Soil initialization strategy for the COSMO model

Francesca Di Giuseppe, Giovanni Bonafé, Davide Cesari

ARPA-Servizio IdroMeteoClima, Bologna, Italy

1 Introduction

Soil initialization is not a straightforward task for a number of reasons. Firstly there are almost no suitable measurements for operational analysis. In situ observations are rare and heterogeneous. The Global Soil Moisture Data Bank project [5] is an example of collection, dissemination, and analysis of gravimetric observations of soil moisture and temperature data from around 600 sites over the globe, but until now has served mainly the scope of creating a dataset for verification and initialization of climate simulation, without attempting to satisfy the real-time needs. Remotely sensed surface soil moisture data are available with mostly daily frequancy, but the detected radiation is directly linked only to the model's uppermost soil layer and therefore can only provide partial information. Moreover soil moisture retrieval from microwave frequencies requires accurate specification of the vegetation cover and soil type at the pixel location which, is not usually possible with the desidered precision. In addition to the lack of useful observations, it has to be mentioned the small representativity of soil measurements, where they exist. Correlation lenght for soil moisture, for example, can be as small as 10 m [8], making the design of an observational network almost an utopia.

The need for soil initialization without representative measurements has led the interest in indirect determination of soil prognostic variables. In this work, three different choices to initialise limited area model soil fields are compared, with the aim of identifying the most suitable strategy which combines ease of implementation, improvement in forecast skills and realistic estimations of soil paremeters, expecially in the light of hydrological applications. The regional model COSMO is used as limited area model forecasting system. The ECMWF, IFS model as global model.

For a three month long integration period spanning September-October-November (SON) 2008, the COSMO soil scheme TERRA is initialised either by a simple interpolation from the IFS soil moisture analysis or by fields generated by a COSMO previous integration or by a local soil analysis implemented with a variational scheme which uses screen level temperature to adjust the soil hydric content to minimises the distance between background and observations.

Extensive comparisons with observations show that most of the improvements are achieved when passing from a simple interpolation from ECMWF experiment to a free running COSMO model, while the soil moisture correction adds a marginal benefit to the prediction of surface fluxes. This suggests that keeping the model soil fields in good equilibrium with the soil scheme, especially for what concerns temperature, is more relevant than the subsequent correction of soil moisture provided by an analysis scheme.

2 Models, data and experiment design

In this section the set-up for the experiments is desccribed in details as well as the validation dataset used for the comparison. Since soil moisture analysis based on atmospheric observations is indirect and relies heavily on the underlying model, a brief summary of the main features of the land surface schemes in the operational ECMWF model (IFS) (used in the experiment named ECMWF) and in the COSMO model is given.

2.1 Experiment strategy

To isolate the effect of soil prognostic variables (i.e. temperature and humidity) on the model analysis, only the methods for initialzing those fields are varying among the three different experiments, while all other model configurations, external fields and boundary fields are kept invariant. The bottom layer soil temperature and humidity and the sea-surface temperature are interpolated from the global model IFS. In all experiments the atmospheric forcing is prescibed from the previous COSMO analysis as in the operational implementation. A three month long integration is performed starting the 1st of September 2008 (SON period). Each cycle lasts 24h starting at 00 UTC, 3-hourly boundary conditions are provided by IFS run at roughly 25 km resolution.

Three soil moisture intialization methods are implemented.

- *Initialization by interpolation from the ECMWF soil analysis (hereandafter ECMWF experiment).* The soil moisture and temperature fields are initialised using IFS analysis. At the beginning of each assimilation cycle the soil temperature and moisture fields are re-initialised with interpolated fields from ECMWF. The interpolation is firstly performed horizontally by taking the nearest IFS land point, and then vertically by a linear fit.
- *Free running soil initialization (hereandafter COSMO experiment).* The initialization is performed using the soil moisture and temperature fields from the previous COSMO run. After the first few days of start up the soil moisture of this experiment represents the equilibrium between the source terms (precipitation, dew, rime, snow) and the sink terms (evaporation from bare soil, transpiration from plant, run-off). The temperature profile is instead calculated with the force-restore method using the energy budget at the surface and it's therefore a direct consequence of the atmospheric model radiative forcing.
- *Free running soil initialization plus variational soil moisture analysis using surface 2 m synop observations (SMA experiment)* A local soil moisture analysis is performed using the variational scheme from [10]. The soil moisture is adjusted to minimize the distance between background and synop observations. The soil temperature is initialized from the previous COSMO run but, being the soil heat capacity a function of the soil moisture, it will predict a soil temperature profile which is different from the one of the COSMO experiment. Moreover a different soil moisture produces a different radiative coupling with the atmosphere. (sono giuste le correzioni qui?)

