# *MEDIUM RANGE HEAT HEALTH FORECASTING*

Paul Becker*, Christina Koppe
Deutscher Wetterdienst, Freiburg, Germany

## 1. INTRODUCTION

Heat Health Warning Systems (HHWS) with a forecast horizon of up to two days have been established in several countries during the last decade. In addition, it seems useful to complement an HHWS by advance heat information that contains a corresponding forecast with lead times of up to 10 days (medium range weather forecasts). This early heat information is useful for the health sector for preparation purposes. Unfortunately, the uncertainty of a forecast increases with its lead time. Probabilistic forecasts can be used to estimate this uncertainty of medium range weather forecasts. A study by Lalaurette (2003) indicates that it was possible to predict one of the most severe heat waves Europe has ever experienced up to 5 days before the onset of the heat wave with the European Centre for Medium Range Weather Forecasts (ECMWF) ensemble prediction system (EPS) forecast.

Apart from an absolute measure for the thermal environment the person's reaction to a certain level of thermal stress depends on how well a human being is adapted to the local meteorological conditions. There are two kinds of adaptation: long-term and short-term. Long-term adaptation is reflected in the fact that many studies find different temperature thresholds upon which negative health effects (in general an increase in mortality) could be detected in different climatic regions (e.g. Keatinge et al., 2000, Hajat et al., 2002). That humans can also adapt in a shorter time range (1 month) is shown by several studies analysing the thermal-environment-mortality-relationship. These studies show a stronger effect of a given level of thermal load earlier in the summer than later in the summer (e.g. Basu and Samet, 2002, Kyselý, 2004). Koppe (2005) was able to show that this behaviour was principally due to short-term adaptation to the local meteorological conditions of the past month using the HeRATE procedure (**He**alth **R**elated **A**ssessment of the **T**hermal **E**nvironment). Therefore, the threshold upon which a thermal situation could be defined as a heat wave depends on the location and on the previous weather conditions.

Therefore, the aim of this study is to develop a method using the HeRATE procedure in combination with an ensemble prediction system which is able to give a hint at the possibility of a heat wave occurring within the next ten days. Probability forecasts were computed based on the ECMWF ensemble prediction system.

A heat health warning requires a warning indicator that somehow is related to negative impacts on health. Quite a broad range of methods is used to identify this warning indicator. Simple methods are merely based on air temperature or a combination of air temperature and humidity. Complex methods are based on heat budget models of the human body or synoptic approaches. In general it could be stated that the more complex a method for the heat warning and the more meteorological parameters that influence the heat budget of a human being (e.g. air temperature, radiant temperature, humidity, wind speed) are taken into account, the better the thermal-environment-mortality-relationship can be modelled. Unfortunately, the uncertainty for forecasting such indexes is greater also than that arises from forecasting just one parameter. The issue of the uncertainty of a forecast becomes even greater with increasing forecast horizon.

In order to provide early information about the probability of the occurrence of a heat wave in the medium range (up to 10 days) we tested several strategies. The first was to use only the 850-hPa temperature as indicator for an imminent heat wave. The second was to use the 2m temperature. The skills of these forecasts were tested using the Brier Score and the Brier Skill Score. In order to identify situations with a potential danger for human health, adaptation to the local weather situation of the previous months was taken into account using the HeRATE approach (Koppe, 2005).

* *Corresponding author address:* Paul Becker, Deutscher Wetterdienst, Medizin-Meteorologie, Stefan-Meier-Str. 4, D-79104 Freiburg, Germany;
e-mail: paul.becker@dwd.de

The strategies mentioned are used to simulate the 2003 heat wave in western Europe. This heat wave caused more than 35 000 deaths all over Europe and was one of the most severe heat waves western Europe has ever experienced.

## 2. MATERIALS AND METHODS

### 2.1. Data

In order to estimate the probability of a heat wave during the following 10 days the ECMWF EPS was used. This EPS consists of 50 ensemble members and one control forecast for each lead time. For the period between 20.05.2003 to 31.08.2003 daily ECMWF ensemble forecasts for three lead times (L060: 60 h = 2.5 days, L132: 132 h = 5.5 days and L204: 204 h = 8.5 days) as well as the respective analyses for the 850 hPa temperature (P130) and the 2m temperature (P167) were provided by the ECMWF (European Centre for Medium Range Weather Forecast). The data were available in a resolution of one degree for the area between longitudes extending from –15° and +45° and latitudes from 80° to 30°.

### 2.2. Verification of probabilistic forecasts

In contrast to a deterministic forecast a probabilistic forecast can never be right or wrong, except when probabilities of 100% or 0% have been stated. Therefore the performance of EPS can only be evaluated from large samples of forecasts (Persson & Grazzini, 2005).

The Brier Score (BS) is the most common verification method for probabilistic forecasts and a measure of the mean square error. It measures the difference between a forecast probability of an event (p) and its occurrence (o), expressed as 0 or 1 depending on wether the event has occurred or not (Persson & Grazzini, 2005):

$$BS = \frac{1}{N} \sum_{n=1}^{N} (p-o)^2 .$$

The Brier Score is bounded by the values 0.0 and 1.0; a lower value represents a better forecast (Legg and Mylne, 2004).

