*Revised version of 27/09/07 (small updated on 07/02/08)*
*First submitted to WAF on 02/03/07, accepted on 09/10/07*

# A Verification of Numerical Model Forecasts for Sounding–Derived Indices above Udine, NE Italy

Agostino Manzato
agostino.manzato@osmer.fvg.it

Keywords: soundings, instability indices, model verification.

*OSMER – Osservatorio Meteorologico Regionale dell'ARPA Friuli Venezia Giulia*
*Via Oberdan 18/a, I–33040 Visco (UD), Italy, tel. +39 0432 934163*

## Abstract

In this work, forty different indices derived from real soundings and the corresponding ECMWF model forecasts for the same location (near Udine, NE Italy) are compared. This comparison is repeated for more than 500 days, from June 2004 through December 2005. The comparison is performed in terms of linear correlation, bias, root mean squared error and relative absolute error. The results show that the quality of agreement varies considerably for the different indices and for the different lead times.

It is found that only fifteen indices have strong correlations ($R^2$ always larger than 0.5) which hold up to the fourth forecast day (+96 h). The indices having best agreement are usually representative of the thermal fields while those having worst agreement are usually representative of the wind fields, particularly in the low levels. It is also found that the correlations for all the forty indices change for the four times of the day (one sounding every 6 h), with the 12 UTC comparison being usually the best one. Also, at 18 and 00 UTC the potential instability is strongly underestimated by the model.

To test if these errors depend on the vertical resolutions a similar comparison has been done between the observed sounding reduced in TEMP format (which are assimilated by the GCM model) and the originals at high vertical resolution. It is found that the indices derived from the TEMP format vary from those with the highest resolution, in particular the indices which depend on the details of the low level profile, such as the level of free convection and the initial height of the lifted parcel. Nevertheless, only the shear has a very low correlation ($R^2 = 0.15$) with the corresponding index derived from the high resolution soundings, because it is very sensitive to the number of vertical levels used.

# 1 Introduction

This work is different from the many already done in the field of model verification because it is based on sounding–derived indices and is end–user oriented. The success of the "European Center for Medium–Range Weather Forecasts" (hereafter ECMWF) model is here assessed based on the errors made by this model when forecasting the sounding–derived indices on a *single horizontal* grid–point (the end–user location). Furthermore, no model reanalysis has been used in this verification, but only true forecast and observed sounding data. Only a few works in the literature deal with the model vertical profile perspective, such as Lesht et al. (2004) for temperatures and mixing ratios or Thompson et al. (2003) for +1 h model forecast of temperatures, mixing ratios and wind speeds, and in particular with sounding indices, such as Hart et al. (1998) for CAPE, CIN and Helicity, but none with all the forty indices studied here. That is particularly important when the model is used for forecasting convective events, which depend on the atmosphere thermodynamic profile. For example, two relatively small errors at the mandatory level temperature, such as a decrease of 0.5 K at 925 hPa and an increase of 0.5 K at 500 hPa, could largely affect the evaluation of the vertical profile instability. Lastly, the statistical study presented here involves hundreds of cases and so expands upon the few case-study verifications reported by others.

Not only the mandatory five levels 925, 850, 700, 500 and 300 hPa (which is also the highest level purchased by OSMER–ARPA FVG) are considered, but also the twenty seven hybrid levels (for a total of *thirty two* vertical levels) forecast below 300 hPa by the 12 UTC run of the ECMWF T511 L60 deterministic model (Simmons et al. 1989, Teixeira 1999, Klinker et al. 2000), for the atmospheric column above LAT=46.0 deg and LON=13.0 deg, are used to build a "pseudo–sounding", sometimes called "pseudo–TEMP". This is done every 6 h, for all the forecast lead times purchased by the OSMER–ARPA FVG meteorological center (starting from +24 and ending at +132 h). These pseudo–soundings have been analyzed with the Sound_Analys.PY software (Manzato and Morgan 2003, hereafter MM03) to derive a multitude of "pseudo–indices".

The same analysis has been performed on the database of the real soundings, made every 6 h by the Italian Aereonautica Militare at the Udine–Campoformido station (ID code 16044), located at LAT=46.04 deg, LON=13.19 deg, ALT=94 m above m.s.l.. This sounding launch base is located in the middle of the Friuli Venezia Giulia (hereafter FVG, NE Italy, shown in Figure 1) plain and its horizontal displacement from the model gridpoint chosen (LAT=46.0 deg, LON=13.0 deg) is about 15 km. Note that the T511 L60 orography has an elevation of 188 m at the studied gridpoint[1], so that there is also a vertical displacement of +94 m, which can influence the comparison between observed and forecasted fields in the very low levels. Figure 1 shows a map of the FVG region with the true orography in filled contours and the ECMWF surface elevation as numerical values, while Figure 2 is a vertical cross–section along the LON=13.0 meridian, that shows how much complex the gradient of the terrain elevation is.

The comparison between real and pseudo indices is necessary because OSMER–ARPA FVG has developed a tool, based on neural networks, which uses the sounding–derived indices to perform thunderstorm and rain forecasts in the FVG plain for the next 6 h (see Manzato 2005 a and Manzato 2007 b for details). In order to extend this work far into the forecast time, one has to use the model pseudo–soundings and the derived pseudo–indices. So, it is necessary to know which are the pseudo–indices that are well predicted by the model, in order to use only those for the development of the new model–based tool (to be done in a following article).

The next section presents in detail the data used in this work. Section 3 shows the results obtained by the first forecast step available (+24 h) for all the indices. Section 4 shows a similar comparison but using the indices derived by the observed soundings reduced in TEMP format, to test the vertical resolution sensitivity. Section 5, shows how the pseudo–

[1]The original T511 L60 orography is not defined in the gridpoint (LAT=46.0 deg, LON=13.0 deg), but on two points very close. One point, about 0.18 deg on the North side, has an elevation of 416 m, while the other, about 0.18 deg on the South, is at -12 m above m. s. l. (because it is near the coast). The 188 m is the value on the point (LAT=46.0 deg, LON=13.0 deg) interpolated by the ECMWF MARS database (Dominique Lucas, ECMWF, personal communication).

index agreement degrade with longer forecast lead times (up to +132 h), while Section 6 expresses some concluding remarks.

## 2 Data and one example

Since June 2004 OSMER–ARPA FVG purchases the thirty two hybrid levels from surface up to 300 hPa above the (LAT=46.0 deg, LON=13.0 deg) gridpoint of the ECMWF T511 L60 model run every day at 12 UTC[2]. The lead times available at the OSMER–ARPA FVG database are from +24 to +84 every 6 h and after that every 12 h up to +132. Since January 2005 the observed soundings and the pseudo–soundings derived by ECMWF deterministic model are plotted as thermodynamic diagrams in the OSMER–ARPA FVG intranet. These diagrams are shown in two formats: the Skew–T function for the "Grid Analysis Display System" (http://grads.iges.org/grads, hereafter GrADS) called Plotskew.GS, developed at the Pennsylvania State University by Robert Hart (see http://moe.met.fsu.edu/~rhart/skew.html), and the Thetaplot.GS function, developed for the GrADS software by Arturo Pucillo at OSMER–ARPA FVG (see http://www.fisica.uniud.it/~osmer/RnD_group/Arturo_HP/Thermodiag.html). On both the diagrams some of the sounding–derived indices computed by the Sound_Analys.PY software (MM03) are added. Looking at these products, local forecasters have sometimes noted large differences between observed and forecast soundings and in particular between the indices displayed. Also to check the consistency of these errors the present statistical study was undertaken.

Figure 3 shows an example of the Thetaplot diagram, first introduced by Morgan (1992). The Thetaplot has been chosen for this figure because it is easier to evaluate the potential instability with this diagram than using the Skew–T, as described also in MM03. Figure 3 a is the "real" Udine–Campoformido sounding measured at 12 UTC of 31 July 2005, made

[2]Since January 2006 also the 00 UTC run has been purchased, but the database collected at OSMER is still too short for a statistical study.

using all the raw data measured by a Vaisala RS-92 radiosonde. Note that these raw data are sampled each two seconds, which means that there are almost $10^3$ vertical levels between the surface and 300 hPa (because it takes about half an hour to do this part of the ascent[3]). This sounding was chosen as one of the many possible examples because it has the second greatest value of CAPE ever measured in the last 14 years (and the greatest of the studied database) and, in spite of that, it is interesting to say that thunderstorms were not observed in the FVG plain, but only in the mountains.

Figure 3 b shows the thirty two levels made by the +24 h forecast of the ECMWF model for the same time: it is possible to see how the forecast profile resembles the general mean profile of the observed sounding, particularly above 900 hPa. But looking at the lowest 100 hPa only the saturated equivalent potential temperature ($\Theta_{es}$) is really similar for the forecast and the observed values, apart from the very thin superadiabatic superficial layer which is about 10 K warmer in the observations. In contrast, the equivalent potential temperature ($\Theta_e$) and the dew–point equivalent potential temperature ($\Theta_{ed}$) are much higher than those forecasted. That is, the mixing ratio is underestimated in the lowest 100 hPa. In particular, the observed 30 hPa thick most unstable parcel (hereafter MUP), that is centered at about 140 m above the surface, has a mean $\Theta_e$ 12 K warmer than that forecasted. That means that the sounding instability is strongly underestimated by the model. In fact, the observed lifted index (Galway 1956) is -7 K while the forecast is -2 K, the MaxBuo (see MM03) is 21 K instead of 7 K and the CAPE is 3.3 kJ kg$^{-1}$ instead of 0.9 kJ kg$^{-1}$. The winds in the lowest 200 hPa are also quite different: the forecast shows a wind blowing from SW, while the observation measures a SE wind.

One may wonder if the differences between the sounding–derived indices and the pseudo–indices are affected by the different vertical resolution of the data ($10^3$ vs. thirty two levels). To test it, Figure 3 c shows the same observed sounding "degraded" in the vertical

[3]Note that the whole sounding takes about one hour and half to be completed and that is why it is usually launched one hour *before* the synoptic reference hour. So, the indices derived from the sounding of "12 UTC" are indeed computed from measures made between 1100 and 1130 UTC. Note also that the Italian local time is one hour more than UTC, which become two hours during Summer Time.

resolution by the TEMP format (FM 35 code, see the WMO Manual on Codes number 306), downloaded from the University of Wyoming upper air database (http://weather.uwyo.edu/upperair/sounding.html). The TTAA (mandatory) plus TTBB and PPBB (significant variations from linear trend) TEMP levels are just a minority of the original raw levels: in this case there are only twenty five levels from surface up to 300 hPa. The Thetaplot diagram of Figure 3 c is very similar to that of Figure 3 a, even if there are some differences, such as the $\supset$ shaped moist advection in the lowest 55 hPa, which is approximated by a straight line in the TEMP, or the temperature cap at 883 hPa, which is completely smoothed out in the TEMP format.

Comparing the TEMP–derived indices with the raw sounding–derived indices it is possible to find that there are significant differences. In fact, if one computes the absolute percentage error $|100\,(tX - X)/X|$ – where $X$ is an index derived from the raw data and $tX$ is the corresponding index derived from the TEMP sounding – for all the indices shown in Figure 3, but the CIN and SWISS (which have values very low respect their climatological values so that small differences can results in excessively large errors), then one obtains a mean error of 19%. On the other hand, if one repeats the same exercise on the pseudo–indices, then the mean $|100\,(pX - X)/X|$ – where $pX$ is the pseudo–index corresponding to $X$ – is 42%, i. e. more than double.

In any case, it is important to note that the indices derived by the TEMP sounding with low vertical resolution could have remarkable differences with respect those derived using the high resolution sounding: e. g. 45% in LFC, -37% in LI, 21% in VFlux, -14% in MaxBuo. In fact, the GCM models usually ingest the TEMP format sounding in their assimilation scheme, so that the errors due to the lower vertical resolution of the TEMP format are propagated through the model initialization.

Of course, the index errors seen in the example of Figure 3 apply only to this *single* case, while, to draw some significant conclusions, one has to perform a statistical study on many cases. For that reason, all the Udine–Campoformido soundings and the pseudo–soundings from June 2004 through December 2005 were analyzed with the SOUND_ANALYS.PY software and the derived real and pseudo indices were compared. The soundings were made

with a Vaisala RS-90 rawinsonde before July 2005 and with a rawinsonde RS-92 after that. The total number of studied cases is about 540 for each of the four synoptic hours of the day, but this number changes with the hour of the day[4], the forecast time and the index considered (different number of missing values). Lastly, to test the sensitivity to the vertical resolution, also the soundings in TEMP format from June 2004 through December 2005 were analyzed and a comparison with the raw sounding–derived indices has been performed.

## 3 The first day forecast

In this section attention will be focused on the +24 h forecast of the ECMWF T511 L60 12 UTC run, which is the first output step purchased at the OSMER–ARPA FVG. Table 1 summarizes the mean errors found between the pseudo and real indices considering all the 12 UTC soundings which have a corresponding +24 h model forecast. For the description of each index the reader is referred to Manzato (2007 b) and in particular to Manzato (2003), which contains also an evaluation of their relative value for local forecasting of different meteorological events. Forty different indices are considered, because not all the indices computed by the Sound_Analys.PY can be calculated from the pseudo–soundings. For example, the model forecast do not have information about the balloon vertical velocity.

For each index (let us call it $X$) it is shown the linear determination coefficient, $R^2$, with the corresponding pseudo–index ($pX$) and also the coefficients, $a$ and $b$, of the least squares fitting line ($X \cong a \cdot pX + b$). The next column shows the mean error between forecast and observation: bias $= \overline{pX - X}$, where the overbar means the mean value. Then there is the root mean squared error (RMSE), that is the square root of the mean squared error,

[4]Since 1986 the Italian Aereonautica Militare performs the Udine–Campoformido soundings every 6 h. Since October 2004, some economic problems caused soundings to be discontinued at 06 and particularly at 18 UTC, or made them with radiosonde (Vaisala RS-80 Navaid) instead of rawinsonde, that is, without measuring the winds. Since February 2006 only 00 and 12 UTC soundings are performed. This problem has been aggravated during the month of April 2006, when also the 00 UTC sounding was suspended.

MSE = $\overline{(pX - X)^2}$. There follows the relative absolute error (RAE), which is the mean absolute error (MAE = $\overline{|pX - X|}$) made relative dividing by the mean absolute deviation of the observations, that is RAE = MAE/($\overline{|X - \bar{X}|}$) . RAE is very similar to MAE, which is considered better than RMSE because less sensitive to large forecast errors (outliers) as discussed by Willmott and Matsuura (2005). Note also that RAE can be larger than 1. Finally there is the total number of valid cases available for both the real and pseudo indices[5].

Table 1 is ordered for $R^2$ decreasing. $R^2$ has been chosen as the most important score because it is invariant under shifts or rescaling, including bias adjustment. The index with the best correlation is Thetae, that is the $\Theta_e$ of MUP (chosen as the 30 hPa thick parcel in the lowest 250 hPa with the maximum $\Theta_e$), which has a $R^2 = 0.97$ (p–value $< 2.2 \cdot 10^{-16}$) [6] and a bias of only -0.4 K. That is very important, because the MUP is the parcel which is adiabatically lifted to compute many instability indices. So, if the initial parcel $\Theta_e$ has a large error, then this error would be propagated to many other indices.