In table 1 a brief summary of the three experiment is reported

2.2 The COSMO soil moisture analysis

The soil moisture analysis used in the SMA experiment is based on a variational approach outlined in [10]. Firstly, a T_{2m}^{an} analysis field is obtained by optimal interpolation of synop observations and model background coming from the previous, T_{2m}^{fg}, COSMO run. Then, increments, ΔT_{2m}, at 12 and 15 UTC are calculated as $T_{2m}^{an} - T_{2m}^{fg}$. Finally, ΔT_{2m} are converted in moisture increments Δwg_i at the various soil levels using a parameterized form of

Exp Name	Temperature	Soil Moisture	Comments
ECMWF	Interpolated from IFS	Interpolated from IFS	
COSMO	From the previous COSMO run	Interpolated from COSMO	Bottom layer from IFS. First initialization from IFS
SMA	From the previous COSMO run	Adjusted with a local soil moisture analysis based on T_{2m} from the synop network	Bottom layer from IFS. First initialization from IFS

Table 1: Summary of the main characteristic of the three experiments under analysis

the jacobian $\frac{\partial w_g}{\partial T_{2m}}$. ΔT_{2m} are evaluated at two instants only when the soil-atmosphere coupling is supposed maximum. One should therefore expect that the major benefit from the soil moisture analysis arises primarily for situations the scheme is designed for, i.e. correcting forecast errors at the time the observations are assimilated, which is around noon. If the bias does not change its sign, the scheme should nevertheless be able to improve screen level errors caused by misspecification of bowen ratio at other lead times. On the other hand, in cases in which the soil is characterised by a substantial heating during daytime and cooling during night, as it happens for example in very dry soil conditions, then the application of this kind of soil moisture correction can exacerbate the reduction in soil thermal inertia with a worsening of the biases during nightime.

All these aspects will be investigated in the following sections.

2.3 The IFS soil analysis

The IFS soil scheme has four prognostic soil layers for moisture and temperature, with a free drainage and a zero heat flux condition at the bottom of the deepest layer. It also includes a precipitation interception layer and a skin temperature level. From the surface to the bottom, the layer thicknesses are, respectively, 0.07, 0.21, 0.72, and 1.89 m. The three top layers correspond to the root zone with a total depth of 1 m. The root density decreases exponentially with depth. The surface evaporation has a bare soil part controlled by soil moisture in the top layer and a vegetation part. The role of the vegetation is represented explicitly, through a transpiration term and an interception loss term corresponding to the evaporation of dew and intercepted rain at the potential rate. The transpiration is controlled by the leaf area index (LAI) and the stomatal conductance, which is regulated by the water availability in the root zone (top three layers) and the photosynthetically active solar radiation.

The IFS implements a soil moisture analysis which employs an Optimal Interpolation (OI) method proposed by [4]. Similarly to the SMA experiments described above, it is based on the analysis increments of 2 m temperature and relative humidity. Every 6 h, corrections applied in each soil layer (analysis increments) are linear combinations of atmospheric increments of 2 m temperature and relative humidity. The details of the method can be found in [1], while for full details on the quality of the IFS soil scheme and comparison with observations we refer to [7].

2.4 The verification dataset

The verification dataset is composed of three sources; surface flux measurements collected by the EU-funded research project CARBOEUROPE Integrated Project (CEIP), soil humidity data collected at the ARPA-SIMC meteo station located at San Pietro Capofiume (SPC) in the middle of the Italian Po Valley and the standard network of synop surface stations (400) which cover the COSMO-I7 domain.

The CARBOEUROPE project has the main aim of quantifying the relationship between carbon fluxes and vegetation characteristics. Therefore, great attention has been posed to locate observing stations over different land use/cover types. Measurements [1] are recorded since 2004 half-hourly on more than one hundred Eddy flux stations over Europe. The collected dataset therefore potentially possesses a good representativity of fluxes over different ecosystem types. The location and the vegetation characteristics of the stations which fall into the COSMO-I7 domain and were active during the validation period (SON 2008) are reported in table 1.