Comparing Brier Scores for different events can be misleading if their climatological probabilities are different. Therefore, the Brier Skill Score (BSS) is often calculated. The Brier Skill Score compares the Brier Score of the forecasting system ($BS_{fc}$) with that obtained by a reference forecast ($BS_{ref}$):

$$BSS = 1 - \frac{BS_{fc}}{BS_{ref}} .$$

Typical reference forecasts used are climatology or persistence. For extreme events, no prior climatological probabilities are available, and since they are rare events, Legg and Mylne (2004) suggest using a null forecast (always forecast the probability to be zero). This shows whether the forecasts are better than the option of never issuing a warning. The BSS measures the reduction in forecast error variance of the forecast compared to the reference forecast.

Brier Score and Brier Skill Score are always calculated for a large sample of forecasts. In this study we determined the BS and BSS variation in time and in space.

### 2.3. Thresholds for warm events

In order to be able to determine the probability of an event, the has first event to be defined. This raises the question of how to define a heat wave or a threshold upon which heat load can be expected. In Koppe (2005) and Koppe et al. (2004) this problem is discussed in detail. An important aspect in this context is that the threshold upon which negative impacts on human health are expected varies within time and space due to the fact that humans adapt to the local meteorological situation of the last 30 days. To solve this problem the HeRATE method was developed (Koppe, 2005; Koppe & Jendritzky, 2005). As HeRATE consists of an absolute part and a relative part which takes short term adaptation into account, the thresholds calculated with this method vary within time and space.

In its original version HeRATE bases upon the Perceived Temperature (PT) an assessment procedure for the thermal environment that takes all relevant mechanisms of heat exchange into account. The calculation of PT requires several input parameters (2m temperature, mean radiant temperature, wind velocity, humidity), therefore, the forecast uncertainty for PT is higher than just for one simple parameter. Thus, it was decided to use in a first step merely the temperature instead of PT for the assessment of the thermal situation from day +3 on. The absolute threshold of 32°C for the temperature at 2m above ground remained the same as when using PT. For the threshold in 850 hPa we used a dry adiabatic cooling.

Based on the HeRATE procedure thresholds that are variable within space and time for heat-events (heat waves) have been calculated for each day and each grid point, so that the probability of exceeding those thresholds could be determined.

## 3. RESULTS

### 3.1. Absolute Thresholds

Brier Score (BS) and Brier Skill Score (BSS) both depend on the threshold that is used to determine the probability of the event in case the null-forecast is used as a reference forecast for the BSS. The higher this threshold is, the lower also the general probability that it will be exceeded and the better the BS (Figure 1a). This is due to the fact that if the threshold is rarely reached and also forecast, for many data points the forecast that there will be no event is correct.

In contrast to the Brier Score, the Brier Skill Score is the better the lower the threshold is set (Figure 1b). The explanation for this behaviour is that, for the calculation of the BSS, the BS of a reference or zero forecast is compared with that of the ensemble forecast. In turn, the reference forecast is the better, the higher the threshold and the lower the real probability that this threshold will be exceeded.

In addition both BS and BSS show better values for shorter lead times, indicating that the skill of a forecast with a shorter lead time is better than the skill of forecasts with longer lead times.

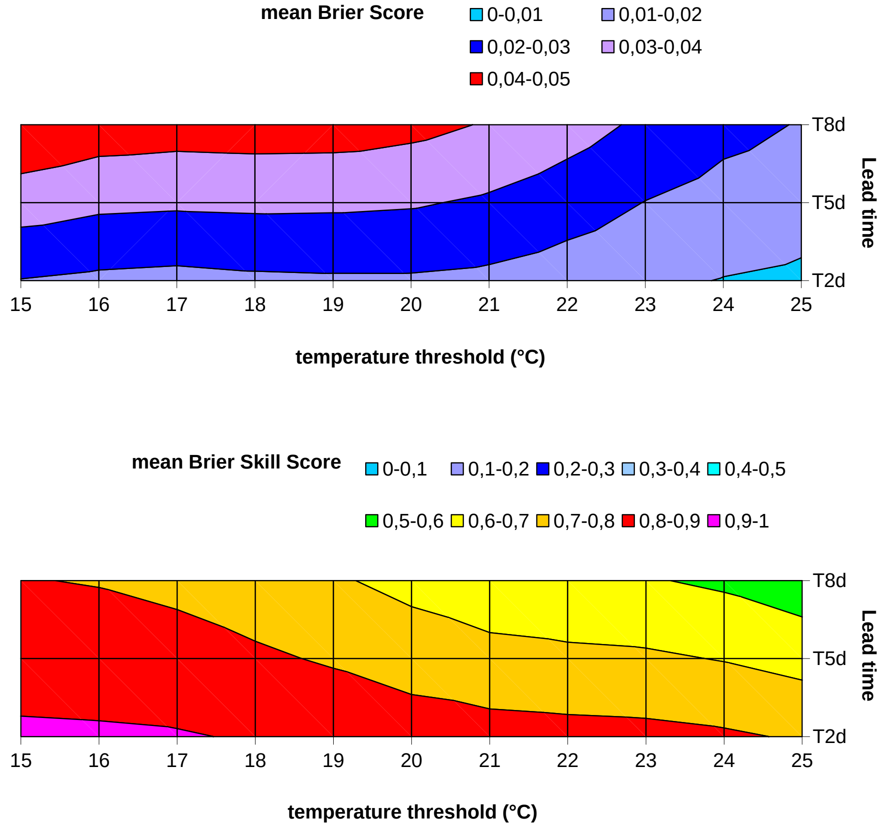



**Figure 1: Dependency of Brier Score (1a: above) and Brier Skill Score (1b: below) on lead time and the threshold that has to be exceeded for the 850 hPa temperature.**

### 3.2. Comparison of PDFs

A comparison of the probability density functions (PDFs) of the analysis and the ensemble mean can provide information about the reliability of the forecasting system. A good reliability means that, for example, if 50 % of the ensemble members predict the exceedance of a certain threshold this threshold is actually exceeded in 50 % of the cases. In Figure 2 (3) the difference between the 50 % (90 %) Percentile of the analysis and the respective percentile of the EPS is shown.

