

IDŐJÁRÁS

*Quarterly Journal of the Hungarian Meteorological Service
Vol. 122, No. 3, July – September, 2018, pp. 237–257*

An assessment of daily extreme temperature forecasts – stations average view

Hristo Hristov* and **Andrey Bogatchev**

*National Institute of Meteorology and Hydrology,
Bulgarian Academy of Sciences,
66 Tsarigradsko shose Blvd., 1784 Sofia, Bulgaria*

* *Corresponding author Email: hristo.hristov@meteo.bg*

(Manuscript received in final form September 6, 2017)

Abstract—The present article is the first one of a couple of articles, related to the assessment of the human forecasts (forecast made by weather forecasters). In this article, we have performed an integral assessment of the human-derived extreme temperature forecasts during 2009–2014. It will give us a more general picture of the forecasts, their accuracy over the years of the period in consideration, and their change through the different months of the year. We will show how the accuracy of the forecasts increases in the assessing period, and also how human forecasts underestimate extreme temperatures. The integral assessment gives us a more clear view on the movement of the various errors in time, but it has a significant shortcoming – the spatial distribution of the information is lost. The spatial distribution would give us a valuable feedback, that could be used to correct the forecasts. It will be discussed in the second article, where an assessment by stations will be made.

Key-words: verification, assessment, evaluation, extreme temperature, mean error, mean absolute error

1. Introduction

The purpose of the assessment of the extreme daily temperatures is to look deeper in and to give a more detailed picture of the short-range forecast (forecast for the next day) of the National Institute of Meteorology and Hydrology, Bulgarian Academy of Science (NIMH-BAS), Bulgaria. The results of the forecast analysis (the assessment) will provide a feedback to the forecaster, that should lead to an improvement of the forecast's quality.

In Bulgaria, meteorological observations has been effectuated for more than 100 years, and civil meteorological forecasts have been drawn up for more than 60 years. Regardless this long period, such extensive and in-depth study of the human weather forecasts issued by the National Weather Service has never been done up to now.

For this aim, different kinds of errors, as well as the percent of correct forecasts are calculated, using the guidance of the World Meteorological Organization (<http://www.wmo.int/pages/prog/amp/pwsp/pdf/TD-1023.pdf>). It uses ± 2 °C or ± 3 °C as a limit when calculating the percent correct. This assessment is very suitable to be used for the general public or other users. This study uses a margin of admissible error up to 2 °C, i.e., ± 2 °C for calculating the percent correct. In order to be trusted by the public, to trust a weather forecast needs to be both: accurate and consistent. Accuracy is important, as public will not trust the forecast that has proven to be inaccurate over a time period, especially if this period is short (just a few days). A consistent forecast, covering usually seven days, is a forecast sequence that does not have large swings (flip-flops) from one forecast to the next (for the same date). Consistency is just as important as accuracy, because the public will not trust to forecasts that flip-flops and changes on a day to day basis (*Lashley et al.*, 2008). These flip flops of the forecast (for one and the same day) are measured with the Ruth-Glahn Forecast Convergence Score (FCS), developed by *David Ruth* and *Harry Glahn* (*Ruth et al.*, 2009).

Brier and *Allen* (1951) point some economical, administrative, and scientific reasons for evaluation. *A.H. Murphy* (1993) says that goodness of a forecast has different faces and determines three types:

- consistency (correspondence between the forecasters' judgment and the forecast),
- quality (correspondence between the forecasts and the matching observations), and
- value (incremental economic and/or other benefits realized by decision makers through the use of the forecasts).

In this work, only the second type (quality) will be discussed.

Globally, there are number of researches that approach the present work. There, the quality of the forecasts for minimum and maximum temperatures and precipitations is calculated. In *Riply* and *Archibold* (2002), several points in Canada (with different climate), in the year 2000, were assessed by calculating the quality of the short-range and medium-range (up to 5 days) forecasts. Determined parameters were: mean error, mean absolute error, Skill score (climate norm was used) for temperatures and Brier score (*Brier*, 1950) for the precipitation probability. It was shown how forecast errors gradually increase from day 1 to day 5.

In *Fajman* (2011), for the area of Omaha Valley during the period 2008–2009, basic errors were calculated, such as: mean error, mean absolute error, Brier score, and forecast convergence score, for maximum and minimum temperatures and precipitation probability for a 3-day and a 7-days period. These statistics were calculated for the forecasts issued from the Meteorological Office (human forecasts) and for the forecasts produced by the GFS model. It was emphasized that the comparison between these two forecasts showed that human forecasts were better. Although, it has to be mentioned, that human forecasts were based mainly on the numerical model predictions.