The worst index studied is the wind shear in the lowest 3 km (Shear3), which has $R^2 = 0.11$ (p–value $= 1.1 \cdot 10^{-15}$). That means that the hodograph path in the boundary layer seems to be not well forecasted by the ECMWF model if compared with the high resolution observed hodograph. In the next section this issue will be discussed in a deeper way. Note that also the helicity in the lowest 3 km, the mean wind intensity in the lowest 500 m (LLWspd) and in particular its North–South component (LLWv) have low determination coefficients: 0.37, 0.35 and 0.26 respectively for Hel, LLWv and LLWspd.

It seems strange that the low level wind problem could be influenced by the 94 m of vertical difference between the surface station elevation and the model orography (shown in Figure 1) since the East–West component has a higher correlation: $R^2 = 0.50$. In the

[5]For the DTC (difference of MUP temperature at the -15 °C parcel level) pseudo–index there are many more missing values than for the others, due to a strong constraint in the SOUND_ANALYS.PY program, which afterward has been relaxed. LFC and EL have fewer cases because they are defined only for potentially unstable soundings.

[6]If not specified differently, all the correlations in the text body have p–values $< 2.2 \cdot 10^{-16}$.

free atmosphere the wind features are well captured by the model, more for the East–West component ($R^2 = 0.86$) than for the North–South component (MLWv has $R^2 = 0.76$). The problem found on the low wind field could be explained by the fact that the Udine–Campoformido location is only 40 km from 2000 m high mountains (Julian and Carnic Alps on the North side). Probably this complex orography (see for example Figure 2) interacts strongly with the wind field and the model is not able to describe well these local effects. Lavagnini et al. (2006) suggest that the low spatial resolution of ECMWF orography was not sufficient to reproduce meso–scale processes and that leads to errors in forecasting wind profiles in the first hundreds of meters.

Figure 4 shows the scatterplot for the indices with the best (Thetae, 2 a) and the worst (Shear3, 2 b) correlation coefficients, to give a visual idea of the whole fitting range. Together with the scatterplots are shown also the two histogram distributions and some scores: the regression line (plotted as a dashed line), R (not $R^2$), and finally the bias and RAE errors. Figure 4 a is an example of good association ($R^2 = 0.97$), while Figure 4 b shows a really low agreement ($R^2 = 0.11$) between the measured and forecasted shear: note how the forecast distribution is clustered around very low values, while the observed one has a long tail of Shear3 $> 10^{-2}\,\mathrm{s}^{-1}$. The Shear3 bias is $-6 \cdot 10^{-3}\mathrm{s}^{-1}$.

A very important aspect of a sounding, in particular for forecasting thunderstorms, is whether or not it is *potentially unstable*, that is, if it is possible to find a Level of Free Convection (LFC) for the lifted parcel. In Manzato (2003) it is said that this condition is equivalent to the condition MaxBuo $\geq 0$. Technically speaking that is not completely correct, since the definition of MaxBuo used is the difference between the maximum $\Theta_e$ in the first 250 hPa layer (MUP $\Theta_e$) and the minimum $\Theta_{es}$ in the "second" 250 hPa layer (typically between 750 and 500 hPa). It may happen that the lifted MUP finds a level with $\Theta_{es}$ lower than its $\Theta_e$ only in the first 250 hPa, but not above, that is, the buoyancy is confined to the low levels because both the LFC and the Equilibrium Level are very low. The existence of soundings potentially unstable only in the first 250 hPa explains why the Udine–Campoformido soundings considered unstable because they have a LFC are about 10% more frequent than those considered unstable because MaxBuo $\geq 0$. This

statistical difference becomes only 6% if one considers only the restricted period from April to September.

If one looks at the soundings with an existing LFC, then one finds that 52% of the real soundings are potentially unstable, while only 38% of pseudo–soundings have a LFC. In particular, only 32% of cases have contemporaneously a real and forecast LFC. So, the model forecasts as potentially stable 52-32=20% of cases which are not (miss cases), and forecasts as unstable 38-32=6% of cases which indeed are stable (false alarms). That means that the model has a Peirce Skill Score (hereafter PSS, e. g. Jolliffe and Stephenson 2003) for forecasting unstable sounding of PSS = POD - POFD = 32/52 - 6/48 = 0.49.

Not only the frequency, but also the magnitude of the instability is underestimated by the model, as can be seen looking at the bias of some instability indices like MaxBuo or DT500 (which is a lifted index evaluated using the MUP parcel instead of the mean lowest 500 m), which are respectively -2.7 K and +0.5 K. Note that DT500 regards negative values as unstable, differently from MaxBuo. The lower instability conditions of pseudo–soundings seems to be a quite common characteristic of the weather models. For example, Elmore et al. (2002) have found a significant bias toward positive values for the lifted index forecasted by the ETA Meso model (Mesinger et al. 1988, Black 1994) in comparison with the ETA reanalysis (taken to be the "true" value).

The instability indices similar to the lifted index have medium to high correlations: from $R^2 = 0.67$ of SWISS (Huntrieser et al. 1997) to $R^2 = 0.83$ of DT500. Strangely, the Showalter Index (ShowI, Showalter 1953) has a much lower correlation ($R^2 = 0.51$), showing the difficulty of forecasting well the initial parcel $\Theta_e$ at the 850 hPa level. The correlations for other frequently used indices are: $R^2 = 0.58$ for CAPE, $R^2 = 0.42$ for CIN and only $R^2 = 0.34$ for the Bulk Richardson Number (BRI). The indices which involve only the humidity field have in general high correlations: from $R^2 = 0.75$ for the mean relative humidity in the lowest 250 hPa (LRH) to $R^2 = 0.95$ for the precipitable water of the environment (PWE).

# 4 Comparing the vertical resolutions

It is interesting to check if the ECMWF pseudo–index errors shown in the previous section can be attributed to the different resolution of the vertical profiles. For that reason in Table 2 the same quantities of Table 1 have been computed comparing the 12 UTC observed soundings in high and low resolution data. To obtain it, all the TEMP ASCII files of the studied period have been downloaded from the University of Wyoming upper air database (http://weather.uwyo.edu/upperair/sounding.html) and analyzed with the Sound_Analys.PY software (MM03). The results obtained in Table 2 fix a sort of "background noise" for the model pseudo–indices performance.

The index with the lowest correlation is again Shear3 ($R^2 = 0.15$, p–value = 0.036) and that means that the low correlation ($R^2 = 0.11$) obtained by the ECMWF pseudo–index can be explained by the vertical resolution difference. The reason why the shear is so sensitive to the number of vertical levels is that it is the length of the hodograph, which is much longer following all the many wind vector variations on a curved path than approximating it by straight segments, as happens when using fewer levels. For that reason the Shear3 from low resolution data is almost always lower than from the high resolution data, as confirmed from its negative bias ($-5 \cdot 10^{-3} s^{-1}$).

All the other thirty nine indices have linear determination coefficients greater or equal to 0.59 (with p–values $< 2.2 \cdot 10^{-16}$) and, in particular, twenty four of them have $R^2 \geq 0.90$. Note that all the indices (but not EL) with $R^2 < 0.5$ in Table 1 are in the lowest part of Table 2, having $R^2 \leq 0.87$. Is it interesting to focus on the indices with lower correlation, that is, those more sensitive to the vertical resolution. In particular, the "altitude" variables in the low levels, like the level of free convection, LFC, the initial altitude of the most unstable parcel, h_MUP, the lifting condensation level, LCL, and the level closer to the melting point, MEL, are among those with the lowest correlations. Other variables with the lowest correlation are related to the detailed profile in the low levels, like the convective inhibition, CIN, the temperature caps, CAP, the wind in the lowest 500 m, LLW, and the Showalter and lifted indices (ShowI and LI respectively). The remaining indices are related

to the bulk shear (BRI) or to the hodograph area (helicity and EHI), which have problems similar to those of Shear3 [7].

Of all these indices, those which have more relevance from an operational forecast perspective are probably CIN, LI, LCL and LFC. Particular attention should be given to the lifted index, because it is often used for forecasting thunderstorms (e. g. Haklander and Van Delden 2003). The vertical resolution difference is due to the choice of the initial parcel, which has the mean characteristics of the lowest 500 m layer.

For the example shown in Figure 3, in the lowest 500 m there are fifty two levels in the high resolution format and *only two* (one at surface and one 27 m above, while the following is already at 670 m above m.s.l.) in the TEMP format. Figure 5 shows how the lifted index derived from the high resolution sounding of Figure 3 a changes when taking only one level every two out of all the originals, or four, or eight, or... only one level every sixty four, which means one measure every 128 seconds instead of every two seconds. To compare these different vertical resolutions with the TEMP one it is possible to say that the thirty two sampling rate has about thirty levels (plus the surface and five mandatory levels, which are never discarded in the test of Figure 5) up to 300 hPa, while the TEMP files analyzed have a mean value of twenty eight levels (plus surface and five mandatory levels) in the same layer, even if it must be noted that the TEMP tries to retain the "significant" levels. It is then possible to state that the thermodynamic features depending on the details of the very low levels are not well captured by the soundings degraded in the TEMP format, which are used to initialize the GCM models.

In conclusion, the sounding derived indices are found to vary with the vertical sampling resolution of the atmospheric profile, in particular those which use all the details of the low levels, like Shear3, LFC and h_MUP that have a correlation of $R^2 \leq 0.65$. Nevertheless, only the Shear3 pseudo–index has a very low correlation with the high resolution sounding–derived index ($R^2 = 0.15$, p–value $< 2.2 \cdot 10^{-16}$), which suggests that it is not wise to

[7] Note that the area inside the hodograph can be larger or smaller when using more levels on a "curved" hodograph than fewer levels on a "straight" hodograph, meanwhile the hodograph length is always larger.

compare the model derived Shear3 with the high resolution derived one.

If one repeats the same study for the soundings of the other three periods of the day (18, 00 and 06 UTC) then one obtains very similar results. In fact, the standard deviation of the $R^2$ coefficients computed for the four periods of the day is $\sigma(R^2) \leq 0.05$ for thirty three indices. The other seven standard deviations are: 0.07, 0.08, 0.09, 0.12, 0.17, 0.021 and 0.21 respectively for LFC, EHI, CIN, LCL, Rel_Hel, SWEAT and BRI. The high standard deviation of the SWEAT index is due to a $R^2$ of only 0.45 at 18 UTC, while that of the BRI is due to a $R^2$ of only 0.27 at 00 UTC.

# 5 Skill degradation with the forecast length

Lorenz (1963) has shown how the predictability of the atmosphere can be challenging, because even small variations of the initial condition can lead to divergent solutions. That is why it is reasonable to expect a quality degradation increasing with the forecast lead time. The question is to quantify this degradation.

Table 3 shows the same results shown in Table 1, but computed on the +132 h forecast (00 UTC of the sixth day), which is the last step purchased at OSMER–ARPA FVG. While in Table 1 there were twenty eight indices with $R^2 \geq 0.5$ (highlighted in bold), here there are only seven. Among these, very important are the MUP $\Theta_e$ ($R^2 = 0.88$) and DT500 ($R^2 = 0.54$). The best wind index is MLWu, which has $R^2 = 0.44$, while the best humidity field is PWE ($R^2 = 0.79$) but the second is LRH with only $R^2 = 0.34$. It is possible to say that in this case the majority of the indices are not useful for forecasting purposes, given their low correlation. That has operational consequences since some indices have been found to be good predictors for forecasting the thunderstorm likelihood in different parts of Europe (e. g. the DT500 found by Manzato 2003, or the SWISS index introduced by Huntrieser et al. 1997, or the lifted index using the mean lowest 100 hPa levels as initial parcel found by Haklander and Van Delden 2003) so that one would like to have a reliable forecast of them.

It is very interesting to note how the degradation is quite different for the different

indices. Figure 6 shows the scatterplot for DT500 at +24 h (4 a) and at +132 h (4 b). The point scatter is much more spread out in Figure 6 b, but the correlation is still high. Note that the two fitting lines are different, since the coefficients $a$ and $b$ differ at the two forecast lead times. The mean relative humidity in the lowest 500 hPa (MRH) decreases at +66 h (6 UTC of the third day) to almost the same R value that DT500 reaches at +132 h. In Figure 7 it is possible to see the scatterplot of MRH at +24 h (5 a) and at +66 h (5 b). $R^2$ decreases from 0.81 to 0.56 (by almost one third) in only 42 h.

Lastly, Figure 8 shows how the speed of the mean wind in the lowest 6 km (MLWspd) has an $R^2$ which decreases from 0.69 to 0.49 (29% less) in only 6 h. In fact, the +30 h forecast (18 UTC of the first day) has even more problems in forecasting the mean wind than the +36 h forecast. That is because not all the four periods of the day (0, 6, 12 and 18 UTC) are forecast with the same skill, as visible also from Table 4, which shows all the $R^2$ values for the twenty three indices with a mean $R^2$ greater than 0.4. The mean is done for all the forecast times available at OSMER–ARPA FVG (every 6 h from +24 to +84 h and every 12 h afterward up to +132 h). Figure 9 shows the time series of $R^2$ drawn only for seven subjectively chosen indices. From this figure and Table 4 it is possible to see that many indices (except MEL and CIN) have a *local minimum* at 18 UTC, that is +30, +54, +78 h.

Since the 18 UTC data are the smallest sample (because the Italian Aereonautica Militare has performed fewer 18 UTC soundings during part of the studied period), one could think that these results are due to the sample size. To check whether or not that is the reason, the tables and diagrams for all the indices relative to the other three times of the day (00, 06 and 12 UTC) have been recalculated on a smaller database (about 60%), sampled randomly from the original one. The results show insignificant differences from those obtained for the whole database (only at 00 UTC there is a slight decrease in $R^2$). Then, it seems that the worse $R^2$s obtained at the 18 UTC forecast are really due to ECMWF model failure.

It may be that the model has more difficulties during the transition between day and night or that the global observation network has a lack of observations at this time, so that

the model has not an adequate description of the atmosphere state at this particular time of the day as well as, for example, at 12 and 00 UTC. In fact, the number of soundings assimilated by the ECMWF model are only about thirty at 06 and 18 UTC, while about six hundred at 12 and 00 UTC[8] .

It is also interesting to study the model skill in forecasting unstable soundings. Looking only at the +30 h forecast (18 UTC of the first day), 66% of real soundings are unstable, 51% of pseudo–soundings are unstable, but only 46% of pseudo–soundings have a corresponding sounding with LFC. That means a PSS = POD - POFD = 46/66 - 5/34 = 0.55. It seems a large improvement with respect to the +24 h, but it must be noted how the "event climatology" (the frequency of real unstable soundings) in this case is farther from the 50% probability and that leads "easily" to higher PSS, as explained in Manzato (2005 b, see also Figure 11 of Manzato 2007 b).

Looking at the unstable soundings at +36 h (00 UTC of the second day) there are 57% of real soundings and only 26% pseudo–soundings with LFC, but 5% of unstable pseudo–soundings are associated with stable soundings. That leads to PSS = POD - POFD = 21/57 - 5/43 = 0.25, which is quite low. That situation is aggravated by the fact that the MaxBuo and DT500 bias are -4.3 K and +1.5 K respectively. So, the ECMWF model strongly underestimates the potential instability in the FVG plain at 00 UTC. Finally, at +42 h (06 UTC of the second day) the soundings with LFC are 50% while the pseudo–soundings are 22%, but 6% of them correspond to stable soundings. That leads to a result even worse than the previous one: PSS = POD - POFD = 16/50 - 6/50 = 0.20. It is not by chance that the lowest PSS is obtained for the sample with one half (50%) probability of occurrence. In any case, the instability magnitude is estimated slightly better than before, since the MaxBuo and DT500 bias are -4.0 K and +1.3 K.