Figure 2 exemplifies the reliability of the EPS for summer 2003 for the two parameters: temperature in 850 hPa and temperature 2m above ground for the three analysed lead times. The yellow colour means that in general the ensemble mean and the 50 % percentile of the analysis correspond quite well. Red colours indicate that the EPS predicted temperatures that were to cold, whereas green and blue colours indicate that the EPS predicted temperatures that were warmer than the analysis.

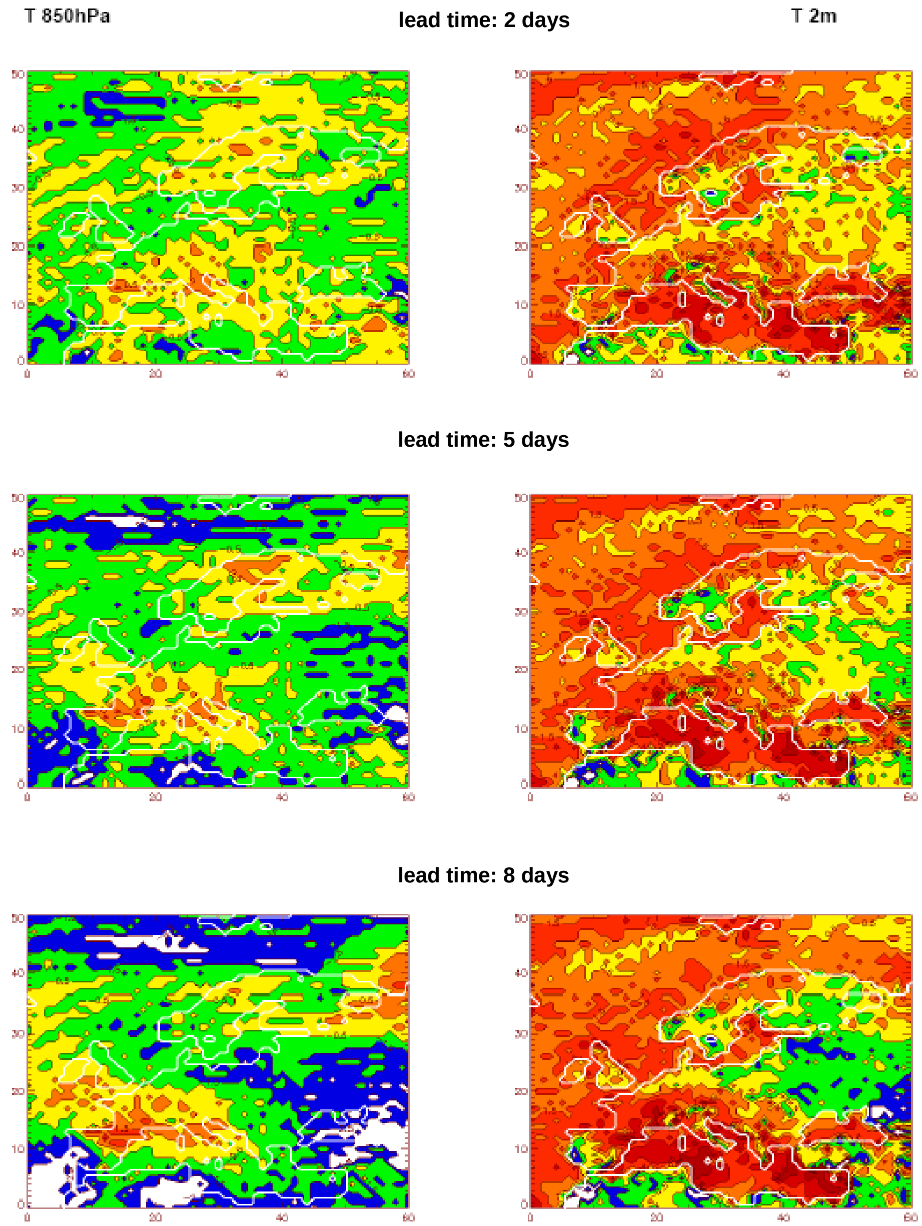



**Figure 2: Differences between the analysed and forecast 50 % Percentile [mean conditions] of temperature in 850 hPa (T 850 hPa, left row) and temperature in 2m (T 2m, right row) for different leadtimes in summer 2003. White: < -2.5 K; blue –2.5 K to –1.5 K; green: -1.5 K to –0.5K; yellow: -0.5 K to 0.5 K; orange: 0.5 K to 1.5 K; bright red: 1.5 k to 2.5 K; dark red: > 2.5 K**

The yellow colour dominates for the 850 hPa temperature and lead times up to 5 days, indicating a relatively good reliability of the EPS during the extremely hot summer of 2003 for mean values. For a lead time of 8 days the ensemble mean underestimates the 50 % percentile of the 850 hPa temperature in southern France and northern Spain. In contrast to the 850 hPa temperature, the 2m temperature estimated by the ensemble mean is generally lower than the analysis for all lead times and large parts of southern, western and central Europe. The differences are most pronounced over the oceans.

