give the values of the vegetation parameters used at the considered grid points. The vegetation parameters corresponding to the forest and the low vegetation tiles are presented for both REF and ECO experiments. The forecasted 2-metre relative humidity also shows a big similarily between REF and ECO experiments and for both forest and low vegetation tiles. The biggest difference appears, however, between days 8 and 11 for both points here considered (see figures 5 and 7, for a zooming of this period). From the SWI evolution (figures 4 and 6 (top)), it is clearly seen that the soil moisture assimilation algorithm is able to minimize differences in 2-metre temperatures and 2-metre relative humidity at the cost of a substantial modification of the total soil moisture, sometimes reaching very unrealistic values. In fact, the biggest change in the total soil moisture is conducted during the assimilation part (vertical steps in the plot).

It seems, therefore, that the robustness of the soil moisture assimilation method is able to cope successfully with variations in the vegetation features. It must be reminded that the soil moisture assimilation algorithm is based on the minimization of 6 hours forecasts errors of 2-metre temperature and relative humidity by adjusting soil water content. There are, however, certain limitations. The soil moisture assimilation algorithm only corrects the first guess value when the actual soil water content falls between field capacity and the product of vegetation cover times the wilting point. Besides, the assimilation algorithm only corrects when certain constraints linked to the synoptic conditions are met (Rodriguez *et al.*, 2003). In particular, when errors of 2-metre temperature and 2-metre relative humidity have different signs. These conditions assure that 2-metre temperature and humidity are strongly coupled to soil moisture.

The zooming appearing in figures 5 and 7 shows, first, that different evolution (mainly of 2-metre relative humidity) is clearly seen for both tiles of the same experiment and, second, that big errors of the screen variables have only effect on soil moisture analysis increments when synoptics constraints are met.

Figure 8 shows a tipical case of evolution of soil water content, 2-metre temperatures and 2-metre relative humidity corresponding to the African verification region. The only tile represented here is the corresponding to bare soil, which is the predominant one over the considered grid square. The soil moisture assimilation algorithm is the main responsible for the drying out of the total soil water reservoir in the case of the REF experiment. The excessively high albedo (see fig.3 and table 3) in the REF experiment gives too cold forecasted (H+6) 2-metre temperatures which in turn give place to a removal of soil water by the soil moisture assimilation algorithm. As water removal hardly has influence on heat fluxes, due to the unefficiency of the evaporation process in cases of very dry soils, the error of the 6 hour forecasted 2-metre temperatures is steadily maintained causing an unrealistic dessication of the soil. The REF experiment shows that 6h forecasted 2-metre temperature errors are rather big, mainly at 18 UTZ. These too cold temperatures, caused by the erroneous albedo, give place to an excessive removal of soil water.

7 Conclusions

The vegetation parameters supplied by the ECOCLIMAP database have been added to the HIRLAM climatic files and used to substitute values assigned by the correspondance tables included in the HIRLAM surface code. So far, the following ECOCLIMAP fields have been explored: leaf area index, vegetation cover, land tile distribution, root depth, vegetation albedo and minimum stomatal resistance.

The effect of the vegetation parameters change has been offset by the soil moisture assimilation algorithm, which is able to compensate differences in vegetation parameters by changes in soil water content. Of course, if vegetation parameters are wrongly specified, it cannot be expected that soil moisture values are realistic. A complementary conclusion of this work is that the optimal interpolation method for soil moisture assimilation is a rather robust approach, which preserves surface heat fluxes rather well against changes in vegetation parameters by minimizing errors of forecasted 2-metre temperature and humidity. The possible errors or misspecification of vegetation parameters is therefore translated to the soil water content, which is an output not very much used of the weather prediction models.

Some further experiments without soil moisture assimilation will be conducted as a continuation of this work to confirm the compensatory role of the soil moisture assimilation algorithm against changes in vegetation parameters. Also other periods will be explored, in particular in spring, when the soil water reservoir is close to the saturation, and in autumn, when soil water recharge starts. Finally the interaction between tiles in the same grid box -through the lowest model layer- will be also further studied.

Acknowledgments. We thank Valéry Masson of Météo-France for putting available the ECOCLIMAP database to the atmospheric modelling community. We also thank Beatriz Navascues and Jose Antonio Parodi for the very fruitful discussions on the role played by soil moisture assimilation. Thanks go also to Han The and Oscar Garcia for his help with the climate files generation.

References

1. Avissar, R. and Pielke, R.A. 1989. A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Mon. Wea. Rev.*, **117**, 2113-2136.

2. Champeaux, J.L., Arcos, D., Bazile, E., Giard, D., Goutorbe, J.P., Habets, F., Noilhan, J. and Roujean, J.L. 2000. AVHRR-derived vegetation mapping over western Europe for use in numerical weather prediction models. *Int. J. Remote Sens.*, **21**, 1183-1199.

3. Douville, H., Viterbo, P., Mahfouf, J.-F. 2000. Evaluation of the Optimum Interpolation and Nudging Techniques for the Soil Moisture Analysis Using FIFE Data. *Mon. Wea. Rev.*, **128**, 1733-1756.