Table 4 and Figure 9 focus only on the $R^2$ metric, while in the first three tables also

[8]Remember that the 4D-var assimilation system implemented at ECMWF (Rabier et al. 2000, Bouttier 2001) uses all the data observed between 3 and 15 UTC to initialize the "12 UTC run of today". This assimilation reduces the forecast errors made by the previous run of "00 UTC", which had assimilated all the data between 15 UTC of the previous day and the 3 UTC of today.

other error measurements are shown. It should be noted that the results obtained for the $R^2$ are not always exactly reproduced in the other metrics. For example, in Figure 10 one can see the RAE time series of the same seven indices shown in Figure 9. Comparing the two figures it is possible to see how the worst index is no longer CIN but the CAP (defined as the maximum increase of $\Theta_{es}$ with height in the lowest 250 hPa), in particular at 00 UTC, when RAE > 1.3. The high overestimation of the CAP intensity at 00 UTC (mean bias of 3.9 K) can partially explain the too low instability forecasted at that time, but not at 06 UTC, when the error is much smaller (mean bias of only 0.9 K). The instability underestimation can be seen also with the CIN index, which has a large RAE at 00 and 06 UTC, because its absolute value is usually much lower than forecasted (with a typical bias of about -190 J $kg^{-1}$).

The CIN has RAEs even better than the VFlux index (the mean North-South water vapour flux in the lowest 3 km), which has pronounced minimum errors only at 12 UTC (+24, +48, +72 and +96 h). Apart from Thetae, which has really low errors, the other variables have RAEs slightly increasing with the forecast time, typically from 0.5 to 0.8, and in general the best results are obtained at 12 UTC (especially for DT500). It is possible to say that the RAE analysis gives results different from the $R^2$ analysis, in particular the worst index rank changes because the worst period of the day seems to be 00 instead of 18 UTC.

Lastly, it is interesting to study also the correlations between the 6 h difference of an index and the same difference of the corresponding pseudo–index. For example, one wants to know how much the LI increases going from +24 (12 UTC) to +30 (18 UTC) and how this difference is forecast by the model. Since the errors made forecasting each index – such as LI(+24) and LI(+30) – could be very different, one expects to see larger errors in their difference [i. e. variables such as D_LI(24,30) = LI(+24) - LI(+30)] than those made forecasting the single index. In fact, only two 6 h difference indices have a correlation $R^2 > 0.3$, and those are the 6 h increase of MLWv between +24 and +30 h ($R^2 = 0.33$) and the 6 h increase of MLWu between +42 and +48 h ($R^2 = 0.32$). All the other 6 h difference indices are forecast with very low agreement, that is $R^2 < 0.3$.

# 6 Conclusions

More than 500 pairs of Udine–Campoformido soundings and ECMWF pseudo–soundings have been analyzed for each of the four periods of the day, deriving forty different indices for each vertical profile. The association between real sounding indices and the pseudo–indices has been studied in terms of linear correlation, bias, RMSE and RAE. It appears that these indices have a very different level of correlation, depending on the indices themselves.

As visible in Table 4, the ECMWF model is able to forecast with high correlation coefficients some MUP characteristics such as $\Theta_e$, mixing ratio (Mix), Lifting Condensation Level temperature (Tbase) and the level where the lifted *parcel* reaches 0 °C (parcel melting level, MEL). Also the *environmental* wet bulb zero (WBZ) shows a good agreement, as is for the environmental precipitable water content in the first 12 km (PWE). Other variables related to the water vapour field (MRH and LRH) are forecast in an acceptable way, let us say with $R^2 \geq 0.45$ always up to +96 h (12 UTC of the fourth day). Of these two important humidity variables, only the first step (+24 h) of MRH has a large correlation, that is $R^2 \cong 0.8$.

Looking at the temperature instability indices, very important results are obtained by DT500 and KI (George 1960), which have $R^2 \geq 0.69$ up to +72 h. In particular, DT500 has $R^2 \geq 0.6$ up to +120 h. MaxBuo, LI and SWISS show slightly worse results, while the similar Showalter index (ShowI) has always $R^2 \leq 0.5$. CAPE has $R^2 > 0.5$ only at +24, +48 and +72 h (that is, always at 12 UTC), while the "inhibition" variables CIN and CAP have always $R^2 \leq 0.44$.

The mean wind in the lowest 6 km (MLW) has high correlations up to +96 h, with $R^2 \geq 0.65$ for the zonal component and $R^2 \geq 0.52$ for the meridional component. Unfortunately, the lowest 3 km North–South wind combined with the vapour density to compute VFlux has $R^2 > 0.5$ only up to +72 h (but not at +54 h) and that limits the use of this index, which was found to be very important for the FVG rain forecast by Manzato (2007 a and 2007 b). The correlations for the wind in the lowest 500 m (LLW) are very low, in particular for the North–South component. Other indices related to the wind field, such as

Relative Helicity (Rel_Hel), Energy–Helicity (EHI) and Bulk Richardson Number (BRI), have low correlation coefficients, which suggest that their model forecast is not very useful for operational purposes.

It is then possible to cluster the forty indices in four main groups (listed for decreasing correlation):

1. indices very well predicted up to +132 h, with a mean (from +24 to +132 h) determination coefficient of $R^2 \geq 0.75$: Thetae, WBZ, PWE, Mix, MEL and Tbase;

2. indices well predicted only up to +96 h, with a mean (from +24 to +96 h) determination coefficient of $R^2 \geq 0.58$: MLWu, DT500, KI, MLWv, MaxBuo, MRH, DTC, LI, LRH;

3. indices which have always $R^2 \geq 0.35$ up to +72 h: b_PBL, SWISS, MLWspd, PWC, VFlux, SWEAT, UpDr, CAPE (apart at +66 h, which has $R^2 = 0.31$), ShowI;

4. the remaining sixteen indices are forecast with low correlation coefficients: for example their mean determination coefficient from +24 to +84 h is $R^2 \leq 0.36$, while all the other twenty four indices had values equal or larger than 0.40.

Part of the low correlation performances can be justified by the vertical resolution sensitivity, as discussed in Section 4. In particular, the Shear3 index computed from the TEMP format, that is using few vertical levels, has a very low correlation with that derived from the high resolution soundings ($R^2 = 0.15$, p–value = 0.036), so that it is not correct to compare the pseudo–Shear3 and the Shear3 indices. All the others indices computed from the TEMP format have a determination coefficient larger than 0.59.

It is interesting to note that the last height indices listed in Table 2 (namely LCL, BRI, CAP, EHI, CIN, h_MUP, LFC and Shear3, which have $R^2 \leq 0.77$) are all in the last category of the previous clustering. That suggests that the model low performances for forecasting these indices could be partially due to the low vertical resolution of the TEMP soundings assimilated by the model. The low level details can strongly affect the evaluation of the CAP, CIN, h_MUP and LFC values, so that it is probable that the model forecast of these

indices (may be also of others, like indicated in Figure 5 for the LI) could benefit from an increase of the levels reported in the TEMP format, particularly in the lowest layer, let us say the lowest 100 or 150 hPa. In fact, the GCM model continuously increases the vertical resolution (e. g. ninety one levels for the last T 799, with about forty below 300 hPa), while the TEMP format is still the same and below 300 hPa of the Udine–Campoformido sounding there are fewer than forty levels.

Another comment can be made about the low correlations obtained by the indices related to the wind fields in the low levels (LLW, Rel_Hel, Hel, EHI and BRI): it is probable that the model low performances are related to the complex orography of the FVG region, because of the strong meridional elevation gradient (e. g. Figure 2), which goes from 2000 m high mountains to sea level in about 70 km.

Particular emphasis has been given to the stability evaluation. The model always forecasts greater stability than that observed (strong bias found for MaxBuo, DT500, CIN and so on) and also the number of potentially unstable pseudo–soundings is underestimated, in particular at 00 and 06 UTC (PSS $\leq 0.25$). At least for that reason, it is possible to say that not all the four periods of the day are forecast with the same skill. In fact, the 12 UTC is generally forecast better than the other output times, while 18 UTC worse. For example, from +66 h (06 UTC) to +72 h (12 UTC) more than half of the indices have an *increase* in their R and the same happens from +78 UTC (18 UTC) to +84 h (00 UTC).

That could be due to a lack of soundings data at 06 and 18 UTC during the model assimilation process, since soundings were the main part of the observing system (Undé at al. 1997, Pailleux et al. 1997)[9]. Since these periods of the day are particular important for the day–night (and vice–versa) transition, it should be desirable that the number of soundings at these synoptic times will be increased from the about thirty assimilated by the ECMWF. In particular, the 18 UTC sounding is considered very important for forecasting

[9] In the last years satellite remote measures are competing with sounding measurements, while previously it was said that "Rawinsondes are unquestionably one of the most important sources of meteorological measurements. They form the backbone of the measurements system worldwide." (Gandin et al. 1993).

thunderstorms, since their frequency has a peak during afternoon and evening (see Manzato 2007 a).

As a final comment, it must be remarked that these results hold only for NE Italy (Udine–Campoformido location) and can say nothing on the ECMWF performance in other parts of its domain. But this is an interesting testing point because it is in the middle of a very populated plain located in a very complex area, having the Grado and Marano lagoon and, after that, the Adriatic sea only 30 km on the South, and the steep Alps chain 40 km on the North. In any case, that is the "end–user point of view" and in particular the point of view of local forecasters, who regard the ECMWF model as the "best one" among the many available at the OSMER–ARPA met office (Marcellino Salvador, Livio Stefanuto and Sergio Nordio personal communication).

Lastly, it should be very interesting to repeat this study in coming years, since in February 2006 ECMWF updated its deterministic model from T 511 L60 to T 799 L91. Also, to repeat in other places of the world the comparison between the indices derived from the soundings with high vertical resolution and those in the TEMP format could generalize the results found in Section 4.

## Acknowledgments

Arturo Pucillo (of OSMER–ARPA FVG, Italy) has done much work extracting the pseudo–soundings from the GRIB binary dataset and producing the pseudo–indices database. Jasna Vehovar (ARSO Slovenia) suggested the use of scatterplots with marginal distributions in R. The author wants to thank his wife and Griffith Morgan for English improvement and encouragement. The anonymous reviewers provided many suggestions to improve an older version of this paper. Last but not least, he thanks the "GNU generation" for having provided such good – and free! – tools as R, Python, LaTeX, Emacs, as well as Linux itself.

# References

Black, T. L., 1994: NMC Notes: The New NMC mesoescale Eta model: description and forecast examples. *Wea. Forecasting*, **9**, 256–278.

Bouttier, F., 2001: The development of 12-hourly 4D-Var. ECMWF Tech. Memo. **348**.

Corbet, J. M., Mueller, C., Burghart, C., Gould, K., and G. Granger, 1994: ZEB: Software for Integration, Display and Management of Diverse Environmental Datasets. *Bull AMS*, **75**, 783–792.

Elmore, K.L., Stensrud, D.J., and K.C. Crawford, 2002: Ensemble cloud model applications to thunderstorm forecasting. *J. Appl. Meteor*, **41**, 363–383.

Galway, J.G., 1956: The lifted index as a predictor of latent instability. *Bull. AMS*, **37**, 528–529.

Gandin, L.S., Morone, L.L., and W.G. Collins, 1993: Two Years of Operational Comprehensive Hydrostatic Quality Control at the National Meteorological Center. *Wea. Forecasting*, **8**, 57–72.

George, J.J., 1960: *Weather Forecasting for Aeronautics.* Academic Press, 673 pp.

Haklander, A. J., and A. van Delden, 2003: Thunderstorm predictors and their forecast skill for the Netherlands, *Atmos. Res.*, 67-68, 273-299.

Hart, R. E., Forbes, G. S., and R. H. Grumm, 1998: The use of hourly model-generated soundings to forecast mesoscale phenomena: Part I: Forecasting warm-season phenomena. *Wea. Forecasting*, **13**, 1165–1185.

Huntrieser, H., Schiesser, H. H., Schmid, W., and A. Waldvogel, 1997: Comparison of Traditional and Newly Developed Thunderstorm Indices for Switzerland. *Wea. Forecasting* **12**, 108–125.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. J. Wiley & Sons, 240 pp.

Klinker, E., Rabier, F., Kelly, G., and J.-F. Mahfouf, 2000: The ECMWF operational implementation of four–dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration. *Quart. J. Roy. Meteor. Soc.*, **126**, 1191–1215.

Lavagnini, A., Sempreviva, A. M., Transerici, C., Accadia, C., Casaioli, M., Mariani, S., and A. Speranza, 2006: Offshore wind climatology over the Mediterranean basin. *Wind Energy*, **9**, 251–266.

Lesht, B. M., Miloshevich, L. M., and J. C. Liljegren, 2004: Comparison of ECMWF Model Analyses with the Observed Upper–Air Temperature and Relative Humidity Climatology at the ARM NSA, SGP, and TWP Climate Research Facility Sites. Proc. of the *Fourteenth Atmospheric Radiation Measurement (ARM) Science Team Meeting*, March 2004 Albuquerque, New Mexico, sponsored by the U.S. Department of Energy.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Manzato, A., 2003: A climatology of instability indices derived from Friuli Venezia Giulia soundings, using three different methods. *Atmos. Res.*, **67–68**, 417–454.

———, 2005 a: The use of sounding derived indices for a neural network short-term thunderstorm forecast. *Wea. Forecasting*, **20**, 896–917.

———, 2005 b: An odds ratio parameterization for ROC diagram and skill score indices. *Wea. Forecasting*, **20**, 918–930.

———, 2007 a: The 6 Hours Climatology of Thunderstorms and Rainfalls in the Friuli Venezia Giulia Plain, *Atmos. Res.*, **83**, 336-348.

———, 2007 b: Sounding–derived indices for neural network based short–term thunderstorm and rainfall forecasts. *Atmos. Res.*, **83**, 349-365.

———, and G. M. Morgan, 2003: Evaluating the sounding instability with the Lifted Parcel Theory. *Atmos. Res.*, **67–68**, 455–473.

Mesinger, F., Janjić, Z. I., Nivckovic, S., Gavrilov, D., and D. G. Deaven, 1988: The step–mountain coordinate: model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Mon. Wea. Rev.*, **116**, 1493–1518.

Morgan, G. M. Jr, 1992: ThetaPlot, an Equivalent Potential Temperature Diagram. *Meteorol. Atmos. Phys.*, **47**, 259–265.

Pailleux, J. 1997: Impact of various observing systems on numerical weather prediction. Proceedings of the CGC/WMO Workshop held in Geneva, 7-9 April 1997 (Editor). WWW Technical Report No 18. WMO/TD No **868**.

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F., and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc*, **126**, 1143–1170.

Showalter, A. K., 1953: A stability index for thunderstorm forecasting. *Bull. AMS*, **34**, 250–252.

Simmons, A. J., Burridge, D. M., Jarraud, M., Girard, C., and W. Wergen, 1989: The ECMWF medium–range prediction models. Development of the numerical formulations and the impact of increased resolution. *Meteor. Atmos. Phys.*, **40**, 28–60.