Figure 3 shows the 90% percentiles of ensemble mean and analysis and gives a hint of the reliability of the system during hot days. For the lead times up to 5 days the accordance between EPS and the analysis for the 850 hPa temperature is good. For lead times of 5 days and more the underestimation of the 850 hPa and the 2m temperature made by the EPS

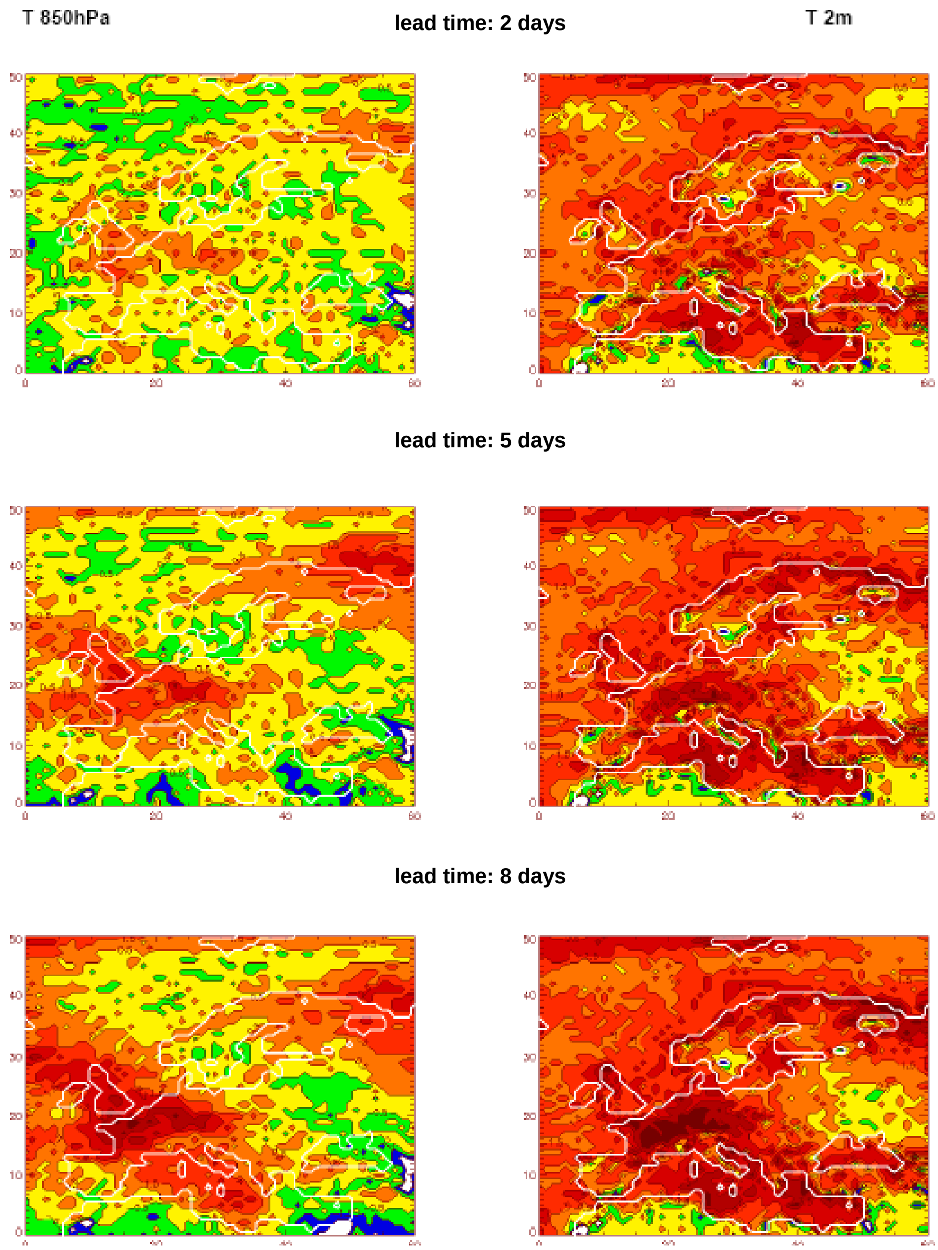



**Figure 3: Differences between the analysed and forecasted 90 % Percentile [extreme conditions] of temperature in 850 hPa (T 850 hPa, left row) and temperature in 2m (T 2m, right row) for different lead times in summer 2003. White: < -2.5 K; blue –2.5 K to –1.5 K; green: -1.5 K to –0.5K; yellow: -0.5 K to 0.5 K; orange: 0.5 K to 1.5 K; bright red: 1.5 k to 2.5 K; dark red: > 2.5 K.**

system in the regions of Europe that were hit by the heat wave becomes even clearer. Again the 850 hPa temperatures matches better with the analysis than the 2m temperature.

### 3.3. Temporal variation

For the following analysis the threshold for a heat wave was set according to the method described in section 2.3. This method of determining the threshold has the advantage that the different regions in Europe can be compared.

With the exception of the BSS of the 2m temperature forecast with a lead time of 204 h on June 23 the values of BSS are positive (figure 4) and the values of BS are lower than 0.2 (not shown). Positive BSS values indicate that the EPS forecast is better than the reference forecast. Hence, the forecasts can be considered as skilful. Nevertheless, it should be borne in mind that the skilfulness of the Brier Score is partly based on the fact that we are looking at the exceedance of extreme values in Europe which is, in our case, displayed by 3111 data points (61 degrees longitude x 51 degrees latitude). Therefore, this skill is partially obtained because the threshold is reached neither by the forecast nor by the analysis over greater parts of Europe.