In 1998 a verification program, named SOLVER, started at the Meteorological Office Corpus Christi in Texas (*Wilk*, 2005). The software package (SOLVER) was developed by Jamie Frederic, a chief forecaster at the meteorological office in Tulsa, Oklahoma at that time. Over the following few years, this software underwent some changes, and since 2003, an overall assessment and personal assessments (sent via e-mail to the weather forecasters) have been made for minimum temperatures, maximum temperatures, and precipitation every month and every six months (for the cold and warm halves of the year). The computer program calculated mean error, mean absolute error as well as percent correct with deviation up to 3 °F for minimum and maximum temperatures. For the precipitation, the Brier score was calculated. These errors were calculated for 6 points using 5 lead-time periods (each having 12 hours) and incorporated human forecasts and the three numerical models in use: GFS (Global Forecast System), NAM (North American Mesoscale Model), NGM (Nested Grid Model). A comparison of errors and results has been made in form of graphics, and placed in the forecasters' restroom. Furthermore, the personal assessments gave possibility for competition between the forecasters, and those with the best results (every six months) received diplomas. The purpose of this verification program was to provide the best forecasts for the area of responsibility of Corpus Christi. In the paper of *Wilk* (2007), the methodology described in *Wilk* (2005) was used. Moreover it was found that the human forecast was about 10% better than the numerical forecast and, furthermore, the Meteorological Office of Corpus Christi had the best results, compared to all other meteorological offices in the USA.

In Bulgaria, in the work of *Spiridonov* (1987), various errors were calculated for minimum and maximum temperatures in Sofia during the period 1983–1984. Beside the human forecast's error, the persistence forecast error was calculated as well. However, the mean error was not determined, which would have given the systematic error of the forecast.

In *Bogatchev* (1988), a scheme was developed for a medium-range forecast of the daily mean temperature and the daily amounts of precipitation for five days (day by day), based on the field AT500 of the ECMWF model. The forecast was made for 8 stations on the territory of Bulgaria. An assessment of the results was made, and it showed that they were satisfactory.

In *Bogatchev* (1994), the work of *Bogatchev* (1988) was described and a predictive scheme (statistical) for forecasting extreme temperatures up to six days for 17 stations in Bulgaria was created. The scheme was developed on the basis of the fields: AT500, surface pressure, and T850, derived from the ECMWF model. Various types of evaluation of the output data have been made; also a comparison with the mean absolute error of the human forecast for the period of November 1991 – October 1992.

In the present work, the method of the continuous variables is used for evaluation; a useful tool is the website of the Joint Working Group on Forecast Verification Research (<http://cawcr.gov.au/projects/verification/>). It gives a good presentation of the modern assessment methods, some of which are used in this study.

2. Methodology

The short-range weather forecast of NIMH is up to 36-48 hours ahead, and it is prepared till 11 a.m. every day. It consists of a forecast for today and a forecast for the next day (that will be assessed). Besides of the text forecast, a numerical forecast is issued as well; it includes minimum and maximum temperatures and a symbol for the phenomena expected (sun, clouds, rain, snow, thunders, fog, etc.). The numerical forecast is elaborated for 68 points in Bulgaria, 37 of which coincide with synoptic stations. For these 37 points (shown in *Fig. 1*) the assessment will be carried out. Four of these 37 points are mountainous. The period for assessing is the years 2009–2014.

In the followings we give the definitions of the error estimations used in this study.

Mean error or bias. It can be interpreted as systematic error. The mean error indicates the difference between the forecast and the observations for the day. For a definite period of time, the sign and the magnitude of the systematic error can be seen.

$$ME = \frac{1}{n} \sum_{i=1}^n (F_i - O_i) \quad (1)$$

where F_i and O_i are the values of the forecast and observations for day i , respectively.

Mean Absolute error. It determines the accuracy of the forecast. It does not take into account the direction of the error, only its magnitude.

$$MAE = A_h = \frac{1}{n} \sum_{i=1}^n |F_i - O_i| \quad (2)$$

where h is human.