SPC meteo station is an intensive observation meteo station managed by ARPA-SIMC. In addition to the conventional meteorological measurements including SYNOP and TEMP variables, since 2007, is operating a Time-Domain Reflectometer (TDR) which measures soil water content and temperature profiles at 8 unevenly spaced levels below the ground between 10 and 100 cm. At the time of the experiments, SPC was not provided with instrumentation for surface fluxes measurements. Finally, global diagnostic are calculated using the synop

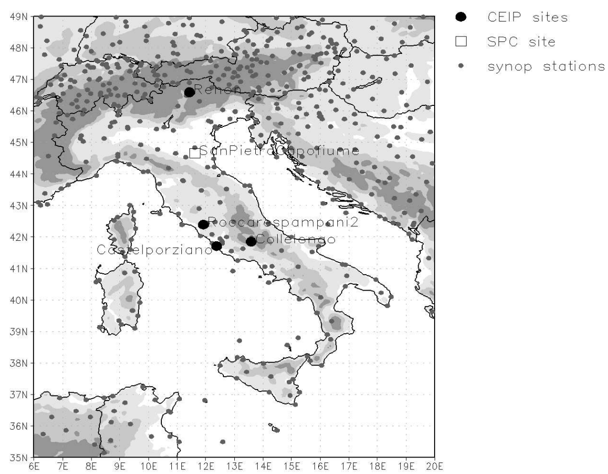

Figure 1: Type and location of the observational dataset used in this study. The displayed area is the operational domain of the COSMO-I7 suite.

network which comprises more then 400 surface stations over land. A map of the location of the stations used for the comparisons is reported in figure 1.

3 Problem diagnosis and impact of soil initialization

The main motivation for investigating a different soil moisture initialization strategy from the simple ECMWF interpolation are systematic annual biases in screen-level temperatures and humidities. As soil variables were initially interpolated from the IFS model it was often

[1] data available at www.carboeurope.org

observed that in the Po Valley, in early spring, temperatures were systematically too high while in summer they were too low. Figure 2 shows the diagnosis of the various problems to be faced. The seasonal variation of the outgoing longwave radiation clearly shows a shift in the annual wave phase, which is superinposed to the delay in the daily cycle. The soil model appears to be too conductinve and thus unable to buid-up thermal enargy during the summer. Moreover the weak daily cycle and its offset of few hours is responsible for a warm nighttime bias and a cold daytime bias. The first prevents the establishment of stable stratified PBL conditions typical of fog formation, the latter inhibits strong daytime mixing with dalay in the triggering of local convection.

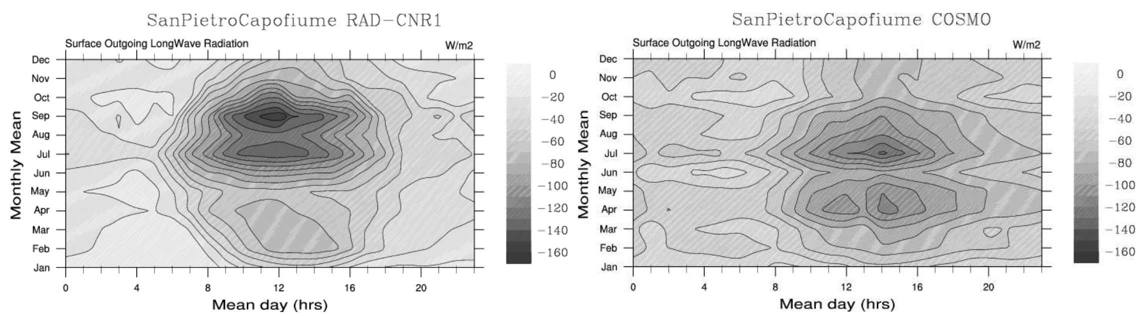

Figure 2: Seasonal versus daily outgoing longwave radiation at SanPietroCapofiume location as measured by the CNR-1 (upper panel) radiometer and as predicted by the COSMO-I7 model (lower panel). The data are averaged over two years 2007-2008.

The September-October-November 2008 period chosen for verification was a typical autumn season with an intermittent series of heavy precipitations and dry spelt days. Figure 3 shows the daily averaged observed precipitation and T_{2m} at the SanPietroCapofiume location. The first 10 days of October experienced no precipitation and have been marked as 'Dry Period'. Between the 24th of October to the 2nd of November several rain bands were moving estlerly from a low pressure minimum located in the middle of the mediterranean sea. Ten days of almost continous precipitation were recorded and have been marked as 'Wet period'. Figures 4 and 5 show the 3-hourly 2m temperature and relative humidity for the 10 days during the 'Dry Period' and the 'Wet period'.