BS and BSS are both better for the temperature in 850 hPa than for the temperature 2m above the ground (Figure 4a and 4b). So the mean BSS for T850 (T2m) are for the different lead times L060 (2.5 days): 0.88 (0.72); L132 (5.5 days): 0.81 (0.66); L204 (8.5 days): 0.73 (0.57).

The variation of the BSS of 2m temperature in summer 2003 is much stronger than that for the 850 hPa temperature. This variation depends also on the lead time, and is more pronounced for longer lead times. During the

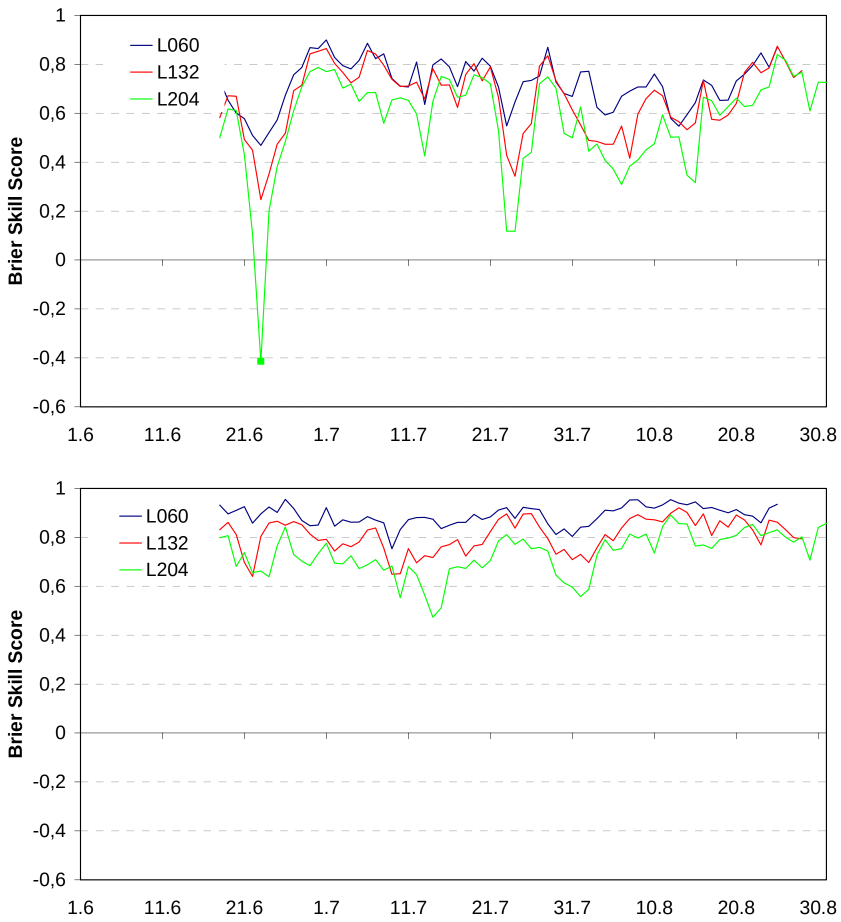



**Figure 4: Brier Skill Score of the temperature 2m above the ground (above, figure 4a) and in 850 hPa (below, figure 4b) for Europe in summer 2003.**

strongest heat event (first week of August), the BSS of both parameters and all lead times show slightly lower values than during the preceding and following days. However, there seems to be no significant temporal correlation between the BSS of the 2m temperature and the BSS of the 850 hPa temperature.

### 3.4. Spatial variation

We also analysed the spatial variation of BS and BSS in summer 2003 (Figure 5 and Figure 6). The aim of the analysis was to see if spatial patterns of BS and BSS exist and if these patterns can be linked to the heat wave in August 2003. BS shows better (lower) values in the northern and southern latitudes of Europe as well as in the UK and southern Scandinavia. In contrast the BSS shows a south–north gradient in its performance with better values in northern Africa and the southern Mediterranean region. Again the values for the UK and southern Scandinavia are quite good.

In general, the skill of the forecast is better for lead times up to 5 days than for longer lead times. However the skill of the forecasts for southern Scandinavia and northern Africa is very good also for lead times of 204 h (8.5 days). In addition, there is a weak hint that the skill for lead times of more than 5 days is slightly worse in those regions that suffered most from heat in summer 2003.

### 3.5. Performance

Based on the values of BS and BSS, it could be assumed that the forecasts of the EPS are quite skilful and even useful for medium-range heat information. Looking at the performance of the ensemble probabilities in August 2003, when Europe was hit by a severe heat wave, it seems that only the forecast up to 132 h (5.5 days) had the skill to show the signs of the imminent heat wave (figure 7). In Figure 7 the orange line