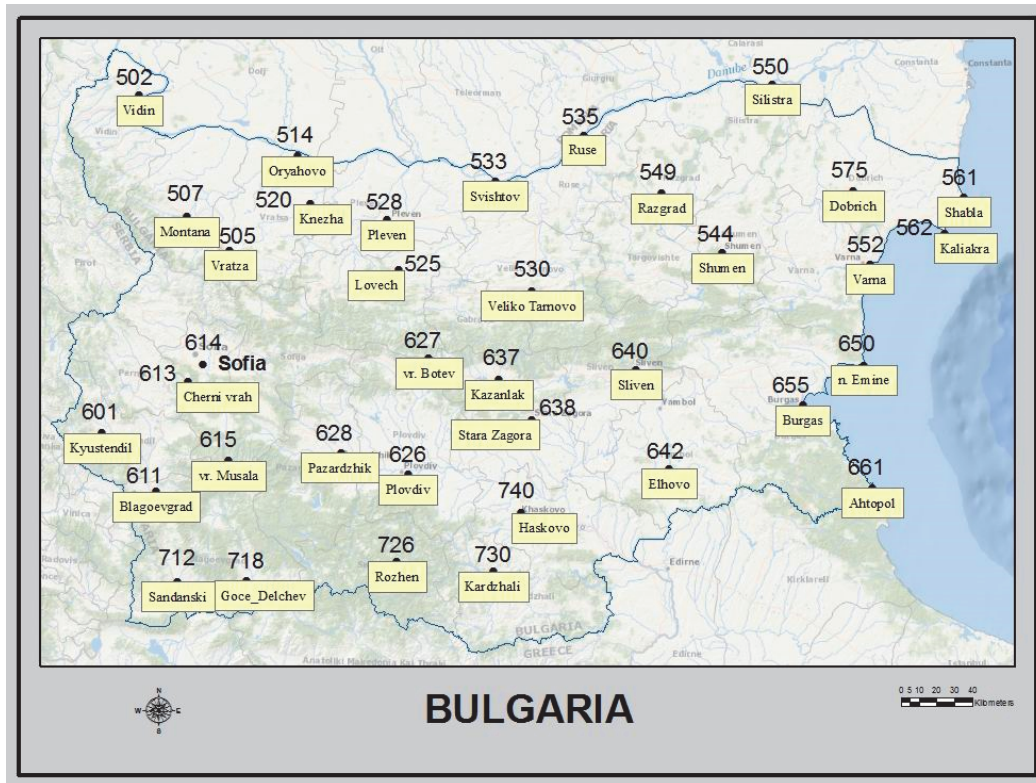


Fig. 1. Map of the points, where the assessment carried out, names of the stations, and their WMO code. First 2 digits (15) are omitted.

Many of the meteorological elements, including the temperature, do not vary very sharply, and they are often close to their previous values. So, if the climatic value or persistence forecast (recent observations) is issued as a forecast, the result would be fairly good (with small errors). It must be acknowledged that a forecaster, using a persistence forecast, would get even better results. To assess the contribution of the persistence forecast to the human forecast of the temperature, it has to be determined. So, to calculate the persistence forecast's error, O_i is used instead of F_i , and O_i is replaced by $O_{(i-1)}$.

$$MAE_p = A_p = \frac{1}{n} \sum_{i=1}^n |O_i - O_{(i-1)}| \quad (3)$$

where p is persistence.

The persistence error is suitable to be used for the forecast of the minimum temperatures, as they are available (at 8 or 9 a. m. in local time, depending on the season) before the forecast is issued (11 a. m.). However, for the maximum temperatures, the last data available are those from the previous day. Thus, there are 2 days difference between observations and forecast. Thus, to get the real error of the persistence forecast, as will be in practice, it should be calculated as follows:

$$MAE_{P_{Tmax}} = A_P = \frac{1}{n} \sum_{i=1}^n |O_i - O_{(i-2)}| \quad (4)$$

In 2012, in the operational forecasting office of NIMH BAS, a program was released in operational mode, that evaluates the predicted temperatures (minimum and maximum) for 37 points (shown on the map above). Every day it calculates the mean error, mean absolute error, maximum error, mean square error, and the percent correct for these 37 stations. Thus, in near-real-time, the forecaster can see these types of errors of his/her own forecast. He can look in more details in the meteorological conditions and analyze the consequences and the reasons for the deviations in his/her forecast.