4. FAO, 1988. UNESCO soil map of the world. World Soil Resources Rep. 60, Rome, Italy.

5. Giard, D. and Bazile, E. 2000. Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997-1015.

6. Gutman, G. and Ignatov, A. 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *Int. J. Remote Sens.*, **19**, 1533-1543.

7. Hall, F.G., Meeson, B., Los, S. Steyaert, L., Brown de Colstoun, E., Landis, D., eds. 2003. ISLSCP Initiative II. NASA. DVD/CD-ROM NASA.

8. Heymann, Y. and Coauthors. 1993. CORINE land cover: Technical guide. Environment, nuclear safety and civil protection series, Commision of the European Communities, Office for Official Publication of the European Communities, Luxembourg, 144 pp.

9. Loveland, T.R. and Belward, A.S. 1997. The IGBP-DIS global 1 km land cover dataset, DISCover: First results. *Int. J. Remote Sens.*, **18**, 3289-3295.

10. Mahfouf, J.-F. 1991. Analysis of soil moisture from near-surface parameters: a feasibility study. *J. Appl. Meteorol.*,**30**, 1534-1547.

11. Masson, V., Champeaux, J.L., Chauvin, F., Meriguet, C. and Lacaze, R. 2003. A Global Database of Land Surface Parameters at 1-km Resolution in Meteorological and Climate Models. *J. Climate*, **16**, 1261-1282.

12. Mucher, C.A. and Coauthors. 2000. Development of a consistent methodology to derive land cover information on a European scale from remote sensing for environmental modelling. PELCOM Report (ENV4-CT96-0315). European Commision. Directorate General XII-D. 299 pp.

13. Noilhan, J. and Lacarrere, P. 1995. GCM grid-scale evaporation from mesoscale modeling. *J. Climate*, **8**, 206-223.

14. Richardson, L.F. 1922. *Weather Prediction by Numerical Process*. Cambridge Univ. Press, reprinted Dover, 1965, 236 pp.

15. Rodriguez-Camino, E. and Avissar, R. 1998. Comparison of three land-surface schemes with the Fourier Amplitude Sensitivity Test (FAST). *Tellus*, **50A**, 313-332.

16. Rodríguez, E., Navascués, B., Ayuso, J.-J. and Järvenoja, S. 2003. Analysis of surface variables and parameterization of surface processes in HIRLAM. Part I: Approach and verification by parallel runs. *HIRLAM Technical Report No 58*, Norrköping, Sweden. [Available from P. Unden, SMHI, S-60176 Norrköping, Sweden].

17. Wilson, M.F. and Henderson-Sellers, A. 1985. Cover and soil datasets for use in general circulation models. *J. Climatol.*, **5**, 119-143.

Table 1: Vegetation and soil parameters used by REF and ECO experiments at the grid point $(46.5^0 N, 9.0^0 E)$. Values are given for both forest and low vegetation tiles.

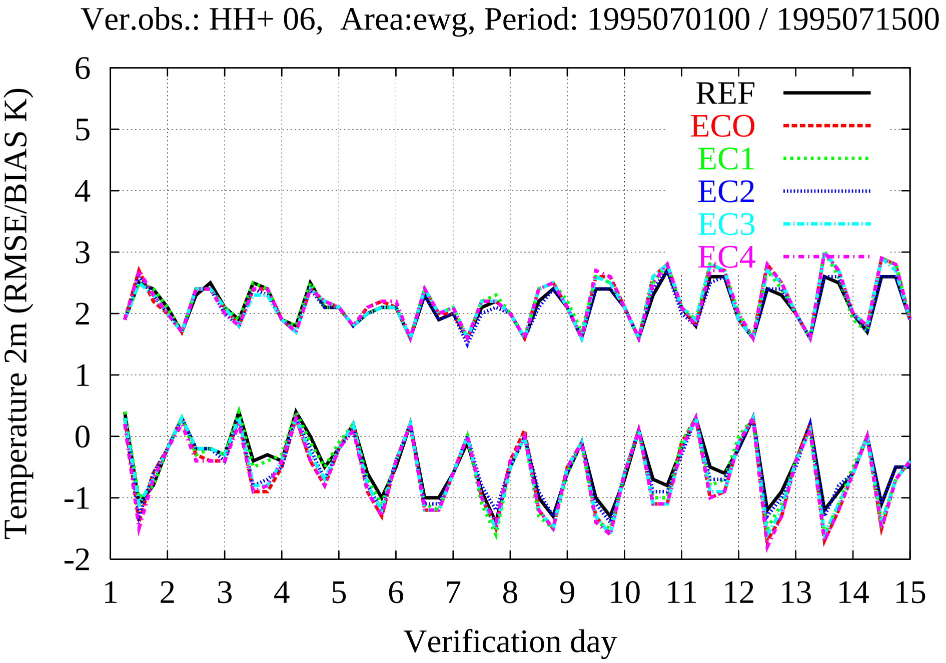
Parameter	REF (low veg.)	ECO (low veg.)	REF (forest)	ECO (forest)
veg	0.8	0.95	0.90	0.95
LAI	2.0	2.96	5.0	3.5
albedo	0.18	0.21	0.16	0.10
Rsmin	40.	40.	250.	150.
root depth	2.0	1.5	3.0	2.0

Table 2: The same as table 1 but for the grid point $(42.5^0 N, 2.0^0 E)$.