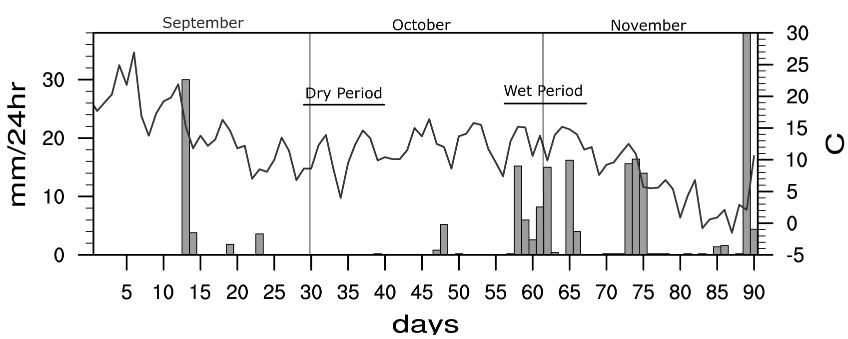

Figure 3: Observed daily precipitation at SPC during the SON period 2008

As expected the largest model discrepancy with the observations is at nightime during the dry period. These biases are exacerbated in the relative humidity due to the fact that the model is evidently also slighly dry. In dry condition, the best performance during daytime is achieved by the SMA initialization as one would expected. Under strong radiative forcing screen-level temperature, used as predictors, strongly depend on the soil moisture content. Moreover soil moisture incremens are only calculated during daytime and it is when the scheme works at its best. If, as it happens, the daytime bias has opposit sign with respect to the nighttime bias, the SMA approach will tend to dry-up a soil which is already too dry exacerbating the nighttime bias as shown, for example, at day 38 in figure 4 where correcting the max 2m temperature brings a worsening of the minimum night temperature. The limitation depited is inherent in the methodology which applays increments calculated at 'noon' to the whole assimilation window.

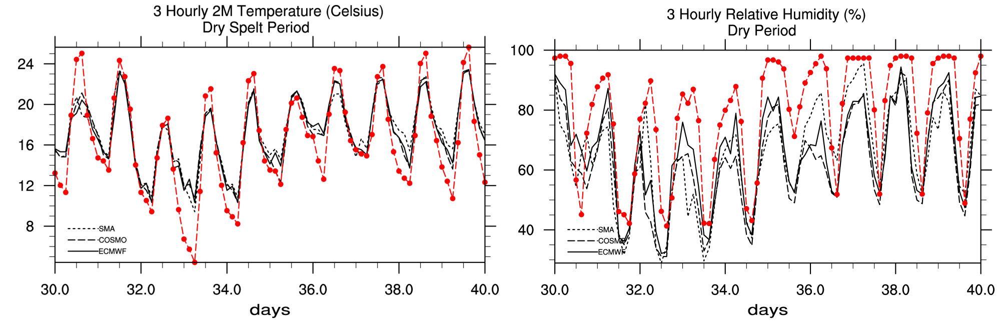

Figure 4: Time series of 3-hourly 2m temperature and relative humidity for the 10 days during the 'Dry Period'.Observations are from the rain gauges while the model simulations are from the three different intializations.

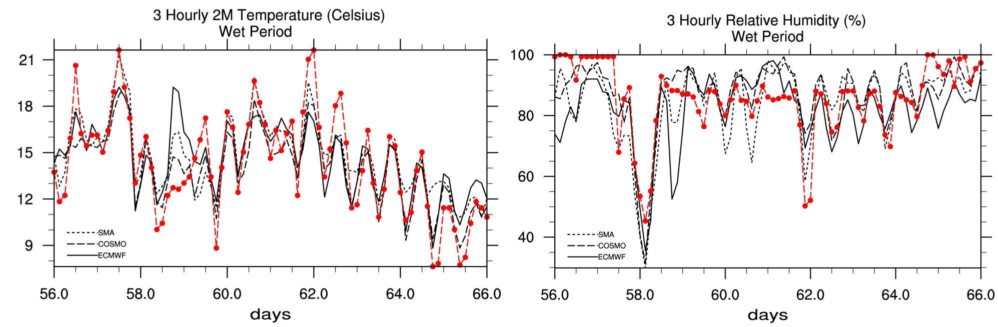

Figure 5: Same as figure 4 but for the wet period

The 'wet period' is characterised by smaller biases, with the exception of day 59 when all the three model simulations missed completly the observed precipitation. It is worth nothing that in this wet regime when the soil moisture is close to its field capacity value the SMA experiment produces unrealistic warm increments. This is a side effect produced by the missing strong coupling between the soil and the boundary atmospheic level. Indeed care has to be taken regarding the applicability of a soil moisture scheme when the information content at screen level is weak. Some thinning is therefore recommended to exclude synoptic situations with weak coupling between soil and atmosphere and when horizontal coupling between neighbour grid points is weak (cloud free, not too strong advection).