Teixeira, J., 1999: The impact of increased boundary layer vertical resolution on the ECMWF forecast system. ECMWF Tech. Memo. **268**, Reading, UK.

Thompson, R. L., Edwards, R., Hart, J.A., Elmore, K.L. and P. Markowski, 2003: Close Proximity Soundings within Supercell Environments Obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261.

Undén, P., Kelly, G., Le Meur, D., and L. Isaksen, 1997: Observing system experiments with the 3D-Var assimilation system. ECMWF Tech. Memo. **244**, Reading, UK.

Willmott, C.J. and K. Matsuura, 2005: Advantages of the mean absolute error (MAE) over the root mean squared error (RMSE) in assessing average model performance, *Climate Research*, **30**, 79–82.

World Meteorological Organization, 1988: *Manual on Codes, International Codes, Vol. I and II*, World Meteorological Organization, Geneva, WMO–No. 306.

# TABLE CAPTIONS

Table 1. *Different error measures between the forty sounding–derived indices and the +24 h pseudo–index forecasts. Measures are linear determination coefficient* $R^2$*, least square fitting a and b coefficients, bias (forecast minus observation), root mean squared error, relative absolute error and number of cases with both observed and forecasted indices.* $R^2 \geq 0.5$ *are highlighted in bold.*

Table 2. *Same as Table 1 but comparing observed soundings in raw and TEMP formats at 12 UTC.* $R^2 \geq 0.5$ *are highlighted in bold.*

Table 3. *Same as Table 1 but for the +132 h forecast time.* $R^2 \geq 0.5$ *are highlighted in bold.*

Table 4. $R^2$ *values for different lead times for all the indices with* $\text{mean}(R^2) > 0.4$. $R^2 \geq 0.5$ *are highlighted in bold.*

# FIGURE CAPTIONS

Figure 1. *The Friuli Venezia Giulia region and its orography (in filled contours) with the Udine–Campoformido (WMO code 16044) sounding base in the middle of the plain and the ECMWF T511 L60 surface elevation gridded values (as kindly provided by Dominique Lucas, ECMWF). Courtesy of Andrea Cicogna (OSMER–ARPA FVG).*

Figure 2. *The vertical cross–section of the Friuli Venezia Giulia region along the LON=13.0 deg meridian (in shaded in gray) with the points representing the ECMWF T511 L60 surface elevation (connected by a black line). It is possible to see how the very complex terrain gradient is smoothed in the model orography.*

Figure 3. 3 a: *Thetaplot and hodograph of the Udine–Campoformido sounding observed on 31 July 2005 at 12 UTC (one measure every two seconds). The three thick lines are (from right to left):* $\Theta_{es}$ *(which intersects the environmental temperature on the isotherms),* $\Theta_e$ *(which intersects the wet bulb temperature on the isotherms) and* $\Theta_{ed}$ *(which intersects the dew point on the isotherms). This diagram is very useful to evaluate the potential instability, as explained in Morgan (1992) or Manzato and Morgan (2003). Also a few instability indices computed by the* SOUND_ANALYS.PY *software are displayed. Courtesy of Arturo Pucillo (OSMER–ARPA FVG).*

3 b: *Thetaplot diagram of the vertical profile forecasted by* ECMWF *run of 30 July 2005 at 12 UTC for the +24 h, that is the same time as in Figure 3 a. There are only thirty two levels from surface up to 300 hPa. Courtesy of Arturo Pucillo (OSMER–ARPA FVG).*

3 c: *Thetaplot diagram of the 31 July 2005 at 12 UTC sounding "degraded" in the vertical resolution by the TEMP format (FM 35, WMO manual on codes number 306), downloaded in text format from the University of Wyoming upper air database. This plot has been done using only 29 levels in the lowest 200 hPa. Note that, differently from the indices studied in the text, the indices shown in fig. 1 have been computed using the* **T** *scheme of MM03, that is, without the virtual correction. Courtesy of Arturo Pucillo (OSMER–ARPA FVG).*

Figure 4. 4 a: *Scatterplot with histogram distributions and linear fit for the Most Unstable Parcel* $\Theta_e$ *derived by sounding observations at 12 UTC* (Thetae) *and derived by model pseudo–sounding forecast* (pThetae) *made the day before (+24 h forecast). The linear*

*fit coefficients, correlation,* bias *and* RAE *shown are the best obtained by all the forty different indices.*
4 b: *Same as in 4 a but for the index with the worst correlation between observed and forecast values, that is the wind shear in the lowest 3 km.*

Figure 5. *How the lifted index changes when the original high resolution data (one level every two seconds) are degraded at different sampling rate. Sampling rate 1 means to use all the original levels while 64 means to retain one original level every 64 (98.4% of data are discarded). The* LI *error is about* 1 K *when using one level every ten, which means to have about 100 levels below 300 hPa. The* LI *derived by the* TEMP *format is* -9.7 K.

Figure 6. 6 a: *Same as before but for the temperature difference between the lifted parcel and the environment at 500 hPa* (DT500) *at +24 h.*
6 b: *Same as Figure 6 a but at +132 h (00 UTC of the sixth day). As usually during night the instability is strongly underestimated* (bias = 1.6 °C)*, anyway the correlation is still acceptable* ($R^2 = 0.56$).

Figure 7. 7 a: *Same as previous figures, but for the mean relative humidity in the lowest 500 hPa* (MRH) *at +24 h. The correlation coefficients (*$a \cong 1$ *and* $b \cong 0$*) show a good calibration, but the high* RAE *show also a high dispersion.*
7 b: *Same as 7 a, but at +66 h (06 UTC of the third day): in only 42 h the correlation and dispersion worsened a lot.*

Figure 8. 8 a: *Same as before but for the mean wind speed in the lowest 6 km* (MLWspd) *at +24 h. Note the forecast understimation* (bias = -2 m s$^{-1}$).
8 b: *After only 6 h (18 UTC of the first day) the correlation is much worsened. Note that at 18 UTC there are many less soundings during the studied period.*

Figure 9. *The time series of the linear determination coefficient* ($R^2$) *for seven indices, for all the time forecasts available at* OSMER–ARPA FVG. Thetae *and* DT500 *conserve high correlations during all the forecasts, with a local minimum at 18 UTC.* MRH *and* MLWv *have acceptable correlations up to +96 h and +84 h respectively.* VFlux *has*

*good* R *up to +72 h, but with minimums at 18 UTC. Note finally the strange behavior of* CIN *which has a local maximum at 18 UTC.*

Figure 10. *Same as the previous figure but for the* RAE *measure. While* Thetae *still outperforms the others,* DT500 *is here more similar to* MLWv*,* CAP *(but not at 00 and 06 UTC) and* MRH. VFlux *has a local minimum (low error) at 12 UTC, while* CAP *has very large errors at 00 UTC, because the night inversions are strongly overestimated* (bias *always larger than* 3.6 °C).

# TABLES

Table 1: results at +24 h (first day, 12 UTC). $R^2 \geq 0.5$ are highlighted in bold.

| rank | index | $R^2$ | a | b | bias | RMSE | RAE | cases |
|---|---|---|---|---|---|---|---|---|
| 1 | Thetae [K] | **0.97** | 0.96 | 13 | -0.4 | 2.9 | 0.16 | 575 |
| 2 | WBZ [m] | **0.95** | 0.92 | -2.4 | 181 | 313 | 0.26 | 576 |
| 3 | PWE [mm] | **0.95** | 0.96 | 0.5 | 0.3 | 2.2 | 0.19 | 577 |
| 4 | Mix [$g\,kg^{-1}$] | **0.88** | 0.99 | -0.02 | 0.08 | 1.31 | 0.29 | 575 |
| 5 | MLWu [$m\,s^{-1}$] | **0.86** | 1.29 | -0.6 | 1.0 | 2.4 | 0.45 | 574 |
| 6 | KI [C] | **0.85** | 0.97 | -1.1 | 1.5 | 6 | 0.36 | 550 |
| 7 | MEL [m] | **0.84** | 0.95 | 346 | -200 | 441 | 0.39 | 575 |
| 8 | DT500 [C] | **0.83** | 0.89 | -0.1 | 0.5 | 2 | 0.41 | 575 |
| 9 | Tbase [C] | **0.82** | 1.00 | -0.5 | 0.5 | 5.5 | 0.35 | 575 |
| 10 | MRH [%] | **0.81** | 0.92 | 3.8 | 0.5 | 7.4 | 0.41 | 575 |
| 11 | MLWv [$m\,s^{-1}$] | **0.76** | 1.21 | 0.1 | -0.2 | 2.6 | 0.51 | 537 |
| 12 | MaxBuo [K] | **0.75** | 0.87 | 2.3 | -2.7 | 4 | 0.71 | 575 |
| 13 | LRH [%] | **0.75** | 0.89 | 5.4 | 1.7 | 9 | 0.48 | 575 |
| 14 | LI [C] | **0.72** | 0.84 | 0.03 | 0.42 | 2.29 | 0.51 | 575 |
| 15 | DTC [C] | **0.72** | 0.86 | -0.2 | 0.5 | 1.8 | 0.57 | 321 |
| 16 | MLWspd [$m\,s^{-1}$] | **0.69** | 1.08 | 1.7 | -2.1 | 3 | 0.82 | 574 |
| 17 | b_PBL [$cm\,s^{-2}$] | **0.68** | 1.01 | 0 | 0 | 0.17 | 0.54 | 575 |
| 18 | SWISS [] | **0.67** | 0.84 | -1.4 | 2.1 | 3.9 | 0.69 | 571 |
| 19 | PWC [mm] | **0.65** | 1.02 | 2.1 | -2.3 | 9.3 | 0.40 | 575 |
| 20 | VFlux [$kg\,s^{-1}\,m^{-2}$] | **0.64** | 0.75 | -0.001 | 0.0003 | 0.010 | 0.60 | 572 |
| 21 | UpDr [$m\,s^{-1}$] | **0.58** | 0.99 | 2.4 | -2.3 | 7.8 | 0.42 | 575 |
| 22 | CAPE [$J\,kg^{-1}$] | **0.58** | 1.28 | 47 | -75 | 280 | 0.45 | 577 |
| 23 | SWEAT [] | **0.58** | 0.74 | 8.7 | 9 | 39.3 | 0.56 | 577 |
| 24 | \|VFlux\| [$kg\,s^{-1}\,m^{-2}$] | **0.55** | 0.71 | 0.003 | 0.001 | 0.009 | 0.69 | 572 |
| 25 | ShowI [C] | **0.51** | 0.66 | 2.0 | -0.2 | 2.9 | 0.75 | 574 |
| 26 | HD [cm] | **0.51** | 1.09 | 0.4 | -0.5 | 1.6 | 0.46 | 575 |
| 27 | LCL [m] | **0.50** | 0.91 | 315 | -134 | 816 | 0.61 | 575 |
| 28 | LLWu [$m\,s^{-1}$] | **0.50** | 1.22 | 0.8 | -0.9 | 2.4 | 0.75 | 571 |
| 29 | h_MUP [m] | 0.48 | 0.69 | 416 | -107 | 737 | 0.48 | 575 |
| 30 | Rel_Hel [$m^2\,s^{-2}$] | 0.45 | 0.77 | 23 | -14 | 59 | 0.69 | 573 |
| 31 | CAP [C] | 0.43 | 0.62 | 3.0 | -1.4 | 4.4 | 0.99 | 480 |
| 32 | EHI [] | 0.42 | 1.42 | 0.02 | -0.02 | 0.12 | 0.52 | 576 |
| 33 | CIN [$J\,kg^{-1}$] | 0.42 | 0.72 | -86 | -65 | 415 | 0.57 | 577 |
| 34 | Hel [$J\,kg^{-1}$] | 0.37 | 0.73 | 4.9 | 1.0 | 57.4 | 0.81 | 571 |
| 35 | EL [m] | 0.35 | 0.73 | 2806 | -815 | 2654 | 0.77 | 184 |
| 36 | LLWv [$m\,s^{-1}$] | 0.35 | 0.51 | -0.5 | 0.3 | 2.3 | 0.99 | 571 |
| 37 | BRI [] | 0.34 | 0.38 | 4.3 | 8.7 | 51.4 | 0.89 | 554 |
| 38 | LLWspd [$m\,s^{-1}$] | 0.26 | 0.70 | 1.4 | -0.6 | 2.4 | 0.92 | 571 |
| 39 | LFC [m] | 0.20 | 0.43 | 925 | 675 | 1263 | 1.22 | 184 |
| 40 | Shear3 [$s^{-1}$] | 0.11 | 0.46 | 0.009 | -0.006 | 0.007 | 1.89 | 571 |

Table 2: comparison between observed soundings in raw and TEMP formats at 12 UTC. $R^2 \geq 0.5$ are highlighted in bold.