Parameter	REF (low veg.)	ECO (low veg.)	REF (forest)	ECO (forest)
veg	0.85	0.83	0.90	0.95
LAI	1.0	2.34	5.0	3.34
albedo	0.20	0.18	0.16	0.13
Rsmin	40	40.	250.	150.
root depth	1.5	1.34	3.0	2.0

Table 3: The same as table 1 but for the grid point $(30.0^0 N, 10.0^0 E)$ and for the bare soil tile.

Parameter	REF (bare soil)	ECO (bare soil)
albedo	0.48	0.27



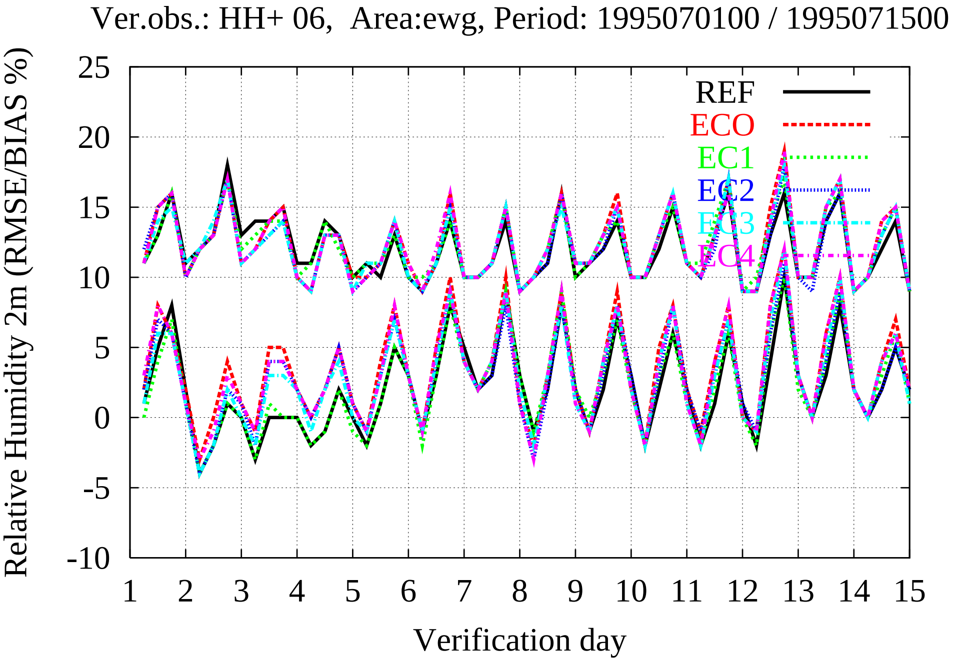
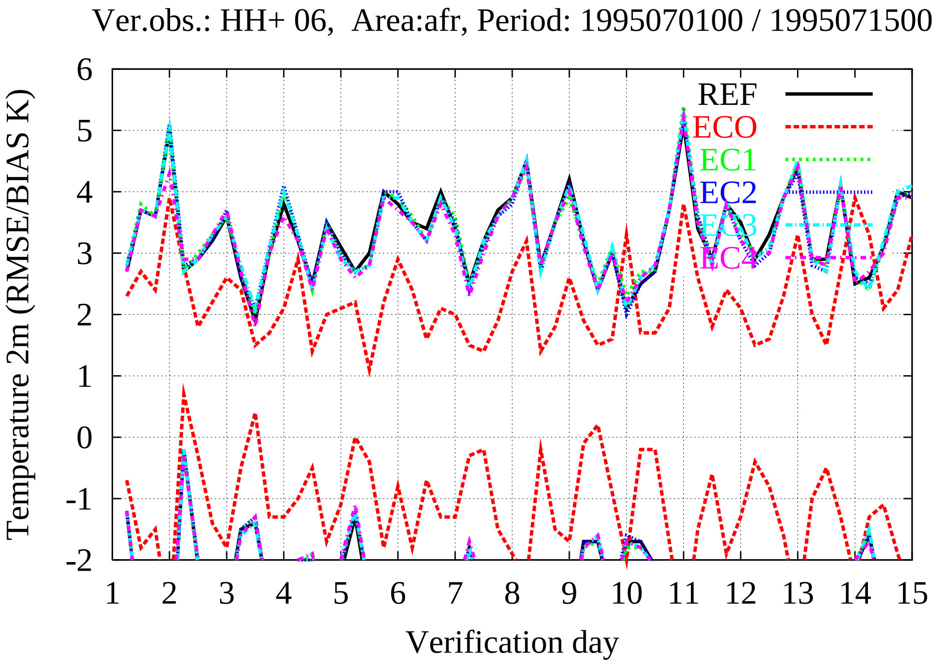


Figure 1: 2-metre temperature (top) and 2-metre relative humidity (bottom) bias/rms error of H+06 forecastings against observations at the EWGLAM stations. Experiments REF, ECO, EC1, EC2, EC3 and EC4 are described in the text. Verification period: 1-15 July 1995