Biases of up to -4 K in the T_{2M} and 20 % in the RH_{2M} highlight problems in the estimations of both turbulent surface fluxes which can be due both to ground temperature and humidity estimations and/or wrong surface turbulent coefficients. Figure 6 shows the mean-day sensible and latent heat fluxes at one of the CARBOEUROPE station taken as example averaged over 10 days during the dry and wet periods identified. Since the other sites show very similar results they have not been included. On average both sensible and latent heat fluxes are always underestiamted (see also figure 8), nevertheless, during nightime there is a substantial over-estimation of both turbulet fluxes. The nightime bias is expecially marked in the ECMWF experiment as was diagnosted by the excess in the outgoing longwave emission of figure 2. The soil–atmosphere interface is too warm and prevents the formation of nightime stratified stable boundary layer conditions.

The diurnal cycle, almost completly absent in the ECMWF experiment, is highly improved by both the COSMO and SMA experiments, showing the role played in turn by the ground temperature and humidity initialization. Moreover this highlights the main danger of external interpolation from global models as a strategy for surface initialization. TERRA in its present implementation is very conductive, so that, deep levels heating is conveyed to upper layers very efficiently, while IFS SVAT scheme is tuned to a much lower conductivity. When the interpolation is performed the initial ground temperature profile given to TERRA is therefore not optimal for the new scheme which then predict a completly unrealistic cycle of the diurnal cycle. As an example, figure 7 shows the daily variation of the ground temperature at two levels as predicted by the three assimilation cycles at Collelongo during the dry period. Despite the fact that SMA and COSMO have a more pronounced diurnal cycle the most stricking feature is the very high temperature predicted by the ECMWF experiment in the deeper ground levels. The use of this value into the TERRA scheme probably causes the unrealistic noctournal heating diagnosted in figure 8. Most of the improvements are achieved when passing from the ECMWF experiment to the COSMO one while the SMA initialization only adds a marginal benefits to the prediction of surface fluxes. This also suggests that a correction of soil temperature is more relevant then the subsequent correction of soil moisture at least in dry condition. In other words, when the soil is close to its wilting point it is hard to estimate the dominating effect between thermal inertia and radiative cooling and possibly simply running the forecast model can furnish the zero-order correction to the soil initialization problem. Near surface variable biases are strongly dependent on the soil conditions. Therefore schemes which for their design only correct for errrors on a selective basis can therefore only be able to 'conditionally' correct them.