Percent correct is determined as followed:

$$PC = \frac{1}{n} \sum_{i=1}^n \begin{cases} \text{correct, if } (F_i - O_i) \leq 2^\circ\text{C} \\ \text{not correct, if } (F_i - O_i) > 2^\circ\text{C} \end{cases} \times 100\% \quad (5)$$

Using Eqs. (2), (3), and (4), the Skill scores of the forecaster can be calculated for the minimum and maximum temperatures using the formula:

$$\text{Skill scores} = \left(\frac{A_p - A_h}{A_p} \right) 100\% \quad (6)$$

where A_h is *MAE* of human forecast (Eq. (2)), and A_p is *MAE* of persistence forecast (Eqs. (3) and (4) – the minimum and maximum temperature respectively). *Skill scores* change from 0% persistence forecast to 100% perfect forecast.

3. Results

To make a stations-averaged (integral) assessment, first the assessment is made in space and then in time, i.e., first, all points for a given day are evaluated (as it is made by the internal verification program), and then they are evaluated for a certain past period. Here, we look at the data by years and by months. The information that will be presented is: mean error (*ME*), mean absolute error (*MAE*), percent correct (*PC*), *Skill scores* and *ME*, *MAE* and *PC* of the persistent forecast.

3.1. Mean absolute error (*MAE*)

We will discuss the mean absolute error, first by years (*Fig. 2*).

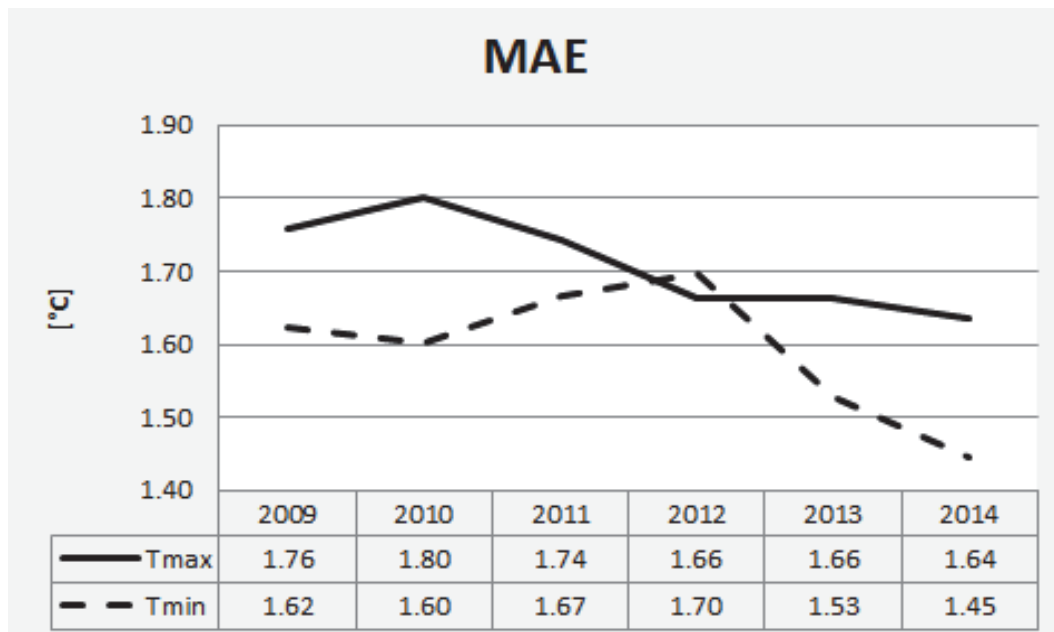


Fig. 2. Mean absolute error of the human forecast for the minimum and maximum temperatures.

It is seen that:

- The minimum temperature *MAE* is less than the maximum temperature *MAE*, i.e., the minimum temperature forecast is more accurate than the maximum temperature forecast.
- In general, the course of *MAE* decreases with the years, due to various factors, such as: improvement of the numerical models, accumulation of more experience from the forecasters, partially probably the release in an operational mode of evaluation system, and others.

At the minimum temperatures, there is a slight increase of the error in 2011 and 2012 with less than 0.1 °C. The maximum temperatures have an increase of the error in 2010 and a delayed decrease of the error in 2013. These fluctuations are normal and can be due to different factors.

In Fig. 3, the *MAE* of the persistence forecast is added to the *MAE* of the human forecast. *Tmax_In1* is also presented just to make the comparison between *MAE* of the persistence forecast for the minimum and maximum temperatures. Later on, the *Tmax_In1* will not be used, and *Tmax_In2* will be discussed instead (this would be the real conditions error).