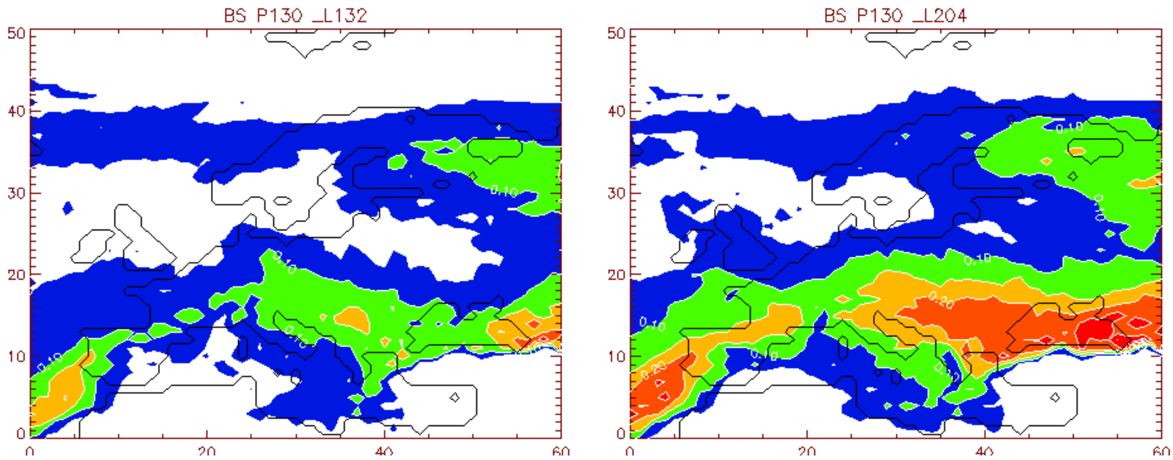



**Figure 5 : Brier Score for 850 hPa temperature for the period from 19.06.2003 – 31.08.2003 for two lead times (left: 5 days, right: 8 days). White: 0.00 to 0.05, blue: 0.05 to 0.10, green: 0.10 to 0.15, orange: 0.15 to 0.20, red: > 0.20.**

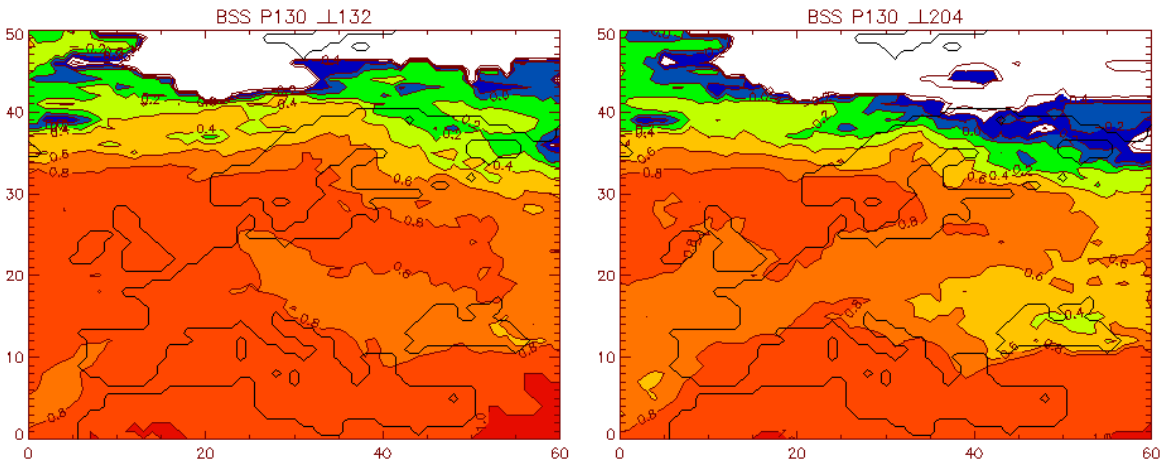



**Figure 6 : Brier Skill Score for the period from 19.06.2003 – 31.08.2003 for two lead times. White: < -0.4, blue: -0.4 to 0.0, green: 0.0 to 0.2, bright green: 0.2 to 0.4, orange: 0.4 to 0.6, bright red: 0.6 to 0.8, red: 0.8 to 1.0, dark red: 1.0**

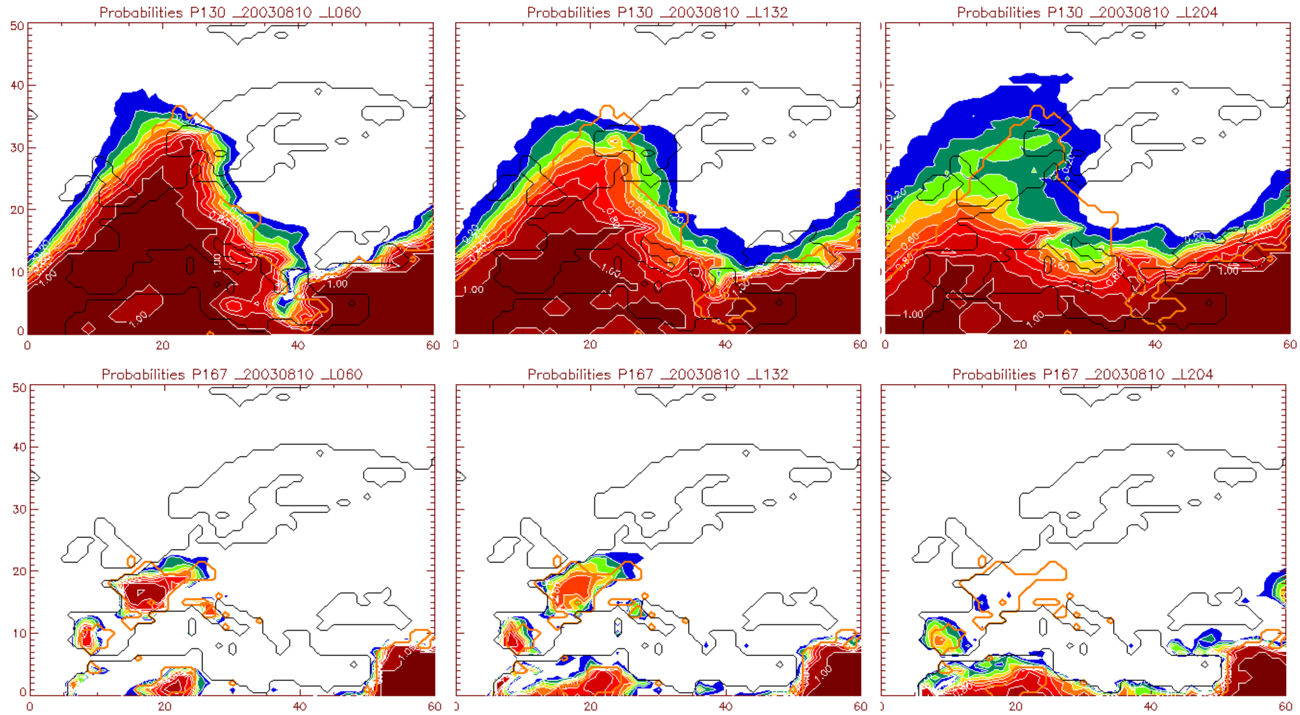
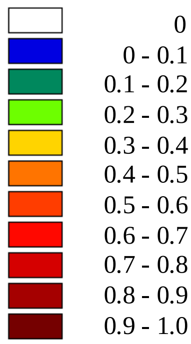