| **rank** | **index** | **$R^2$** | **a** | **b** | **bias** | **RMSE** | **RAE** | **cases** |
|---|---|---|---|---|---|---|---|---|
| 1 | PWE [mm] | **0.99** | 1.00 | 0.1 | -0.1 | 0.7 | 0.03 | 550 |
| 2 | KI [C] | **0.99** | 0.98 | 0.3 | -0.1 | 1.2 | 0.07 | 547 |
| 3 | Thetae [K] | **0.99** | 0.99 | 4.2 | -1 | 1.6 | 0.08 | 546 |
| 4 | MRH [%] | **0.99** | 0.99 | -0.06 | 0.6 | 1.8 | 0.09 | 540 |
| 5 | LRH [%] | **0.98** | 0.98 | 0.5 | 0.4 | 2.6 | 0.12 | 546 |
| 6 | MLWv [$m\,s^{-1}$] | **0.97** | 1.09 | 0.06 | -0.08 | 1.0 | 0.18 | 546 |
| 7 | MLWu [$m\,s^{-1}$] | **0.96** | 1.04 | -0.1 | 0.2 | 1.0 | 0.17 | 546 |
| 8 | Mix [$g\,kg^{-1}$] | **0.96** | 1.06 | -0.2 | -0.2 | 0.8 | 0.16 | 546 |
| 9 | WBZ [m] | **0.96** | 0.92 | -22 | 210 | 319 | 0.23 | 550 |
| 10 | PWC [mm] | **0.96** | 1.08 | 0.6 | -1.3 | 3.6 | 0.13 | 546 |
| 11 | CAPE [$J\,kg^{-1}$] | **0.95** | 1.13 | 16 | -34 | 105 | 0.17 | 550 |
| 12 | DT500 [C] | **0.95** | 0.93 | -0.3 | 0.5 | 1.2 | 0.20 | 545 |
| 13 | SWEAT [] | **0.95** | 0.96 | 2.5 | -0.2 | 12.7 | 0.13 | 546 |
| 14 | HD [cm] | **0.94** | 1.11 | 0.09 | -0.19 | 0.6 | 0.17 | 546 |
| 15 | VFlux [$kg\,s^{-1}\,m^{-2}$] | **0.94** | 1.00 | 0.0001 | -0.0001 | 0.004 | 0.22 | 545 |
| 16 | UpDr [$m\,s^{-1}$] | **0.94** | 1.05 | 0.7 | -1 | 3.1 | 0.15 | 546 |
| 17 | EL [m] | **0.93** | 1.00 | 295 | -265 | 891 | 0.15 | 255 |
| 18 | Tbase [C] | **0.93** | 1.04 | 0.1 | -0.2 | 3.4 | 0.20 | 546 |
| 19 | MaxBuo [K] | **0.92** | 0.88 | 1.3 | -1.6 | 2.4 | 0.37 | 546 |
| 20 | DTC [C] | **0.92** | 0.94 | -0.2 | 0.3 | 1 | 0.26 | 456 |
| 21 | LLWu [$m\,s^{-1}$] | **0.92** | 1.08 | 0.3 | -0.4 | 1 | 0.31 | 442 |
| 22 | \|VFlux\| [$kg\,s^{-1}\,m^{-2}$] | **0.91** | 0.99 | 0.0003 | -0.0001 | 0.004 | 0.29 | 545 |
| 23 | b_PBL [$cm\,s^{-2}$] | **0.90** | 1.03 | 0.02 | -0.02 | 0.1 | 0.27 | 546 |
| 24 | MLWspd [$m\,s^{-1}$] | **0.90** | 1.09 | -0.09 | -0.44 | 1.25 | 0.31 | 546 |
| 25 | SWISS [] | **0.87** | 0.85 | -0.9 | 1.5 | 2.7 | 0.43 | 392 |
| 26 | LLWspd [$m\,s^{-1}$] | **0.87** | 1.06 | 0.4 | -0.6 | 1.1 | 0.43 | 442 |
| 27 | Rel_Hel [$m^2\,s^{-2}$] | **0.87** | 1.16 | 4.5 | -10.6 | 29.7 | 0.37 | 545 |
| 28 | LI [C] | **0.86** | 0.82 | 0.5 | -0.1 | 1.7 | 0.36 | 546 |
| 29 | LLWv [$m\,s^{-1}$] | **0.84** | 1.09 | -0.05 | 0.11 | 0.99 | 0.42 | 442 |
| 30 | ShowI [C] | **0.81** | 0.83 | 0.8 | 0.2 | 1.7 | 0.36 | 546 |
| 31 | MEL [m] | **0.80** | 0.87 | 711 | -380 | 585 | 0.53 | 546 |
| 32 | Hel [$J\,kg^{-1}$] | **0.80** | 1.06 | 2.3 | -3.4 | 30.2 | 0.41 | 546 |
| 33 | LCL [m] | **0.77** | 1.03 | -31 | -28 | 549 | 0.36 | 546 |
| 34 | BRI [] | **0.74** | 1.11 | 0.04 | -1.4 | 24.2 | 0.26 | 501 |
| 35 | CAP [C] | **0.74** | 0.65 | 1.2 | 1 | 3.3 | 0.60 | 510 |
| 36 | EHI [] | **0.71** | 0.91 | 0.02 | -0.02 | 0.09 | 0.27 | 549 |
| 37 | CIN [$J\,kg^{-1}$] | **0.70** | 0.82 | -63 | -29 | 300 | 0.33 | 550 |
| 38 | h_MUP [m] | **0.65** | 1.06 | -164 | 95 | 564 | 0.46 | 546 |
| 39 | LFC [m] | **0.59** | 0.70 | 43 | 815 | 1030 | 1.24 | 255 |
| 40 | Shear3 [$s^{-1}$] | 0.15 | 0.48 | 0.008 | -0.005 | 0.006 | 1.42 | 546 |

Table 3: results at +132 (6 days, 00 UTC). $R^2 \geq 0.5$ are highlighted in bold.

| **rank** | **index** | **$R^2$** | **a** | **b** | **bias** | **RMSE** | **RAE** | **cases** |
|---|---|---|---|---|---|---|---|---|
| 1 | Thetae [K] | **0.88** | 0.96 | 15 | -1.8 | 6 | 0.32 | 547 |
| 2 | WBZ [m] | **0.87** | 0.84 | 246 | 112 | 487 | 0.34 | 548 |
| 3 | PWE [mm] | **0.79** | 0.92 | 2.7 | -1.1 | 5.2 | 0.41 | 548 |
| 4 | Mix [$g\,kg^{-1}$] | **0.74** | 0.97 | 0.6 | -0.4 | 2.1 | 0.46 | 547 |
| 5 | Tbase [C] | **0.69** | 0.92 | 1.1 | -1.1 | 7.3 | 0.53 | 547 |
| 6 | MEL [m] | **0.63** | 0.99 | 282 | -251 | 634 | 0.61 | 547 |
| 7 | DT500 [C] | **0.54** | 0.82 | -0.8 | 1.6 | 3.5 | 0.77 | 541 |
| 8 | b_PBL [$cm\,s^{-2}$] | 0.45 | 0.77 | -0.06 | 0.09 | 0.22 | 0.90 | 547 |
| 9 | MLWu [$m\,s^{-1}$] | 0.44 | 1.05 | -1.0 | 1.1 | 4.0 | 0.79 | 548 |
| 10 | KI [C] | 0.39 | 0.71 | 3.4 | 0.1 | 13.6 | 0.72 | 269 |
| 11 | MaxBuo [K] | 0.39 | 0.70 | 3.2 | -4.5 | 6.6 | 1.12 | 547 |
| 12 | LI [C] | 0.38 | 0.63 | -0.01 | 1.02 | 3.65 | 0.86 | 547 |
| 13 | EL [m] | 0.36 | 0.82 | 2627 | -1263 | 2901 | 0.86 | 113 |
| 14 | SWEAT [] | 0.35 | 0.65 | 28 | -2 | 55 | 0.71 | 548 |
| 15 | DTC [C] | 0.35 | 0.68 | -0.5 | 1.3 | 3.0 | 0.95 | 282 |
| 16 | LRH [%] | 0.34 | 0.59 | 29 | -2 | 16 | 0.85 | 547 |
| 17 | MLWv [$m\,s^{-1}$] | 0.33 | 0.73 | -0.3 | 0.7 | 4.2 | 0.82 | 548 |
| 18 | SWISS [] | 0.32 | 0.62 | -1.7 | 3.2 | 5.9 | 1.05 | 547 |
| 19 | PWC [mm] | 0.32 | 1.01 | 6.8 | -6.9 | 15.9 | 0.60 | 547 |
| 20 | EHI [] | 0.30 | 2.63 | 0.05 | -0.07 | 0.24 | 0.60 | 547 |
| 21 | CAPE [$J\,kg^{-1}$] | 0.30 | 1.54 | 123 | -153 | 419 | 0.62 | 548 |
| 22 | LCL [m] | 0.29 | 0.71 | 580 | 43 | 990 | 0.83 | 547 |
| 23 | UpDr [$m\,s^{-1}$] | 0.25 | 0.98 | 5.6 | -5.5 | 11.9 | 0.64 | 547 |
| 24 | MRH [%] | 0.25 | 0.56 | 29 | -3 | 18 | 0.92 | 547 |
| 25 | CIN [$J\,kg^{-1}$] | 0.24 | 0.62 | -51 | -190 | 495 | 0.81 | 548 |
| 26 | HD [cm] | 0.23 | 1.41 | 0.8 | -0.9 | 2.3 | 0.64 | 547 |
| 27 | MLWspd [$m\,s^{-1}$] | 0.19 | 0.56 | 4.0 | -2.0 | 4.0 | 1.10 | 548 |
| 28 | ShowI [C] | 0.19 | 0.45 | 2.4 | 0.4 | 4.0 | 1.01 | 543 |
| 29 | VFlux [$kg\,s^{-1}\,m^{-2}$] | 0.17 | 0.40 | -0.003 | 0.007 | 0.020 | 1.18 | 548 |
| 30 | CAP [C] | 0.16 | 0.28 | 5.8 | 4.2 | 8.6 | 1.53 | 546 |
| 31 | h_MUP [m] | 0.14 | 0.40 | 779 | 212 | 905 | 0.91 | 547 |
| 32 | LLWspd [$m\,s^{-1}$] | 0.11 | 0.39 | 2.3 | -0.3 | 2.7 | 0.99 | 547 |
| 33 | LLWv [$m\,s^{-1}$] | 0.10 | 0.28 | 0.1 | 1.5 | 3.2 | 1.48 | 547 |
| 34 | Rel_Hel [$m^2\,s^{-2}$] | 0.10 | 0.37 | 41 | -9 | 84 | 0.96 | 548 |
| 35 | Hel [$J\,kg^{-1}$] | 0.09 | 0.37 | 17 | 4 | 75 | 1.09 | 547 |
| 36 | \|VFlux\| [$kg\,s^{-1}\,m^{-2}$] | 0.09 | 0.29 | 0.007 | 0.002 | 0.015 | 1.11 | 548 |
| 37 | LLWu [$m\,s^{-1}$] | 0.07 | 0.63 | 1.3 | -0.8 | 3.0 | 0.96 | 547 |
| 38 | Shear3 [$s^{-1}$] | 0.06 | 0.40 | 0.01 | -0.01 | 0.01 | 2.28 | 543 |
| 39 | BRI [] | 0.03 | 0.16 | 16 | -2 | 97 | 0.88 | 539 |
| 40 | LFC [m] | 0.003 | 0.03 | 2000 | 1525 | 2221 | 2.50 | 113 |

Table 4: $R^2$ values for different lead times for all the indices with mean($R^2$) > 0.4. $R^2 \geq 0.5$ are highlighted in bold.

| **Index** | **+24** | **+30** | **+36** | **+42** | **+48** | **+54** | **+60** | **+66** | **+72** | **+78** | **+84** | **+96** | **+108** | **+120** | **+132** | **Mean** | **St. Dev.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 12 | 18 | 00 | 06 | 12 | 18 | 00 | 06 | 12 | 18 | 00 | 12 | 00 | 12 | 00 | | |
| Thetae [K] | **0.97** | **0.94** | **0.96** | **0.96** | **0.96** | **0.92** | **0.96** | **0.95** | **0.95** | **0.91** | **0.94** | **0.94** | **0.93** | **0.91** | **0.88** | **0.94** | 0.02 |
| WBZ [m] | **0.95** | **0.91** | **0.93** | **0.94** | **0.95** | **0.91** | **0.93** | **0.93** | **0.93** | **0.88** | **0.9** | **0.91** | **0.89** | **0.89** | **0.87** | **0.92** | 0.03 |
| PWE [mm] | **0.95** | **0.89** | **0.94** | **0.93** | **0.93** | **0.86** | **0.92** | **0.90** | **0.90** | **0.84** | **0.89** | **0.86** | **0.85** | **0.81** | **0.79** | **0.88** | 0.05 |
| Mix [$g\,kg^{-1}$] | **0.88** | **0.80** | **0.85** | **0.85** | **0.86** | **0.78** | **0.83** | **0.83** | **0.84** | **0.76** | **0.80** | **0.83** | **0.80** | **0.78** | **0.74** | **0.82** | 0.04 |
| MLWu [$m\,s^{-1}$] | **0.86** | **0.78** | **0.84** | **0.83** | **0.84** | **0.74** | **0.77** | **0.75** | **0.76** | **0.65** | **0.72** | **0.65** | **0.60** | 0.49 | 0.44 | **0.71** | 0.13 |
| KI [C] | **0.85** | **0.80** | **0.82** | **0.80** | **0.79** | **0.69** | **0.78** | **0.71** | **0.69** | **0.63** | **0.62** | **0.64** | **0.60** | 0.48 | 0.39 | **0.69** | 0.13 |
| MEL [m] | **0.84** | **0.87** | **0.76** | **0.72** | **0.82** | **0.83** | **0.78** | **0.70** | **0.77** | **0.80** | **0.75** | **0.76** | **0.73** | **0.72** | **0.63** | **0.77** | **0.06** |
| DT500 [C] | **0.83** | **0.77** | **0.81** | **0.77** | **0.78** | **0.69** | **0.74** | **0.71** | **0.72** | **0.63** | **0.69** | **0.67** | **0.64** | **0.60** | **0.54** | **0.71** | 0.08 |
| Tbase [C] | **0.82** | **0.76** | **0.80** | **0.79** | **0.78** | **0.72** | **0.78** | **0.75** | **0.75** | **0.69** | **0.75** | **0.75** | **0.73** | **0.71** | **0.69** | **0.75** | 0.04 |
| MRH [%] | **0.81** | **0.67** | **0.77** | **0.71** | **0.72** | **0.61** | **0.68** | **0.56** | **0.57** | **0.52** | **0.56** | **0.50** | 0.42 | 0.38 | 0.25 | **0.58** | 0.15 |
| MLWv [$m\,s^{-1}$] | **0.76** | **0.65** | **0.73** | **0.73** | **0.71** | **0.66** | **0.69** | **0.69** | **0.62** | **0.62** | **0.56** | **0.52** | 0.47 | 0.40 | 0.33 | **0.61** | 0.13 |
| MaxBuo [K] | **0.75** | **0.71** | **0.67** | **0.62** | **0.72** | **0.65** | **0.60** | **0.59** | **0.69** | **0.63** | **0.58** | **0.64** | **0.50** | **0.53** | 0.39 | **0.62** | 0.09 |
| LRH [%] | **0.75** | **0.51** | **0.74** | **0.71** | **0.68** | **0.50** | **0.69** | **0.61** | **0.58** | 0.45 | **0.56** | **0.53** | **0.50** | 0.45 | 0.34 | **0.57** | 0.12 |
| LI [C] | **0.72** | **0.62** | **0.68** | **0.66** | **0.66** | **0.52** | **0.62** | **0.60** | **0.59** | 0.45 | **0.55** | **0.53** | **0.51** | 0.46 | 0.38 | **0.57** | 0.1 |
| DTC [C] | **0.72** | **0.72** | **0.63** | **0.60** | **0.66** | **0.65** | **0.54** | 0.48 | **0.60** | **0.61** | **0.58** | **0.58** | **0.50** | 0.47 | 0.35 | **0.58** | 0.1 |
| MLWspd [$m\,s^{-1}$] | **0.69** | 0.49 | **0.56** | **0.53** | **0.59** | 0.42 | 0.49 | 0.45 | 0.48 | 0.38 | 0.37 | 0.36 | 0.33 | 0.28 | 0.19 | 0.44 | 0.13 |
| b_PBL [$cm\,s^{-2}$] | **0.68** | **0.53** | **0.60** | **0.64** | **0.65** | **0.50** | **0.55** | **0.58** | **0.60** | 0.43 | **0.50** | **0.56** | 0.49 | 0.48 | 0.45 | **0.55** | 0.07 |
| SWISS [] | **0.67** | **0.54** | **0.65** | **0.64** | **0.61** | 0.36 | **0.59** | **0.52** | 0.48 | 0.29 | 0.49 | 0.45 | 0.46 | 0.38 | 0.32 | **0.50** | 0.12 |
| PWC [mm] | **0.65** | **0.58** | 0.48 | 0.48 | **0.67** | **0.50** | 0.44 | 0.45 | **0.58** | 0.47 | 0.42 | **0.54** | 0.42 | 0.45 | 0.32 | **0.50** | 0.09 |
| VFlux [$kg\,s^{-1}\,m^{-2}$] | **0.64** | **0.53** | **0.56** | **0.60** | **0.57** | 0.43 | **0.54** | **0.51** | **0.51** | 0.36 | 0.42 | 0.36 | 0.32 | 0.21 | 0.17 | 0.45 | 0.14 |
| UpDr [$m\,s^{-1}$] | **0.58** | **0.52** | 0.43 | 0.40 | **0.59** | 0.47 | 0.38 | 0.35 | **0.54** | 0.42 | 0.33 | 0.49 | 0.30 | 0.39 | 0.25 | 0.43 | 0.1 |
| CAPE [$J\,kg^{-1}$] | **0.58** | 0.42 | 0.43 | 0.39 | **0.57** | 0.43 | 0.36 | 0.31 | **0.52** | 0.35 | 0.28 | **0.50** | 0.33 | 0.36 | 0.03 | 0.41 | 0.1 |
| SWEAT [] | **0.58** | 0.44 | **0.58** | **0.57** | **0.50** | 0.38 | **0.52** | **0.54** | 0.44 | 0.33 | 0.45 | 0.42 | 0.38 | 0.34 | 0.35 | 0.46 | 0.09 |