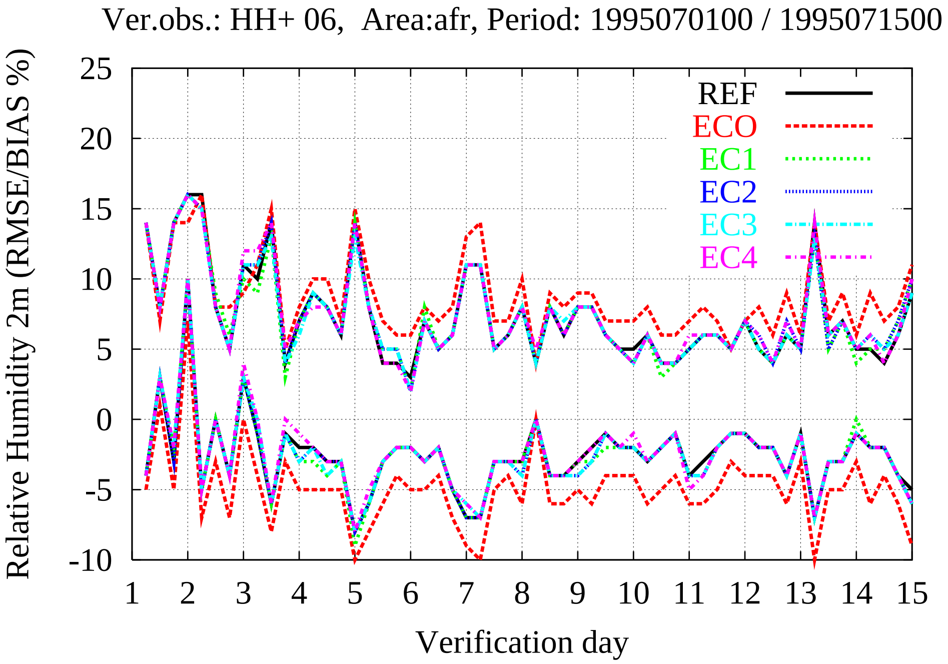


Figure 2: The same as Fig. 1, but for the African region ($30^0 N$, $30^0 E$, $20^0 N$, $10^0 W$)

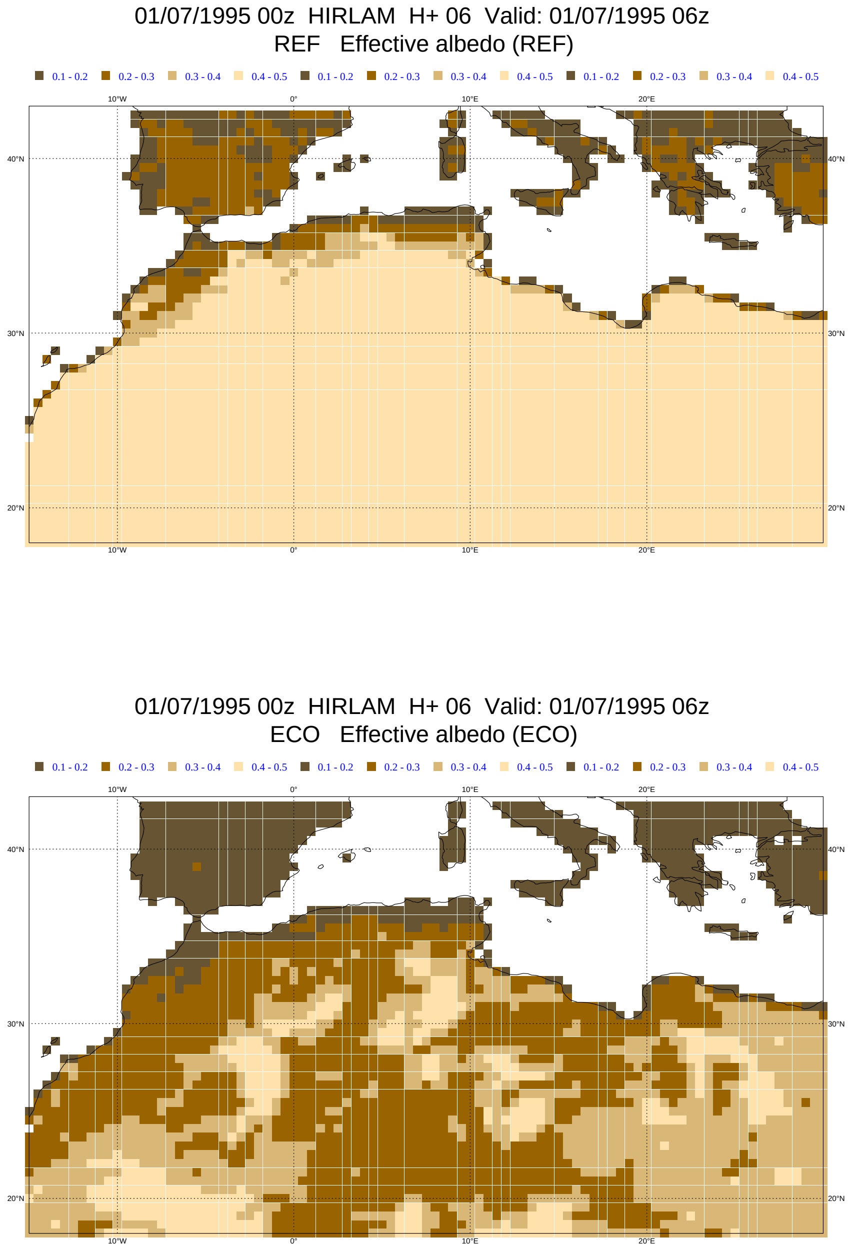


Figure 3: Effective albedo for the REF (top) and ECO (bottom) experiments over the African region $(30^0N, 30^0E, 20^0N, 10^0W)$