To compare, considering *Tmax_In1* and *Tmin_In1*, it is seen (Fig. 3) that the error in the minimum temperatures is smaller, that leads to the conclusion that the fluctuations (day to day) in the minimum temperatures are smaller.

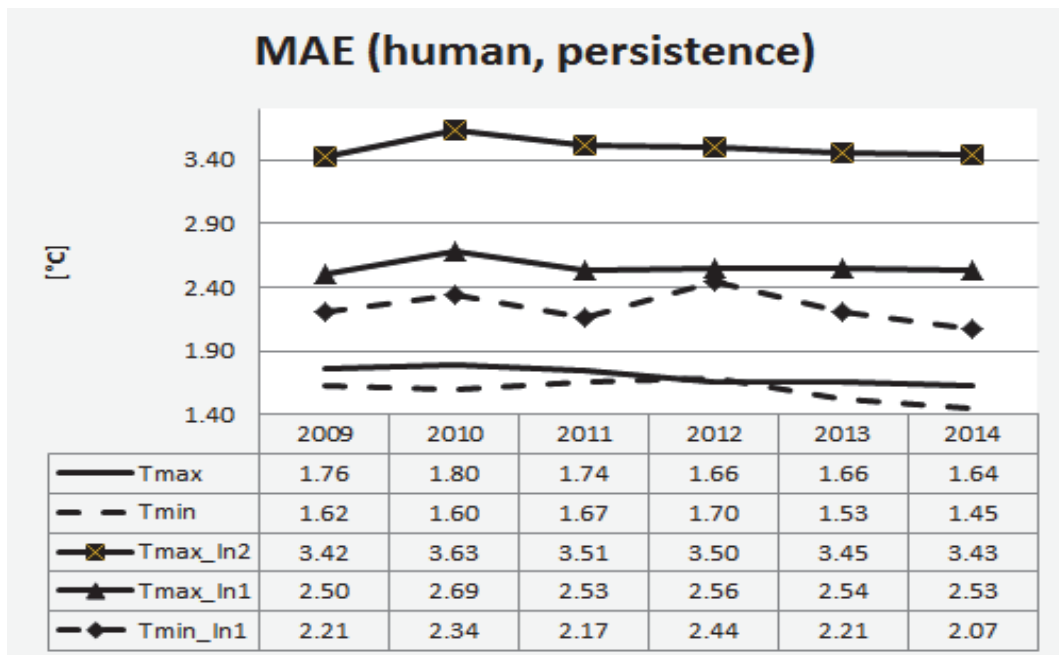


Fig. 3. MAE for the human and persistence forecasts. $Tmax_In2$ is MAE of the persistence forecast for the maximum temperatures calculated in Eq.(4), i.e. the forecast and the observations difference by two days. $Tmax_In1$ and $Tmin_In1$ MAE of the persistence forecast for the maximum and the minimum temperatures calculated by Eq.(3), or one day difference between the prediction and the observation.