# FIGURES

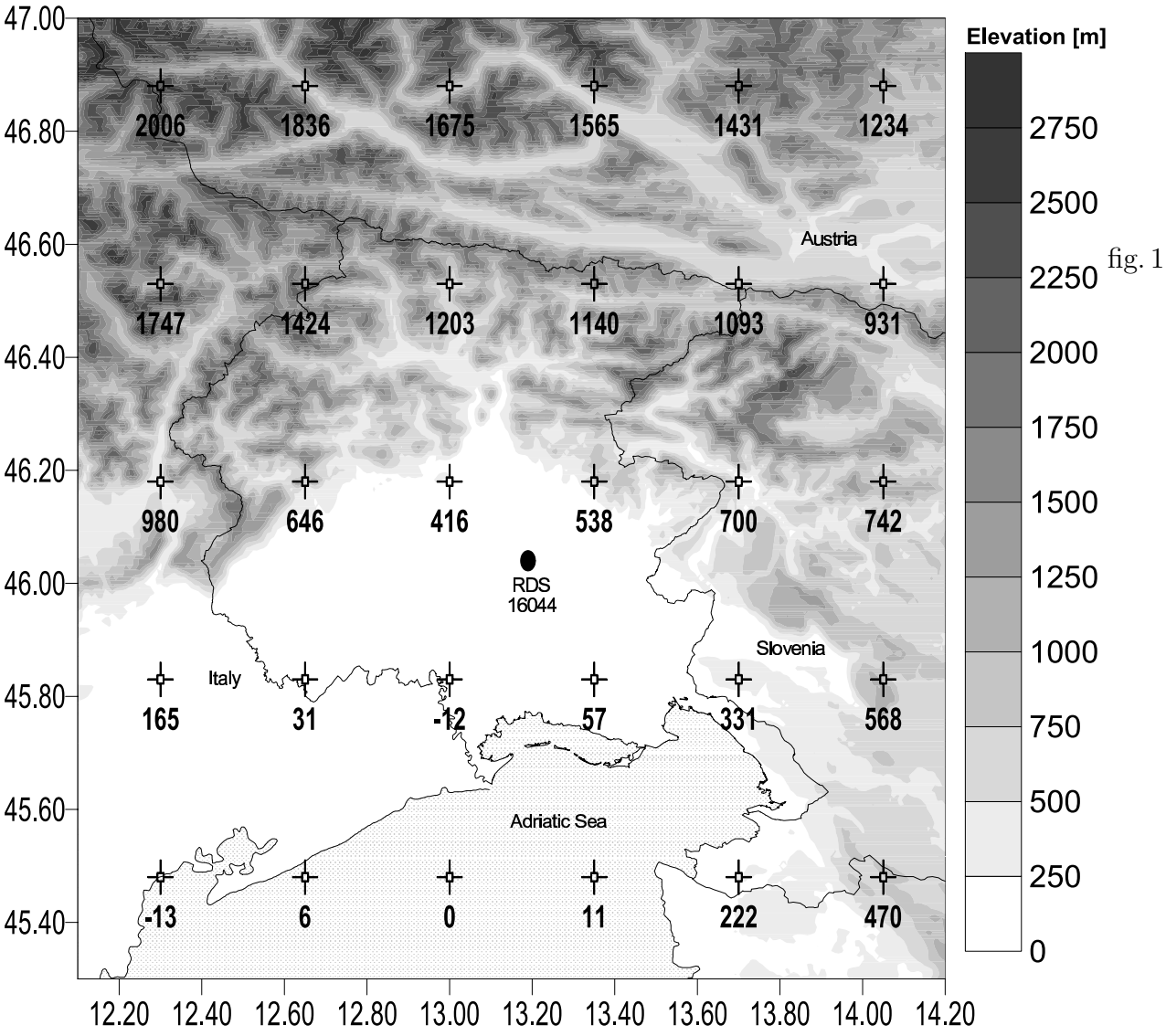


fig. 1

fig. 1: *The Friuli Venezia Giulia region and its orography (in filled contours) with the Udine–Campoformido (WMO code 16044) sounding base in the middle of the plain and the ECMWF T511 L60 surface elevation gridded values (as kindly provided by Dominique Lucas, ECMWF). Courtesy of Andrea Cicogna (OSMER–ARPA FVG).*

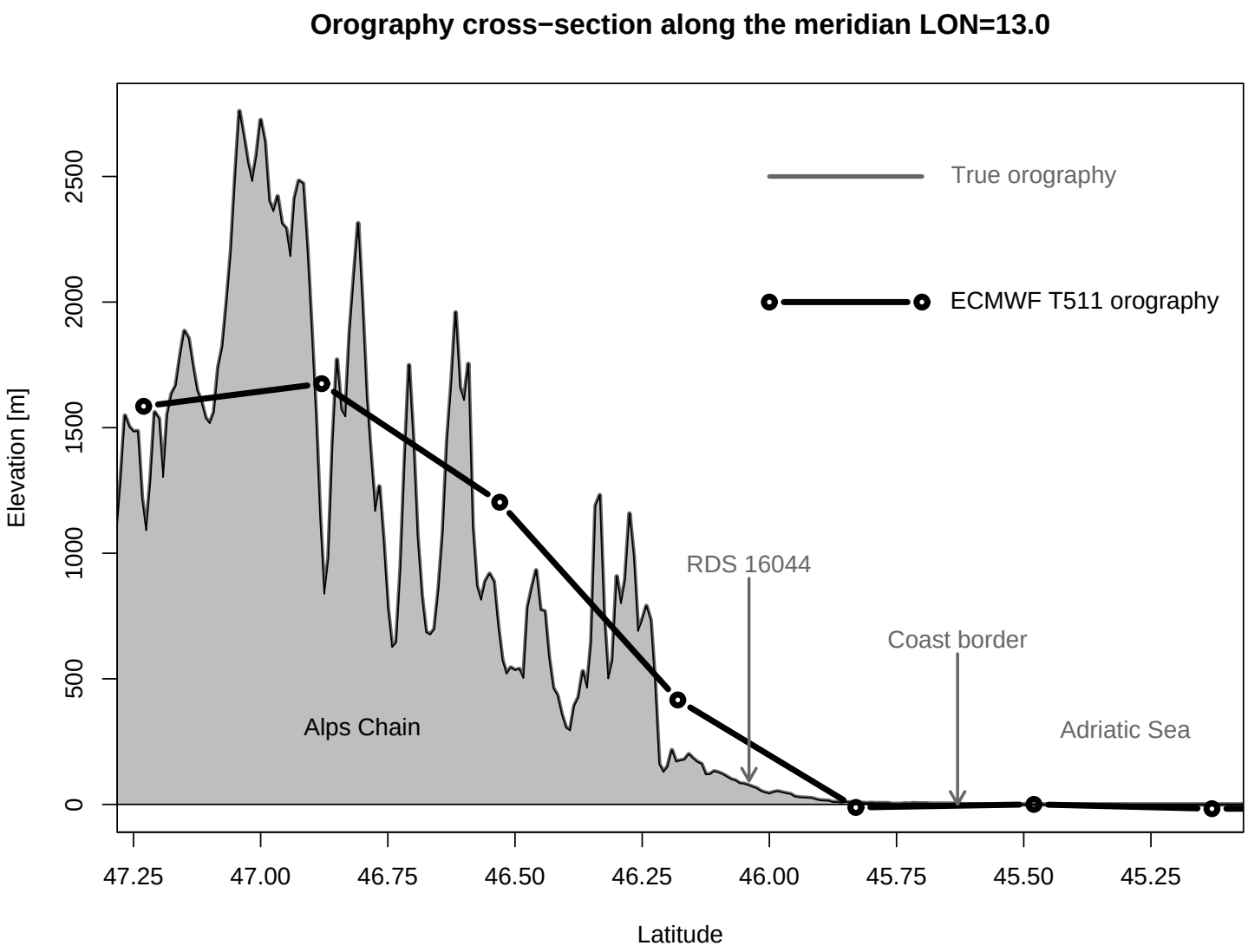


fig. 2

fig. 2: *The vertical cross–section of the Friuli Venezia Giulia region along the LON=13.0 deg meridian (in shaded in gray) with the points representing the ECMWF T511 L60 surface elevation (connected by a black line). It is possible to see how the very complex terrain gradient is smoothed in the model orography.*

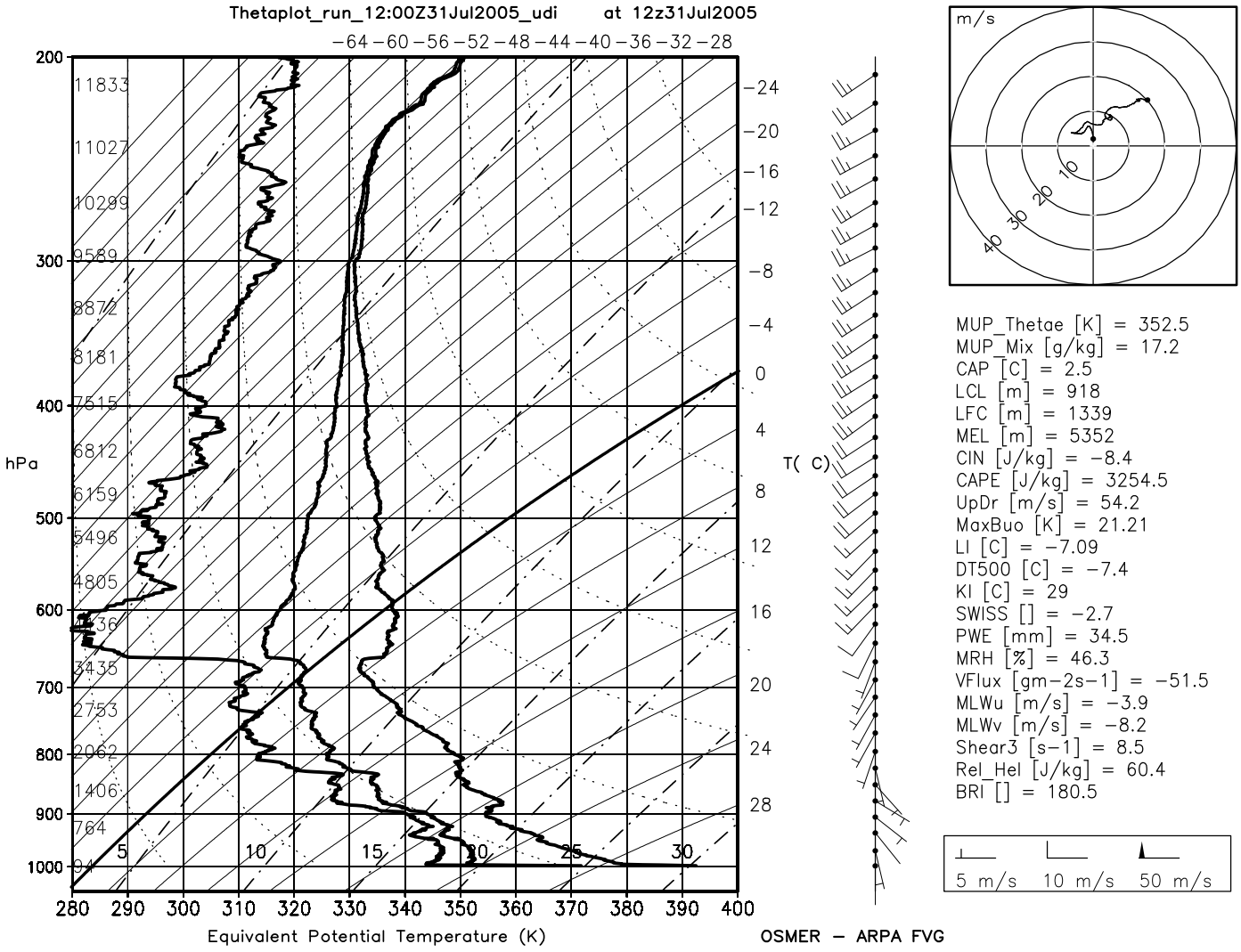


fig. 3 a: *Thetaplot and hodograph of the Udine–Campoformido sounding observed on 31 July 2005 at 12 UTC (one measure every two seconds). The three thick lines are (from right to left):* $\Theta_{es}$ *(which intersects the environmental temperature on the isotherms),* $\Theta_e$ *(which intersects the wet bulb temperature on the isotherms) and* $\Theta_{ed}$ *(which intersects the dew point on the isotherms). This diagram is very useful to evaluate the potential instability, as explained in Morgan (1992) or Manzato and Morgan (2003). Also a few instability indices computed by the* SOUND_ANALYS.PY *software are displayed. Courtesy of Arturo Pucillo (OSMER–ARPA FVG).*