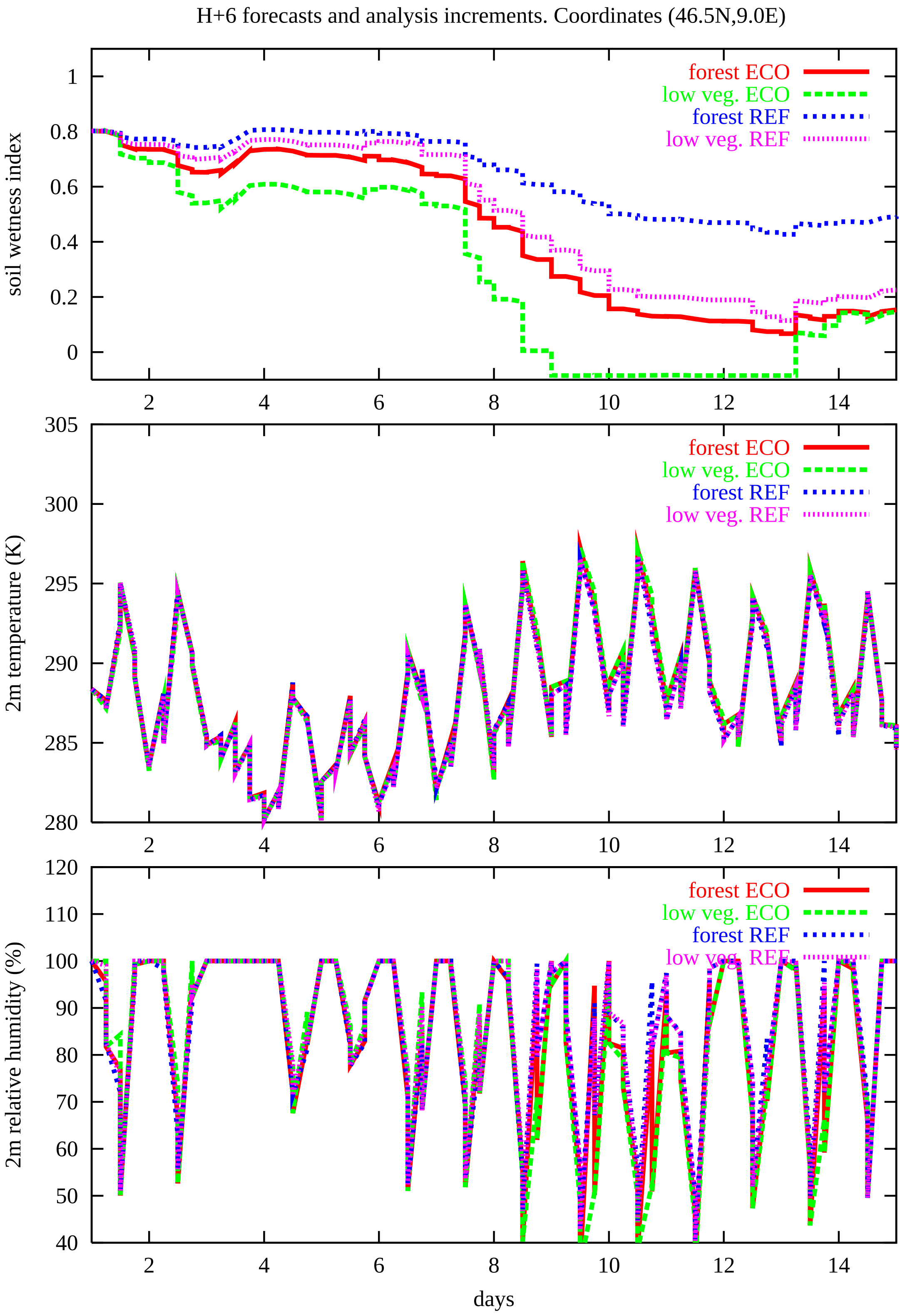


Figure 4: Evolution of total soil wetness index (top), 2-metre temperature (middle) and 2-metre relative humidity (bottom) for the grid point $(46.5^0 N, 9.0^0 E)$. The values corresponding to the forested and low vegetation tiles are plotted for the ECO and REF experiments. Both first guess (H+6) values and analysis increments (vertical lines) are represented. Evolution period: 1-15 July 1995