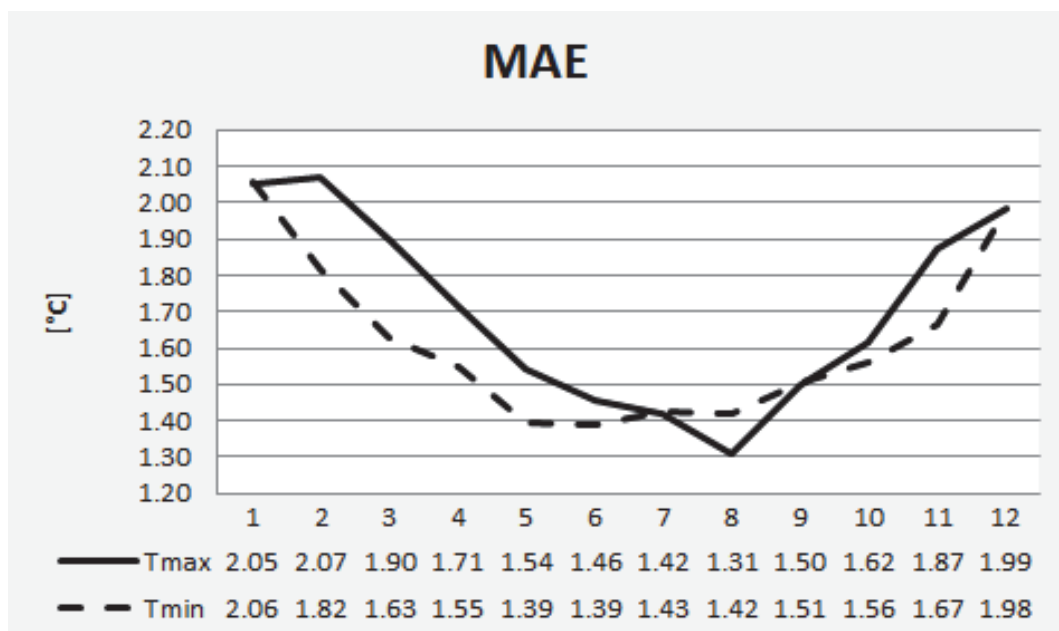


Fig. 4. MAE for the minimum and maximum temperatures of the human forecast.

Looking now at $Tmax_In2$ and $Tmin_In1$ (Fig. 3), it is seen that the persistence forecast error is significantly greater than the forecast error of the meteorological office. It is especially true for the maximum temperature, where the difference is almost 2 °C, and this is a double, in some years even, greater

error. For the minimum temperature, the difference is about 0.5 °C. Besides, the correlation between $Tmin_In1$ and $Tmin$ is high, and this leads to the conclusion that the influence of the persistence forecast on the minimum temperatures is quite big and moreover, it is bigger than the maximum temperatures.

Let us now consider the mean absolute error changes in months. The monthly averaged absolute error for the minimum and maximum temperatures is presented in *Fig. 4*.

One can see in *Fig. 4* that, during the warm months, *MAE* decreases and during the cold months, it grows. The course of the maximum temperatures shows a minimum in August and a maximum in February. The difference between August and February is nearly 0.8 °C, and this is a significant difference in the absolute error. The results are almost identical to those shown by *Wilk (2005)*. At the minimum temperatures, there is a minimum in May and June (*Fig. 4*) and a maximum in January; the difference between them is nearly 0.7 °C, i.e., the fluctuations in the minimum temperatures' absolute error are slightly smaller than those in maximum temperatures.

Let us add now the persistence forecast error (*Fig. 5*).

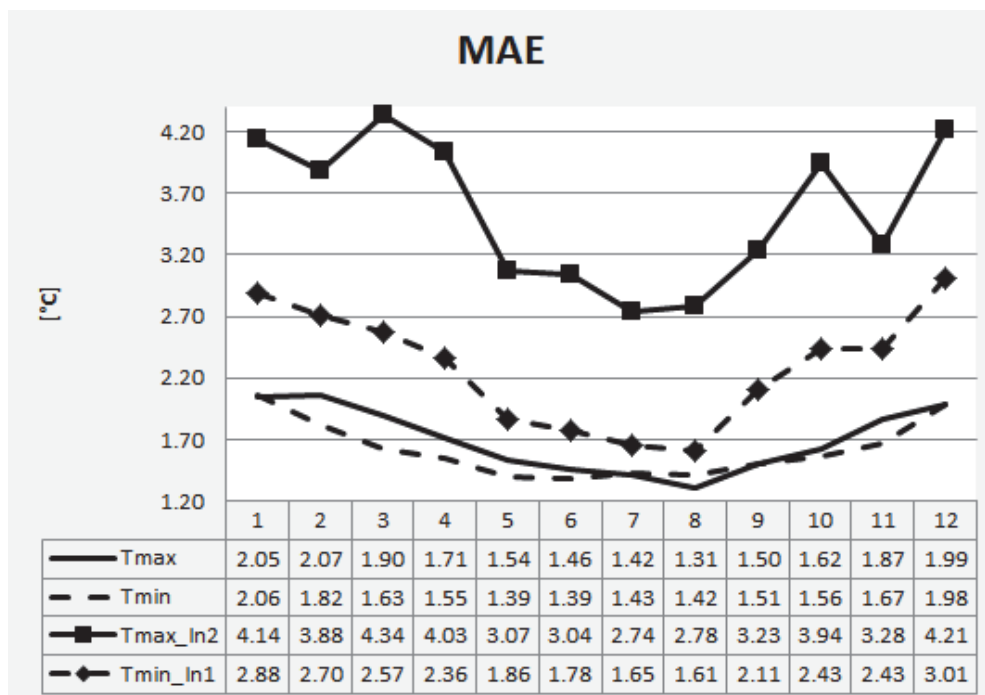


Fig. 5. MAE of the human and persistence forecasts. Tmax_In2 is MAE of the persistence forecast for the maximum temperatures calculated by Eq. (4). Tmin_In1 is MAE of the persistence forecast for minimum temperatures calculated by Eq. (3).