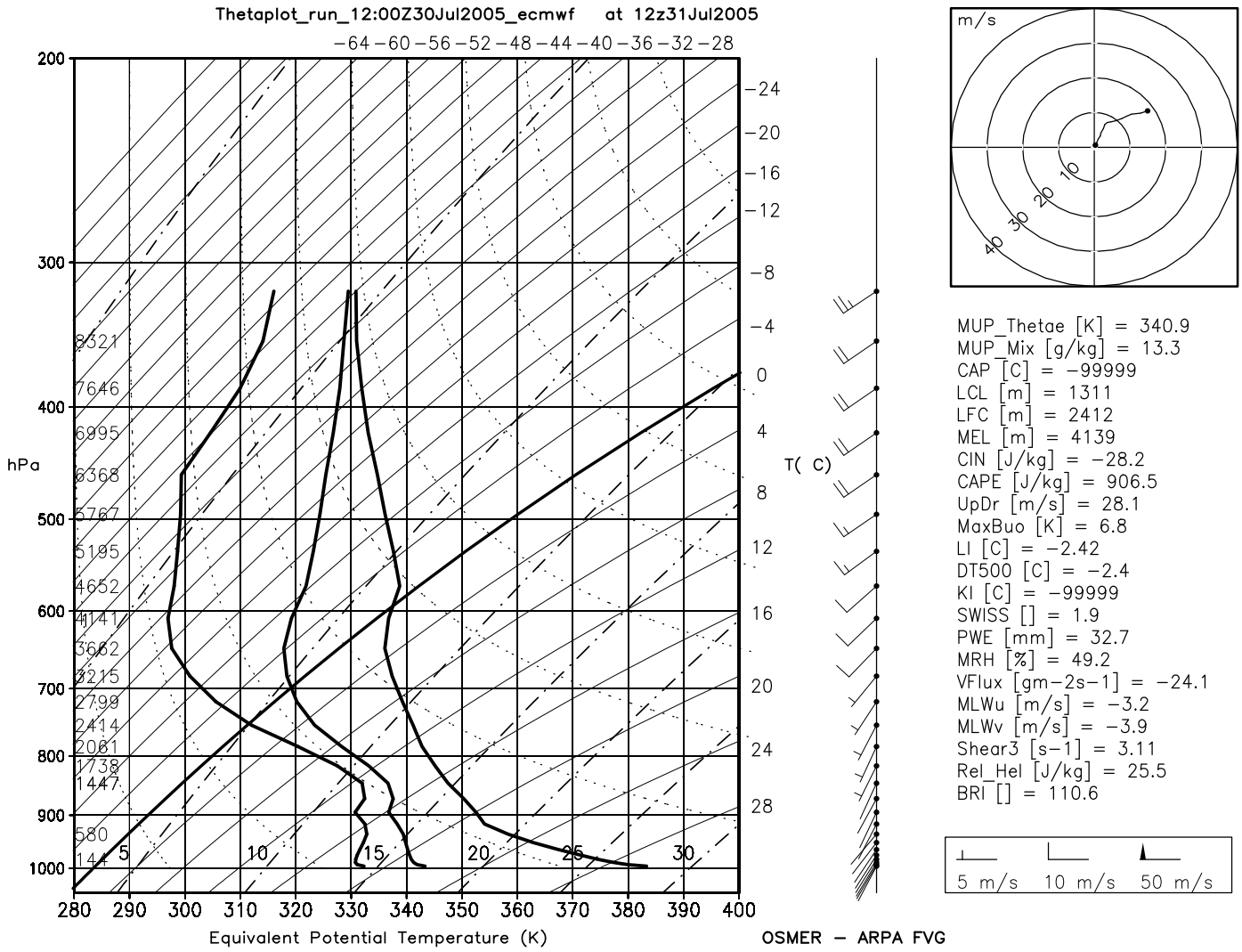


fig. 3 b: *Thetaplot diagram of the vertical profile forecasted by* ECMWF *run of 30 July 2005 at 12 UTC for the +24 h, that is the same time as in figure 1 a. There are only thirty two levels from surface up to 300 hPa. Courtesy of Arturo Pucillo (OSMER–ARPA FVG).*

fig. 3 c: *Thetaplot diagram of the 31 July 2005 at 12 UTC sounding "degraded" in the vertical resolution by the TEMP format (FM 35, WMO manual on codes number 306), downloaded in text format from the University of Wyoming upper air database. This plot has been done using only 29 levels in the lowest 200 hPa. Note that, differently from the indices studied in the text, the indices shown in fig. 1 have been computed using the* **T** *scheme of MM03, that is, without the virtual correction. Courtesy of Arturo Pucillo (OSMER–ARPA FVG).*

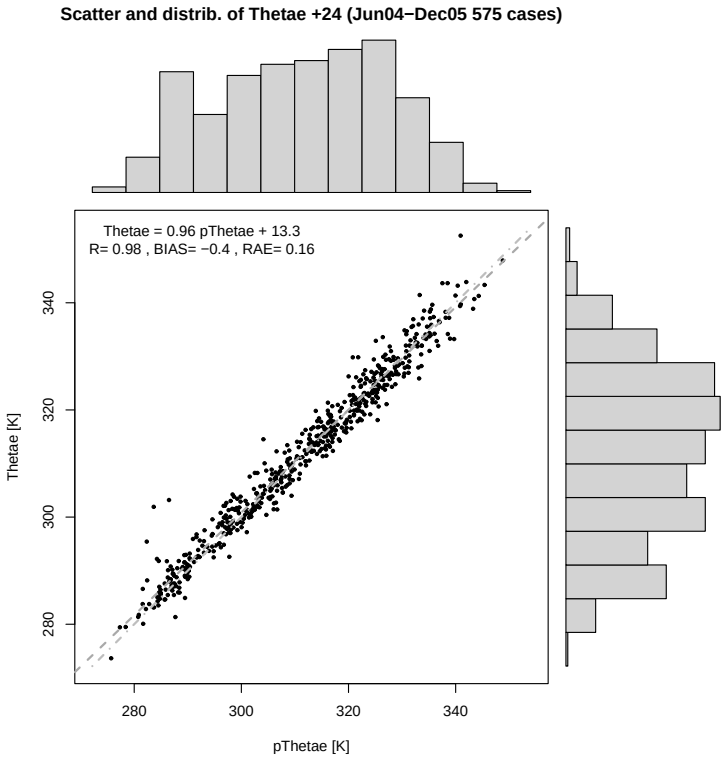


fig. 4a

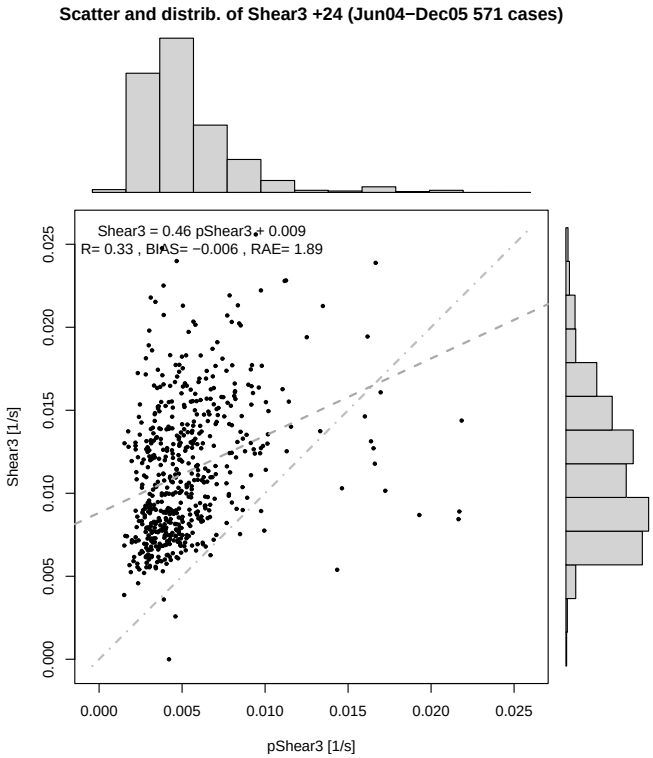


fig. 4b

fig. 4 a: *Scatterplot with histogram distributions and linear fit for the Most Unstable Parcel* $\Theta_e$ *derived by sounding observations at 12 UTC* (Thetae) *and derived by model pseudo–sounding forecast* (pThetae) *made the day before (+24 h forecast). The linear fit coefficients, correlation,* bias *and* RAE *shown are the best obtained by all the forty different indices.*

fig. 4 b: *Same as in 4 a but for the index with the worst correlation between observed and forecast values, that is the wind shear in the lowest 3 km.*

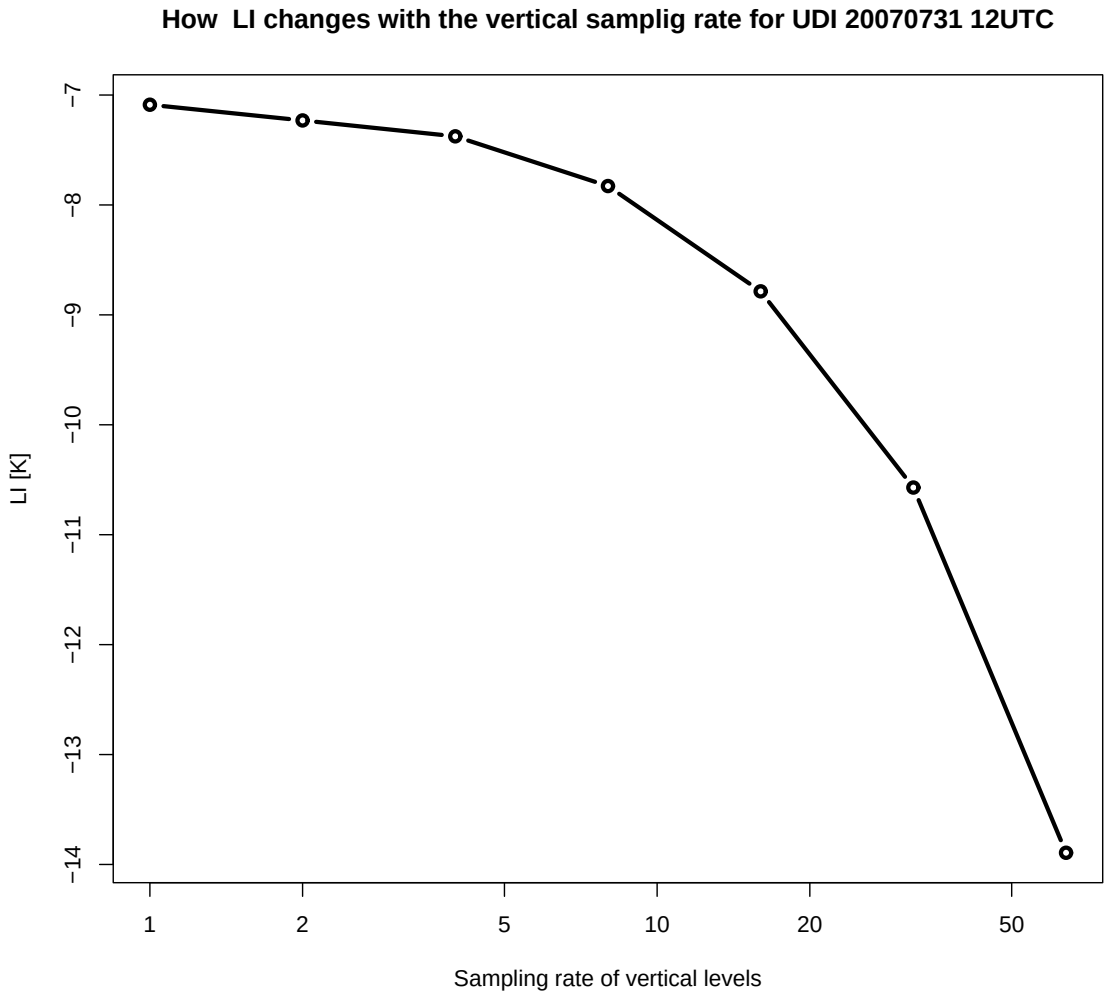


fig. 5

fig. 5: *How the lifted index changes when the original high resolution data (one level every two seconds) are degraded at different sampling rate, from 1 to 64. Sampling rate 1 means to use all the original levels while 64 means to retain one original level every 64 (98.4% of data are discarded). The* LI *error is about* 1 K *when using one level every ten, which means to have about 100 levels below 300 hPa. The* LI *derived by the* TEMP *format is* -9.7 K.

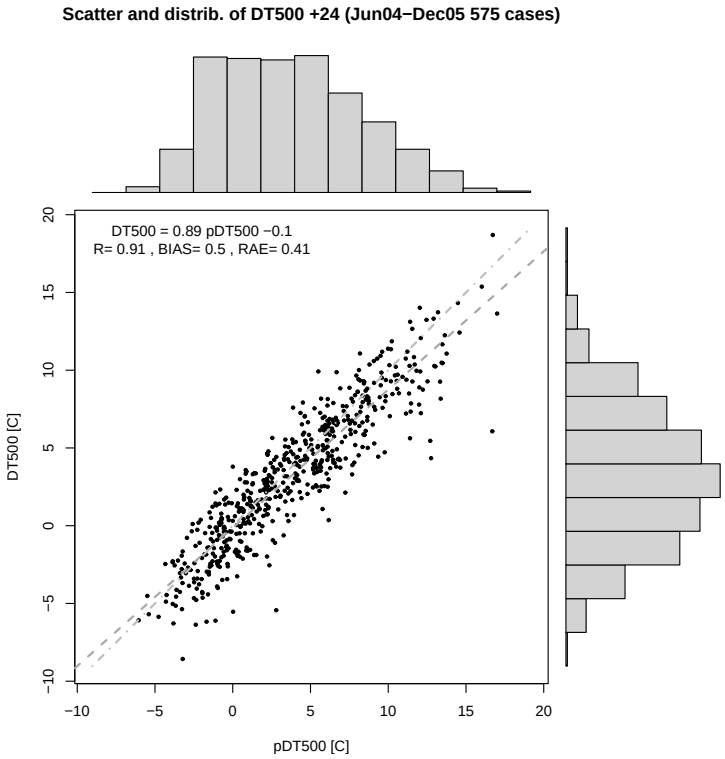


fig. 6a

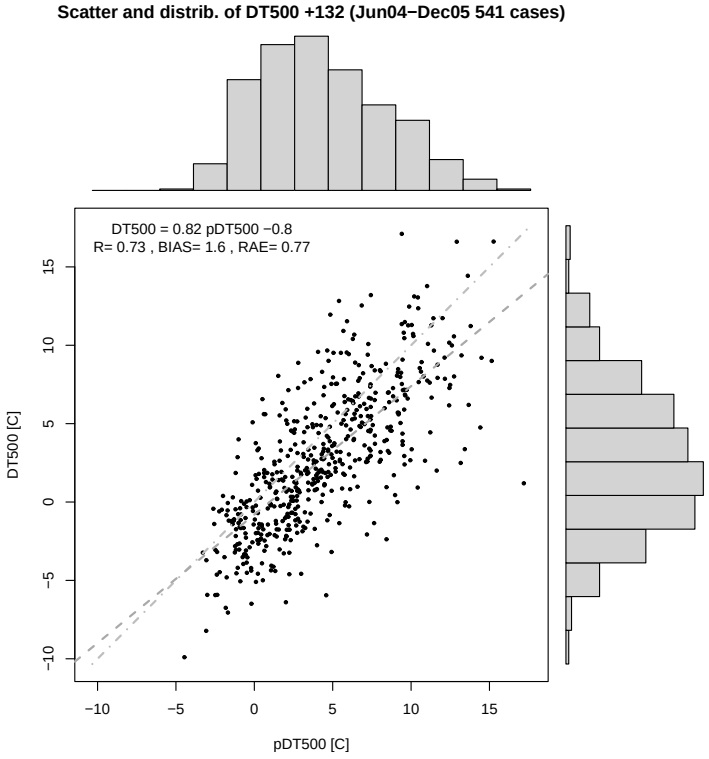


fig. 6b

fig. 6 a: *Same as before but for the temperature difference between the lifted parcel and the environment at 500 hPa* (DT500) *at +24 h.*
fig. 6 b: *Same as figure 6 a but at +132 h (00 UTC of the sixth day). As usually during night the instability is strongly underestimated* (bias = 1.6 °C)*, anyway the correlation is still acceptable* ($R^2 = 0.56$).