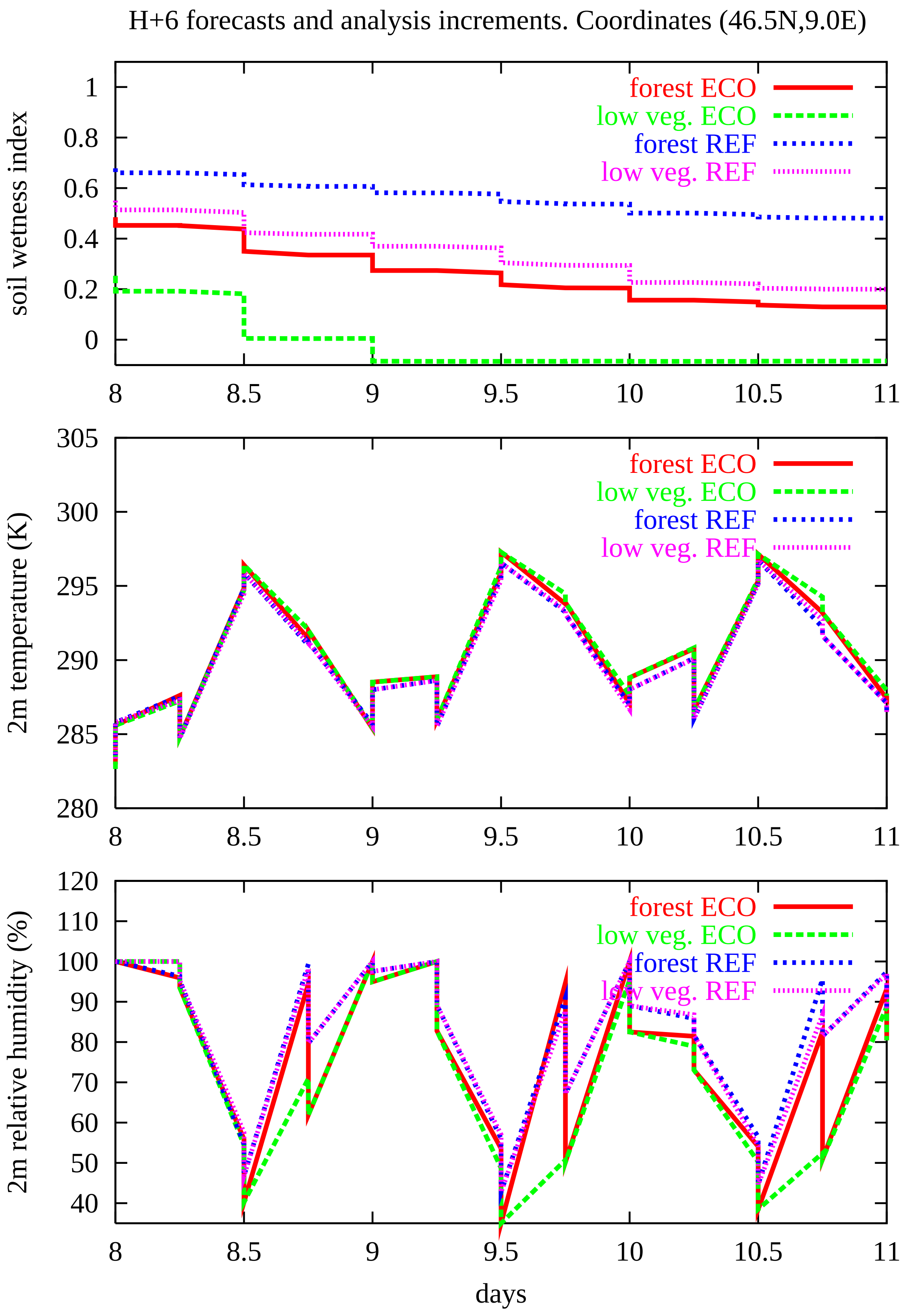


Figure 5: The same as Fig. 4 but zooming between days 8 and 11 July 1995.