**Figure 7: Example probability forecasts of 850 hPa (P130, above) and 2m temperature (P167, below). Orange line: exceedance of the HeRATE threshold by the analysis.**




shows the contour of the region in which the threshold for a heat wave was exceeded, based on the data of the operational analysis. The colours indicate the probability with which the ensemble members predict an exceedance of the threshold. White means that none of the ensemble members predicts an exceedance and dark red means that an exceedance of the threshold was predicted by all the ensemble members.

While the EPS predicted the 850 hPa (P130) temperature up to 5.5 days quite well, the signs for the 2m temperature (P167) with lead times 132h and 204h were relatively weak or not existent in the EPS. As a consequence, an early warning or information system based only on 2m temperature would have failed to warn the health system of the imminent danger, even with a very low exceedance probability.

## 4. DISCUSSION AND CONCLUSION

In this study the skill of the ECMWF medium-range ensemble prediction system was analysed with respect to its ability to forecast extreme situations such as the one that occurred in summer 2003 in Europe. The skill to predict the exceedance of several absolute thresholds for three lead times (2, 5 and 8 days) was tested for summer 2003. In general, the skill of a forecast is better for shorter lead times than for longer ones. Brier Score (BS) and Brier Skill Score (BSS) are better for the 850 hPa temperature than for the 2m temperature. The skill expressed by the BS is better the shorter the lead time and the higher the threshold to be reached. An explanation for this behaviour could be that the higher the threshold, the rarer this threshold is exceeded, either by the operational analyses and / or by single ensemble members, and the higher the number of grid points with a right null forecast. These findings indicate that for extreme events a good Brier Score is not necessarily an indicator for a highly skilful forecast. With BSS it is the opposite way round: the lower the threshold, the better the skill indicated by the Brier Skill Score. As the Brier Skill Score is the ratio of the BS and the BS of a reference, in our case the null forecast, the BS of the null forecast gets worse as it becomes more probable that the threshold will be exceeded. Therefore, it could be questioned wether BS and BSS are the right measures for the EPS forecast skill with respect to evaluating extreme events.

Thus, it was necessary to look at the summer heat wave of 2003 and especially the heat wave event in August in more detail.

For an ideal probabilistic forecasting system of all occasions when a probability of e.g. 50 % is assigned to an event, that event will occur on 50 % of the occasions (Legg and Mylne, 2004). For summer 2003 probabilities of 50 % and 90 % were analysed. The 50 % probability was analysed to give a hint of the reliability during mean conditions. In summer 2003 for lead times of 2 days and the 850 hPa temperature the reliability for the mean conditions was quite good. For longer lead times there was a slight bias: the EPS predicted mean conditions that were too cold in the northern Spain and Italy and in southern France (the regions most affected by the heat wave). This bias was even stronger (partially more than 2.5 K) in all lead times for the 2 m temperature. In addition, it was even amplified for all lead times and both temperatures when testing the reliability of forecasting extremes (90 % probability). But again the reliability of the 850 hPa temperature forecast was much better than the one for the 2m temperature. These findings indicate that the ECMWF EPS underestimated extreme events for lead times of longer than 2 days in summer 2003.

Lalaurette (2003) stated that the ECMWF medium-range forecast was quite successful in predicting the large-scale flow expressed by the 850 hPa temperature associated with the heat wave up to 5 days in advance. In the short-range (lead times up to 3 days) the model was even able to predict correctly the onset, duration and decrease of the anomaly (Laurette, 2005). In addition, the extreme forecast index (EFI) for the 2 m temperature showed very high values over large areas of Europe. Lalaurette did not define the term heat wave explicitly. When forecasting health relevant heat events the question“what is a heat wave?” always arises. Due to the ability of humans to adapt to local climatic conditions only a relative definition of a heat wave seems to be appropriate. One issue of this study was therefore to analyse the possibility to identify the heat wave signs even for lead times longer than 3 days by using another indicator for a heat wave than the extreme forecast index or an arbitrarily set absolute threshold. Therefore, in this study the threshold for a heat wave was defined based on the HeRATE approach.