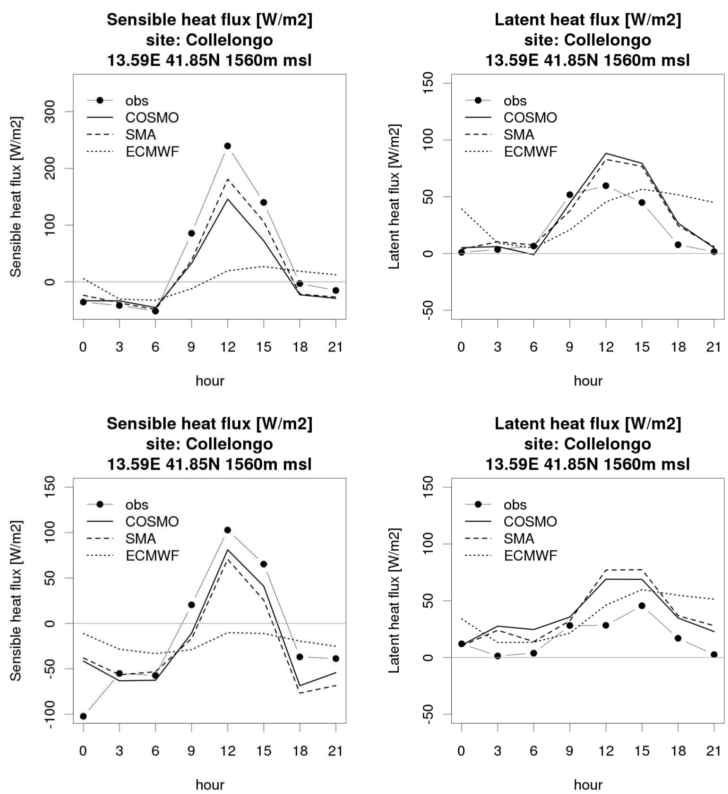

Figure 6: Top panel.Comparison of model and observed meanday sensible and latente heat fluxes for the dry period under study. Low panel. as on the top panel but for the wet period. Similar results have been found in the other CARBOEUROPE sites (not shown)

Keeping these caveats in mind it is nevertheless important to assess the global statistical performance of the various experiments in terms of forecast scores. Figure 8 shows the comparison between H and L_E at two CARBOEUROPE locations for the three methods as

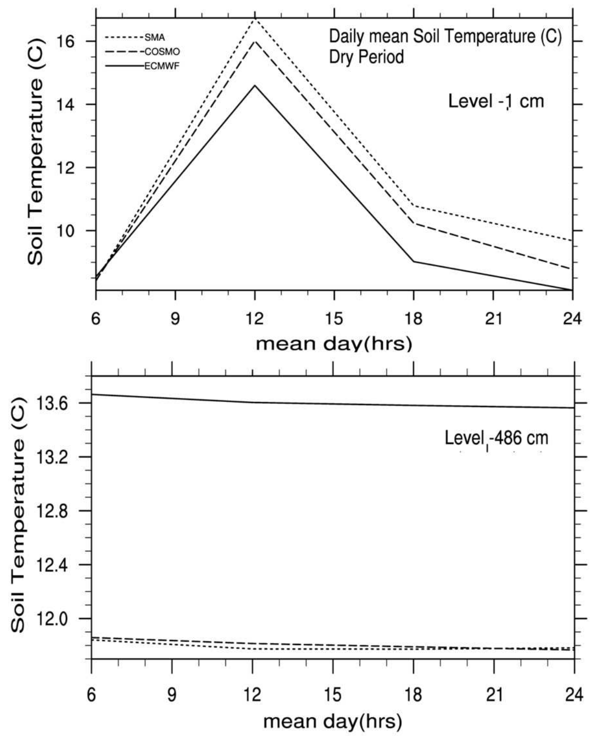

Figure 7: Mean day temperature profiles at two model levels as predicted by the three initialization scheme.

compared to the observations for the whole period. The fit accross the data is performed locally. That is, for the fit at given point x, the fit is made using points in a neighbourhood of x, weighted by their distance from x following [9]. These diagnostics are calculated for the whole period so different soil condition are sampled. On average the model underestimates both sensible and latent heat during daytime and nighttime. It is clear that even on domain mean statistics the improvements produced by the COSMO experiment with respecty to ECMWF interpolation largely offset the addictional benefit of also performing a soil moisture analysis. Similar conclusions can be drawn if looking at 2m diagnostics evaluated using the synop network over the whole COSMO-I7 domain and reported in figure 9. Most of the bias contribution is due to the cold bias at noon which is reduced by both the COSMO and SMA experiments. The nightime warm bias, even if smaller in size, is nevertheless very significative from the point of view of weather inpact. It is likely infact that these biases cause the missing prediction of local weather phenomena such as fog and brine formation.

4 Conclusions

Given the lack of any representative soil measurement network and the often insufficient knowledge of surface pedology, lithology, and vegetation characteristics, at the present soil analysis should be considered simply as a *tuned* lower boundary condition to drive at its best lower level atmospheric processes. The first and more obvious consequence of this strong assertion is that soil outputs from atmospheric models are not appropriate to drive for example hydrological models, as absolute soil moisture content strongly depends on the design of the soil model [6]. Onces this is given for acquired, there are a number of practical considerations which directly stems from this acknowledgment. They have been proved along this work in which different soil initialization methods have been compared in the quest for the best strategy to provide a soil analysis in regional models. They main findings are now