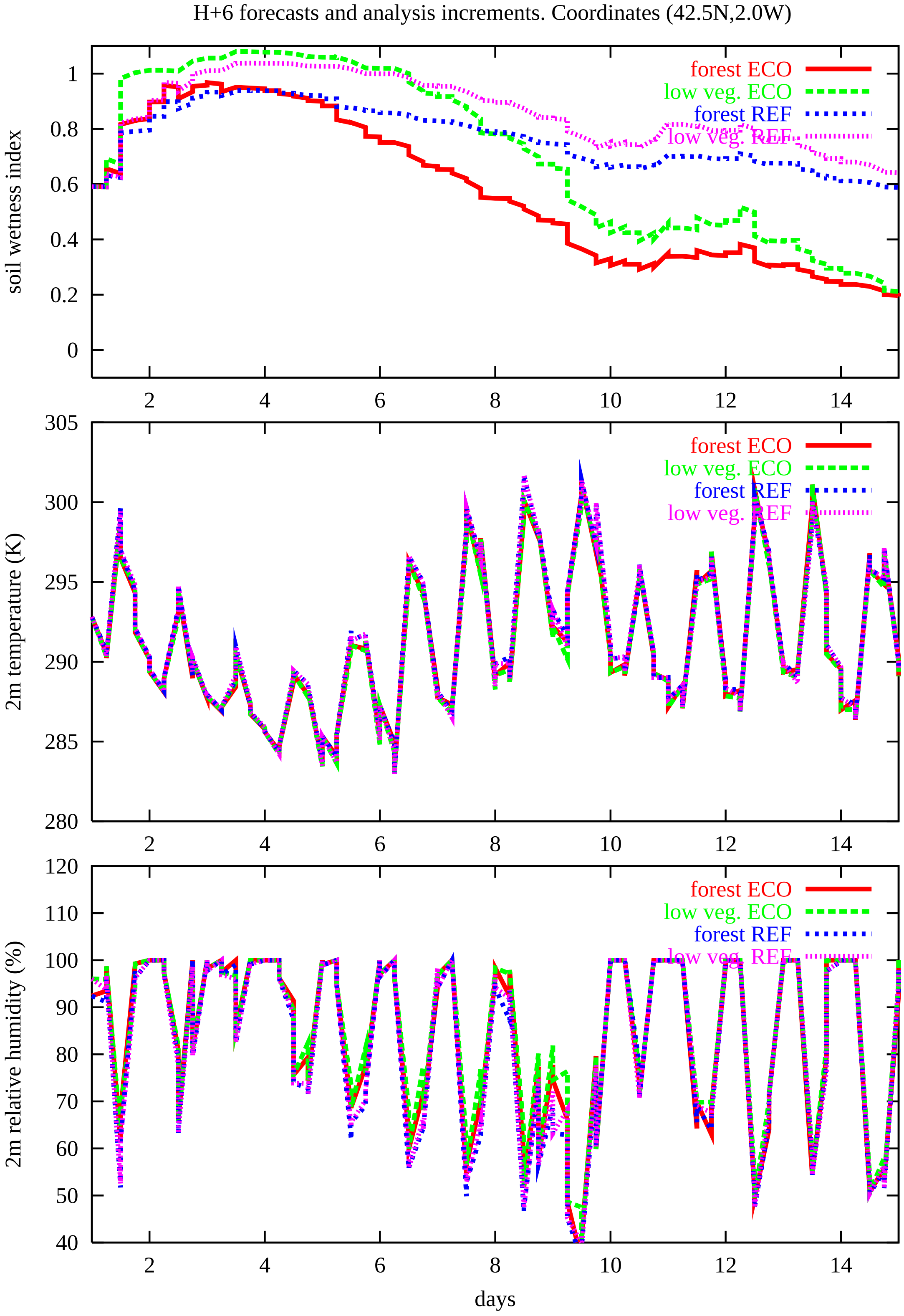


Figure 6: The same as Fig. 4 but for the point $(42.5^0N, 2.0^0W)$

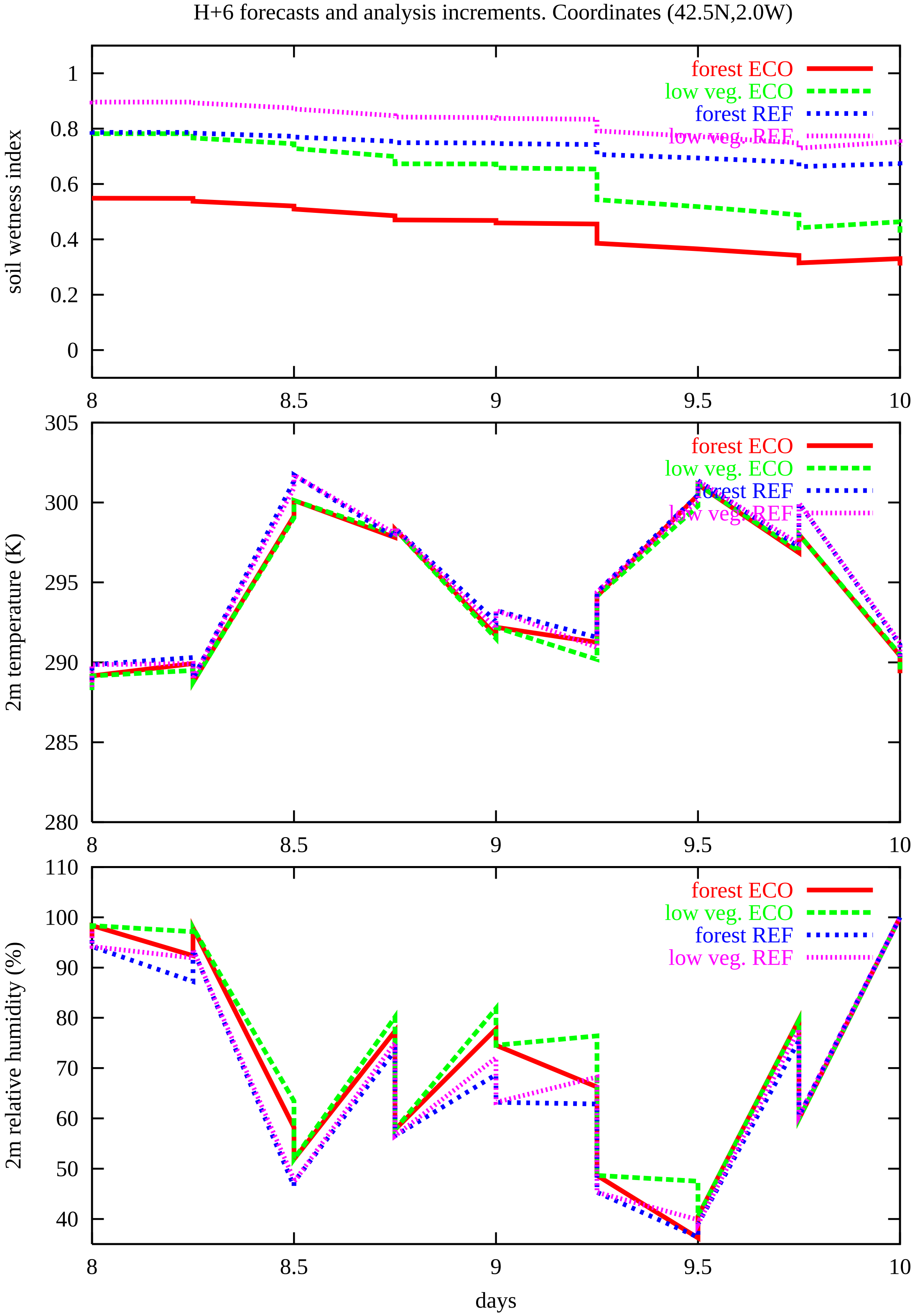


Figure 7: The same as Fig. 6, but zooming between days 8 and 10 July 1995.