The skill of the EPS to predict an exceedance of the HeRATE threshold was tested with respect to temporal and spatial variations. The variation of BSS for the 850 hPa temperature was much lower than the BSS of the 2 m temperature. The skilfulness of the 850 hPa temperature during the whole summer was quite good for all of the lead times but the better the shorter the lead time. The analysis of the spatial variation of the BS and BSS of the 850 hPa during summer 2003 indicated that both were worse in the regions that suffered most from the heat wave for the lead time of 8 days.

When issuing warnings or early information about severe weather based on an ensemble prediction system, it is important to define the probability that should be reached or exceeded for the issuing of the warning / information. In general this probability is based on the ratio between the costs for reacting to a warning to the loss if no measures are taken. Is the cost-loss-ratio low, it is better to issue a warning even if the probability for the event is low. In contrast for a high cost-loss-ratio, warnings should only be issued if there is a high probability. In the case of early heat information, it is very difficult to determine a cost-loss-ratio, because there are many users and it is a more or less ethical question to determine the value of a life saved. We estimate that for early heat information the cost-loss-ratio is low, because it only aims at preparing the health system and not at initiating concrete interventions. The later will be initiated after a heat warning based on a short-range forecast. Our analysis showed that there were signs of a heat wave in the 850 hPa forecast for lead times up to 8 days in summer 2003. However, these signs were relatively weak in the northern parts where the heat wave occurred. There were also signs in the 2 m temperature forecast for lead times up to 5 days. However for the lead time of 8 days no signs could be identified.

The underestimation of the extreme events in summer 2003 and the relatively weak (850 hPa) respective missing (2 m temperature) heat wave signs show us that, in order to catch the signs of an imminent heat wave for lead times of 5 days or more, relatively low exceedance probabilities should be used. This could be also justified by the low-cost-loss ratio for early heat information.

In this study the skills of the 2 m temperature and the 850 h Pa temperature were tested. In general the skill and the reliability of the 850 hPa temperature for mean and extreme conditions is better than for the 2m temperature. In addition there is a stronger spatial variation in the temperature near the ground. This is caused by the modifying effects of orography and land-use. The models for the medium-range forecasts have a resolution that is too coarse to catch all these effects. The 850 hPa temperature, on the other hand, does not depend on these effects but displays better the general state of the atmosphere. Therefore, it seems to be appropriate to use, in a first step, the 850 hPa temperature for the medium-range heat information and to use adiabatic cooling to estimate the temperature at ground level.

A third strategy, that will be tested at a later stage of this project, is to combine the ensemble forecasts with a complete heat budget model of a human being which takes into account all heat exchange mechanisms. The input variables for the heat budget model include air temperature, dew point, wind speed, and cloud coverage, in order to calculate the radiant fluxes. The performance of PT temperature has not been analysed yet. Because the calculation of PT is based on the 2m temperature and other parameters near the surface, it could be expected that the performance of PT is even worse than that of the 2 m temperature. This uncertainty leads to the question as to whether it makes sense at all to use complex indices for heat information in the medium-range.

## ACKNOWLEDGEMENTS

This study was funded by the European Commission DG SANCO for the EuroHEAT project [agreement no. 2004322]. We should also like to thank Dr. Martin Göber for the very fruitful discussions.

## REFERENCES

Basu, R., Samet, J.M., 2002: Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiol. Rev.* **24**, 190-202.

Hajat, S., Koats, R.S., Atkinson, R.W., Haines, A., 2002: Impact of hot temperatures on death in London: a time series approach. *J. Epidemiol. Community Health* **56**, 367-372.

Keatinge, W.R., Donaldson, G.C., Cordioli, E., Marinelli, M., Kunst, A.E., Mackenbach, J.P., Nayha, S., Vuori, I., 2000: Heat related mortality in warm and cold regions of Europe: observational study. *BMJ* **321**, 670-673.

Koppe, C.; Kovats, S.; Jendritzky, G., Menne, B. et al., 2004: Heat-waves: risks and responses. World Health Organization. Regional Office for Europe. *Health and Global Environmental Change*, Series, No. **2**. Kopenhagen, Denmark.

Koppe, C., 2005: Gesundheitsrelevante Bewertung von thermischer Belastung unter Berücksichtigung der kurzfristigen Anpassung der Bevölkerung an die lokalen Witterungsverhältnisse. Dissertation. *Berichte des Deutschen Wetterdienstes* **226**. Deutscher Wetterdienst, Offenbach am Main.

Koppe, C. & Jendritzky, G., 2005: Inclusion of short-term adaptation to thermal stresses in a heat load warning procedure. *Meteorologische Zeitschrift* **14**(2), 271-278.

Lalaurette, F., 2003: The exceptional warm anomalies of summer 2003. *ECMWF Newsletter* No. **99** – Autumn / Winter 2003, 2-8.

Kyselý, J., 2004: Mortality and displaced mortality during heat waves in the Czech Republic. *Int. J. Biometeorol.* **49**, 91-97.

Legg, T.P. & Mylne, K.R., 2004: Early Warnings of Severe Weather from Ensemble Forecast Information. *Weather and Forecasting*, **19**, 891-906.

Persson, A. & Grazzini, F., 2005: User Guide to ECMWF forecast products. Version 4.0, 13 July 2005, ECMWF 2005.