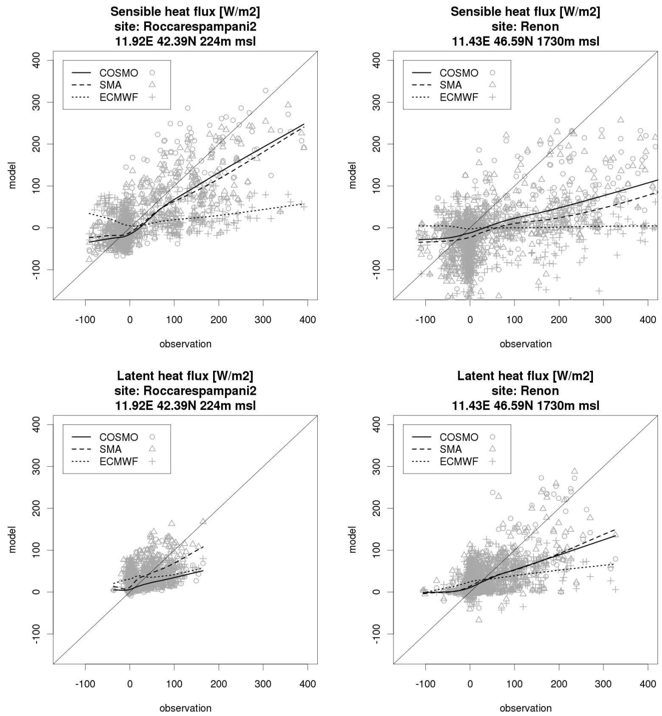

Figure 8: SON 2008 period statistics for the latent and sensible heat fluxes at two Carboeurope locations.

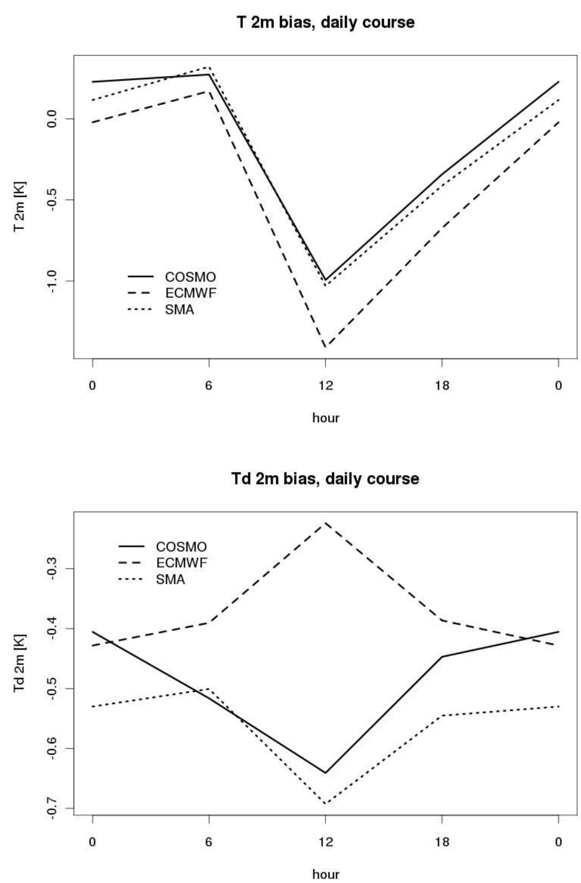

Figure 9: Global domain statistics of the T_{2m} variable as predicted by the three experiments. The bias is calculated using the 400 stations of the synop network.

synthesized in this conclusive section.

Firstly, there are serious problems in trying to use soil analysis derived from a global model into a limited area model if the two models are different. While this is certainly a suitable option for the atmospheric fields which to some extend represent the true atmospheric state it looses any validity for the soil prognostic variables if, as it is at the present, they only represents parameters to be tuned to compensate for various model biases. It is clear that SVAT models with different characteristics will also adjust to different thermal and hydrological equilibria with little connection to the reality. If the global model and the regional one have varying assumptions, the consequences on the surface fluxes estimation can be severe especially under selective circumstances (i.e. dry condition, little advection, large cloud cover, etc). In our study we found, for example, that initializing the regional model COSMO SVAT moudule TERRA by interpolation from the global IFS model induce in COSMO a strong warn nighttime bias and a weakening of the diurnal cycle in both latent and sensible heat surface fluxes. The cause was identified in the different specification of the soil thermal conductivity between the two models. What therefore could be appropriate (.i.e. 'tuned') for IFS is probably not optimal when used into another model.