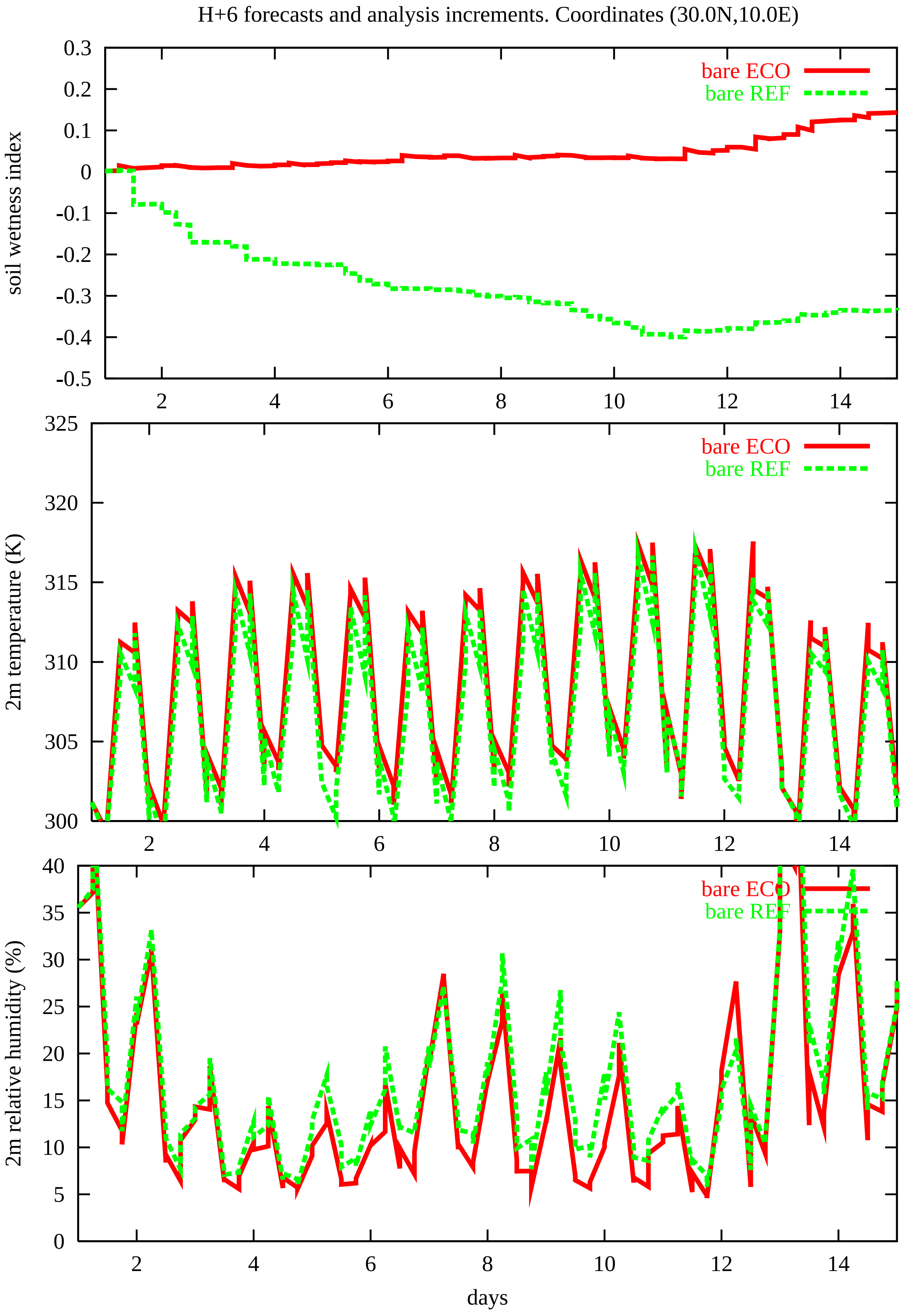


Figure 8: The same as Fig. 4 but for the point $(30.0^0 N, 10.0^0 E)$. The values correspond here only to the bare ground fraccion, which is the predominant one over the considered grid square.