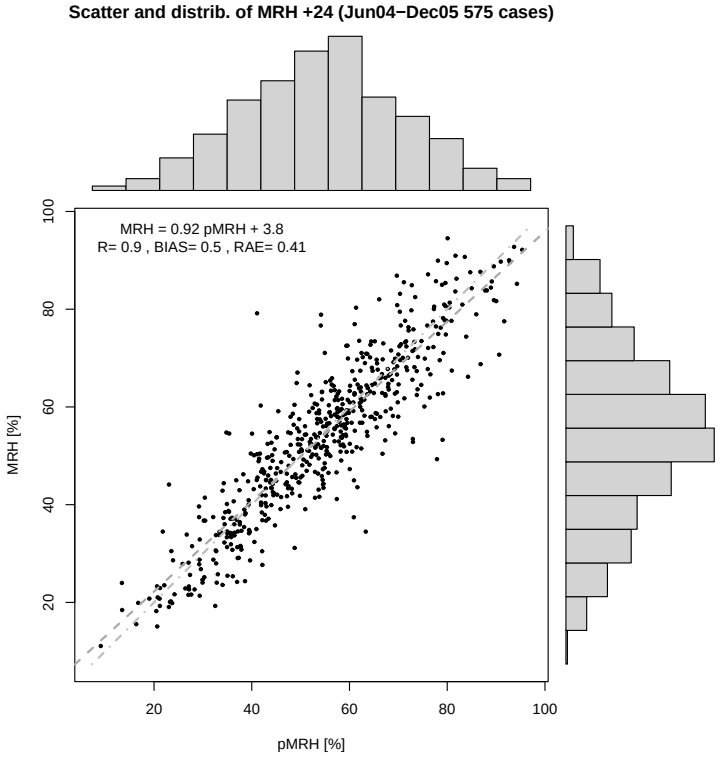


fig. 7a

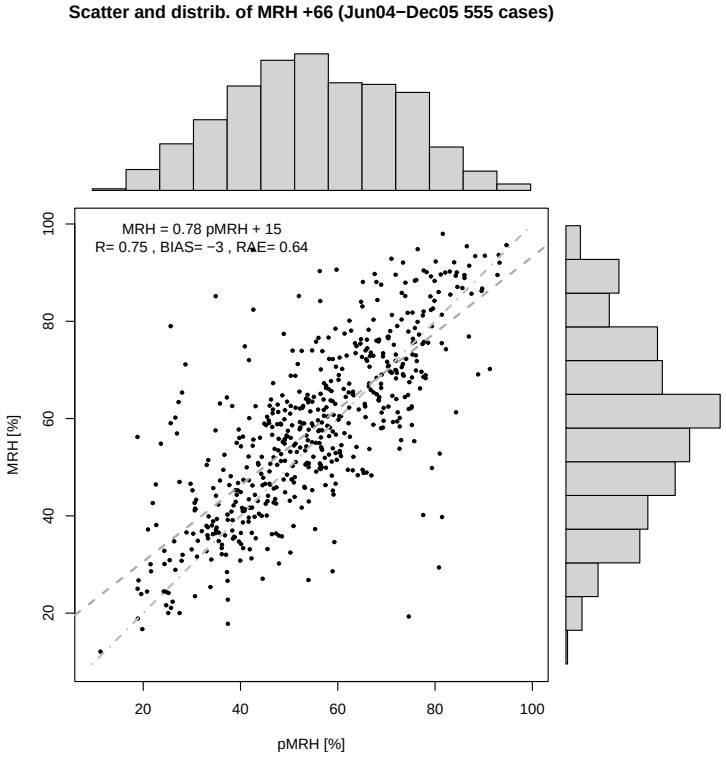


fig. 7b

fig. 7 a: *Same as previous figures, but for the mean relative humidity in the lowest 500 hPa* (MRH) *at +24 h. The correlation coefficients (*$a \cong 1$ *and* $b \cong 0$*) show a good calibration, but the high* RAE *show also a high dispersion.*
fig. 7 b: *Same as 7 a, but at +66 h (06 UTC of the third day): in only 42 h the correlation and dispersion worsened a lot.*

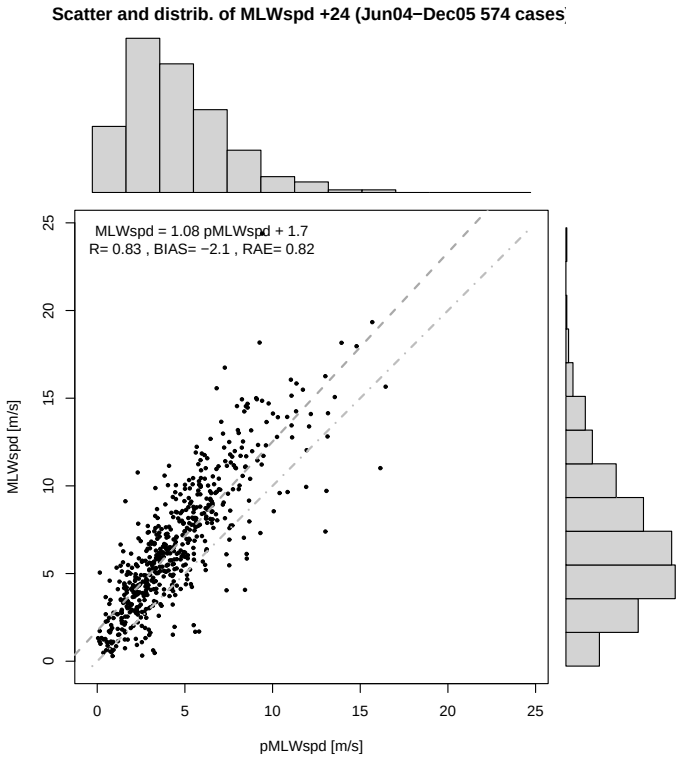


fig. 8a

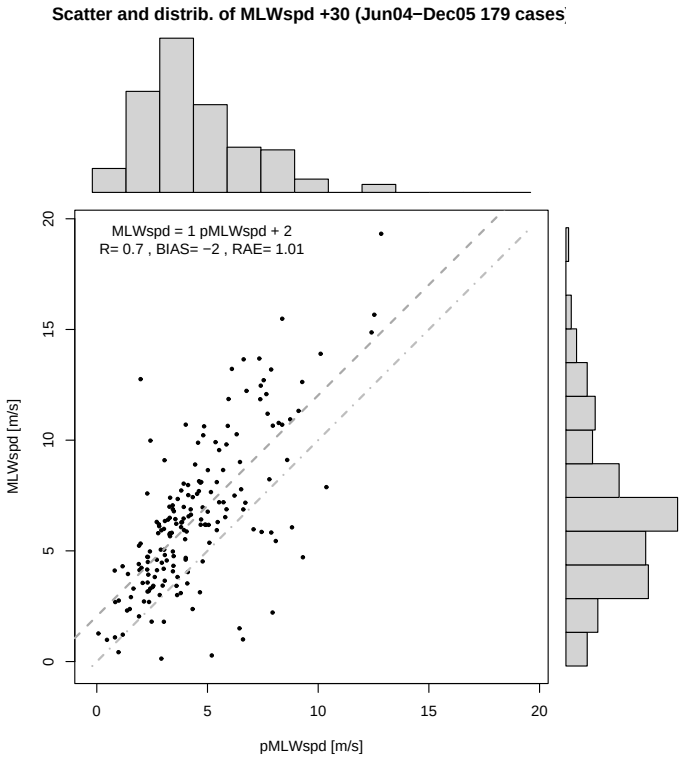


fig. 8b

fig. 8 a: *Same as before but for the mean wind speed in the lowest 6 km* (MLWspd) *at +24 h. Note the forecast understimation* (bias = -2 $\mathrm{m\,s^{-1}}$).
fig. 8 b: *After only 6 h (18 UTC of the first day) the correlation is much worsened. Note that at 18 UTC there are many less soundings during the studied period.*

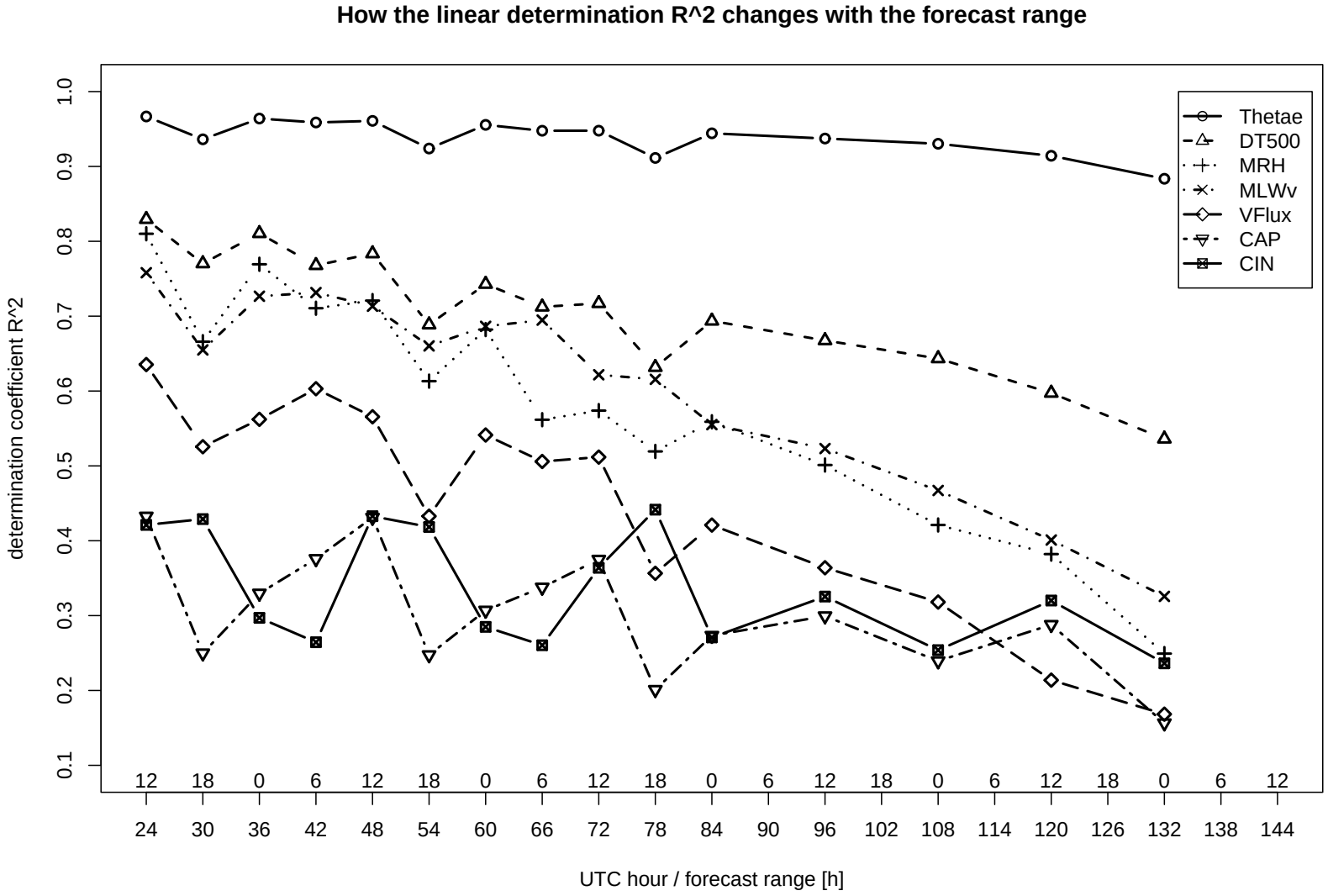


fig. 9

fig. 9: *The time series of the linear determination coefficient* ($R^2$) *for seven indices, for all the time forecasts available at* OSMER–ARPA FVG. Thetae *and* DT500 *conserve high correlations during all the forecasts, with a local minimum at 18 UTC.* MRH *and* MLWv *have acceptable correlations up to +96 h and +84 h respectively.* VFlux *has good* R *up to +72 h, but with minimums at 18 UTC. Note finally the strange behavior of* CIN *which has a local maximum at 18 UTC.*

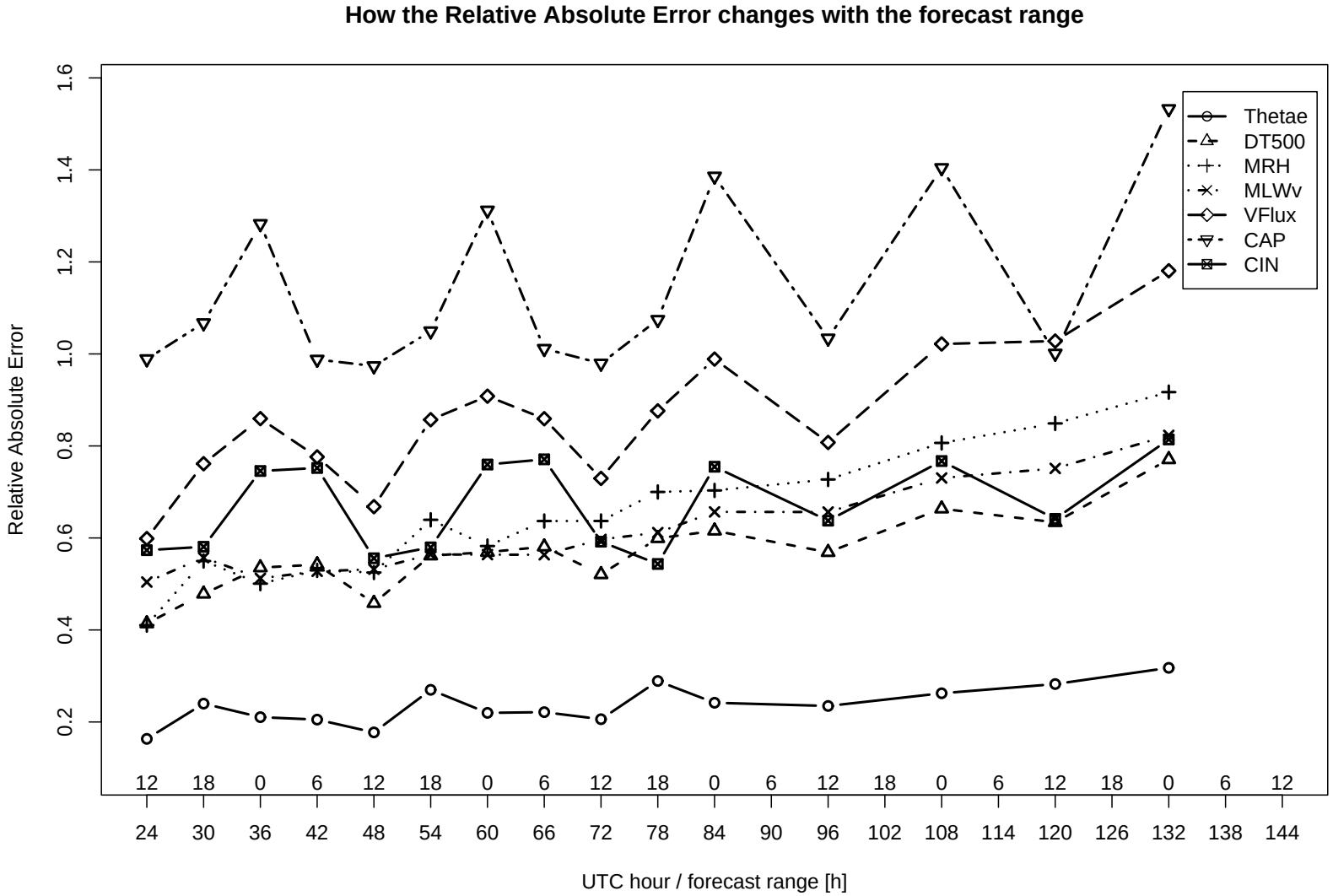


fig. 10

fig. 10: *Same as the previous figure but for the* RAE *measure. While* Thetae *still outperforms the others,* DT500 *is here more similar to* MLWv, CAP *(but not at 00 and 06 UTC) and* MRH. VFlux *has a local minimum (low error) at 12 UTC, while* CAP *has very large errors at 00 UTC, because the night inversions are strongly overestimated* (bias *always larger than* 3.6 °C).