Fig. 5 shows that *MAE* of the persistence forecast also decreases during the warm months, and this is logical, since during the warm months the atmosphere is more stable. In addition, the correlation between the *MAE* of the human and

persistence forecasts is good enough (especially for the minimum temperatures), so that a conclusion can be drawn that the contribution of the persistence forecast in the human forecast is big, and it is bigger at the minimum temperatures. Moreover, the figure shows that the *MAE* of the persistence forecast for minimum temperatures (*Tmin_In1*) during the warm months is less than the human forecast error during the cold months, i.e., the persistence forecast in summer is more accurate than human forecast in winter.

At the maximum temperatures (*Tmax_In2*), the situation is a little different. The error is much greater than the human error. But it is normal to expect, since when calculating *Tmax_In2*, a two-day difference is used, moreover, the fluctuations (from day to day) of the maximum temperatures are higher.

3.2. Percent correct (PC)

As mentioned before, a forecast with an error up to 2 °C is considered to be correct. *Fig. 6* presents the *PC* for maximum and minimum temperatures, by years.

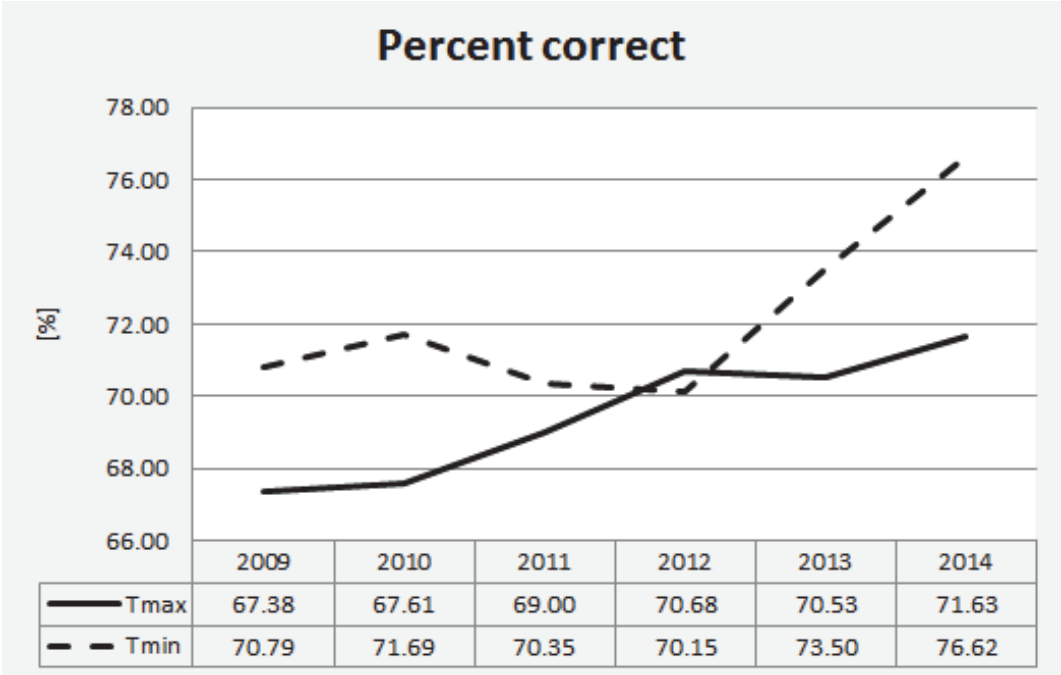


Fig. 6. Percent correct for the human forecast

Fig. 6 indicates that the *PC* for minimum temperatures is greater than for the maximum temperatures. Furthermore, the *PC* increases over the years (contrary to *MAE*); for the maximum temperatures this rise is a little more than 4%, and for the minimum temperatures nearly 6%. At the minimum

temperatures, there is a slight decrease of the *PC* in 2011 and 2012, which fully corresponds with the increase of the *MAE* in these two years.

At the maximum temperatures, there is a drop in 2013 for the *PC*, but the *MAE* does not have a rise (2012 and 2013 have the same error). In addition, the *MAE* has an increase in 2010, but it does not correspond to the *PC* (also increasing). It is completely normal to get, if *MAE* is more evenly distributed and more stations fall in this interval of 2 °C. So, it is seen that the *MAE* and the *PC* are not necessarily always opposite.