Simply using COSMO in a freely running configuration can avoid these imbalances and it is found a winning and therefore recommended strategy especially for situation when the soil is close to its wilting point. This condition are particularly challenging since higher amplitudes in soil heating during daytime and cooling during night are expected. The correct prediction of the soil vertical temperature profile becomes then fundamental. Soil temperatures profiles which are in balance with the soil scheme are the zero order factor for correct surface flux estimations. Any subsequent improvement in the soil moisture estimation performed, for example, using indirect measurements such as T_{2m} only adds a marginal benefit. This apparently striking result is justified by the design of current soil mositure schemes which, using near surface observations, as predictors for soil moisture increments, assume that atmosphere is informative about soil moisture. However, forecast errors of atmospheric temperature and humidity do not always contain useful information. For instance, during rain, at nighttime, and with low solar insulation, this method is likely to fail. In the COSMO implementation the soil moisture correction is evaluated at two instants around noon when the soil-atmosphere coupling is supposed maximum. The major benefit from the soil moisture analysis arises therefore primarily for situations the scheme is designed for i.e. correcting forecast errors at the time the observations are assimilated which is around noon. In situation in which the bias does change its sign during the day, it is intuitive that the scheme can dangerously act to exacerbate the bias already present. A possible, and already proposed solution [2] would be a selective application of soil moisture increments only on those cases for which the underline hypothesis are stickily verified. This is nevertheless again an 'ad hoc' solution which doesn't help envisage a substantial improvement of the absolute quality of soil analysis.

The use of other observations such precipitation, satellite derived surface soil moisture can help on future soil moisture analysis. Nevertheless it is clear from this study that a substantial benefit can immediately arise from including temperature as a control variable in the assimilation scheme and from a concreat refinement of the model itself as was also outiled by the intercomparison work done by [3].

Acknowledgment We are in debt with Martin Lange for providing the soil moisture analysis scheme and a series of usufull comments which helped clarifying the scope of the paper. Riccardo Valentini (University of Tuscia, Viterbo), Giorgio Matteucci e Vincenzo Magliulo (CNR ISAFoM) are thanked for providing the data from CARBOEUROPE sites. This work is part of the activities of ARPA-SIMC funded by the Italian National Department for Civil Protection (Protezione Civile Nazionale, Roma).

References

[1] Douville, H., P. Viterbo, J. Mahfouf, and A. Beljaars, 2000: Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data. *Monthly Weather Review*, **128 (6)**, 1733–1756.

[2] Hess, R., M. Lange, and W. Wergen, 2008: Evaluation of the variational soil moisture assimilation scheme at Deutscher Wetterdienst. *Quarterly Journal of the Royal Meteorological Society*, **134 (635)**.

[3] Jacobs, C., et al., 2008: Evaluation of European Land Data Assimilation System (ELDAS) products using in situ observations. *Tellus A*, **60 (5)**, 1023–1037.

[4] Mahfouf, J. F., 1991: Analysis of Soil Moisture from Near-Surface Parameters: A Feasibility Study. *Journal of Applied Meteorology*, **30 (11)**, 1534–1547.

[5] Robock, A., K. Vinnikov, G. Srinivasan, J. Entin, S. Hollinger, N. Speranskaya, S. Liu, and A. Namkhai, 2000: The global soil moisture data bank. *Bulletin of the American Meteorological Society*, **81 (6)**, 1281–1299.

[6] Teuling, A., R. Uijlenhoet, B. van den Hurk, and S. Seneviratne, 2009: Parameter Sensitivity in LSMs: An Analysis Using Stochastic Soil Moisture Models and ELDAS Soil Parameters. *Journal of Hydrometeorology*, **10 (3)**, 751–765.

[7] Viterbo, P. and A. Beljaars, 1995: An Improved Land Surface Parameterization Scheme in the ECMWF Model and Its Validation. *Journal of Climate*, **8 (11)**, 2716–2748.

[8] Western, A., G. Bl
”oschl, and R. Grayson, 1998: Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment. *Journal of Hydrology*, **205 (1-2)**, 20–37.

[9] Cleveland, W., E. Grosse, W. Shyu, J. Chambers, and T. Hastie, 1991: Statistical models in S. *Wadsworth and Brooks/Cole, Pacific Grove, Ch. Local regression models*, 309–376.

[10] Lange, M., 2009: Parametrisation of the sensitivity $dT2m/dw$ in soil moisture analysis. Tech. Rep. -, Deutscher Wetterdienst, available at Deutscher Wetterdienst Frankfurter Str. 135 63067 Offenbach Germany.