Now, let's add the *PC* of the persistence forecast (*Fig. 7*).

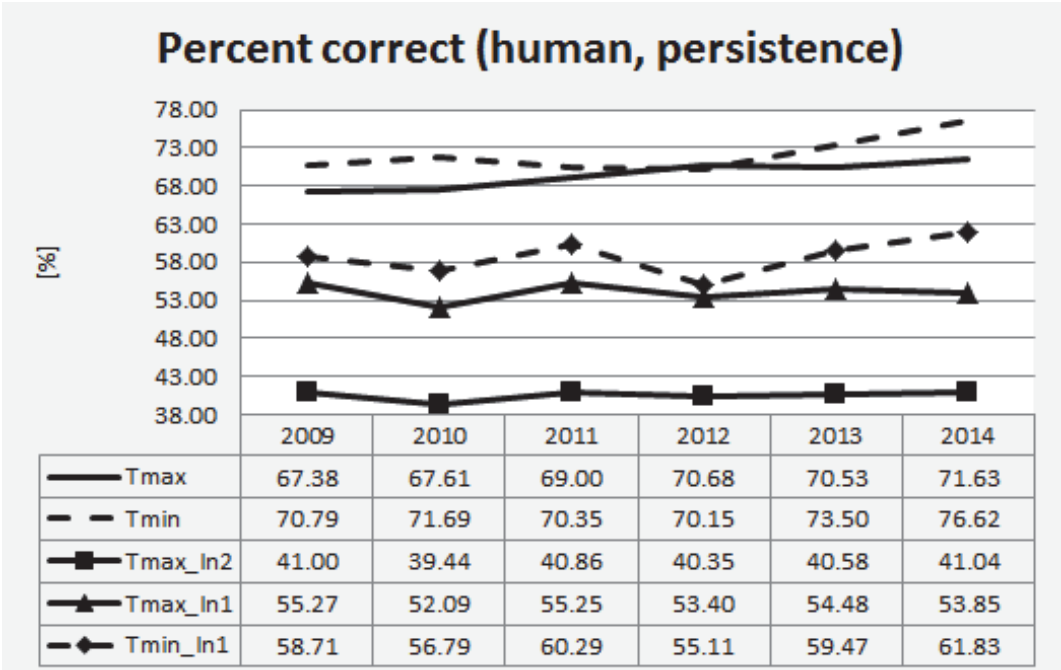


Fig. 7. Percent correct for the human and persistence forecast. *Tmin_In1* and *Tmax_In1* are the *PC* for the minimum and maximum temperatures with one day difference between forecast and observation. *Tmax_In2* is the *PC* for the maximum temperature with two days difference.

Here *Tmax_In2*, *Tmax_In1*, and *Tmin_In1* are described below *Fig. 3*, and *Tmax_In1* is used only for an initial comparison, from which it is seen that the maximum temperature has a smaller *PC* (53–55%) compared to the minimum temperature (55–61%). Moreover, *Tmin_In1* decrease in 2012 coincides with the fall in *Tmin*, proving once again the big influence of the persistence forecast on the human forecasts.

We made the difference between the *PC* of the human forecasts and the *PC* of the persistence forecast: $Tmin - Tmin_In1$ and $Tmax - Tmax_In2$, with the aim to check approximately what percentage from the *PC* is not due to the persistence forecast. These are presented in *Fig. 8*.

Fig. 8 indicates that at the minimum temperatures, from 10% to 15% of the errors are NOT due to the persistence forecast. In fact, some of these 10–15% are due to the persistence forecast with correction, so the forecaster’s experience is of great importance.

Can we state that at minimum temperatures between 85 and 90% of the human forecasts are due to persistence forecast? The answer is NO.

The reason is that the *PC* of the human forecast, which is about 70–76% for the minimum temperatures (Fig. 7), is a superposition of three elements: persistence forecast, numerical models, and the forecasters’ decisions (strongly depending on his/her experience). Till now, we studied just one of these three elements, so in this ”equation” still two unknown remain. In a next work, we shall investigate the numerical models and find one more of factors. Then, only one unknown element, the forecaster’s decision will remain; and we would be, to a great extent, able to say what parts of the human forecast is due to the numerical models and to the persistence forecast.

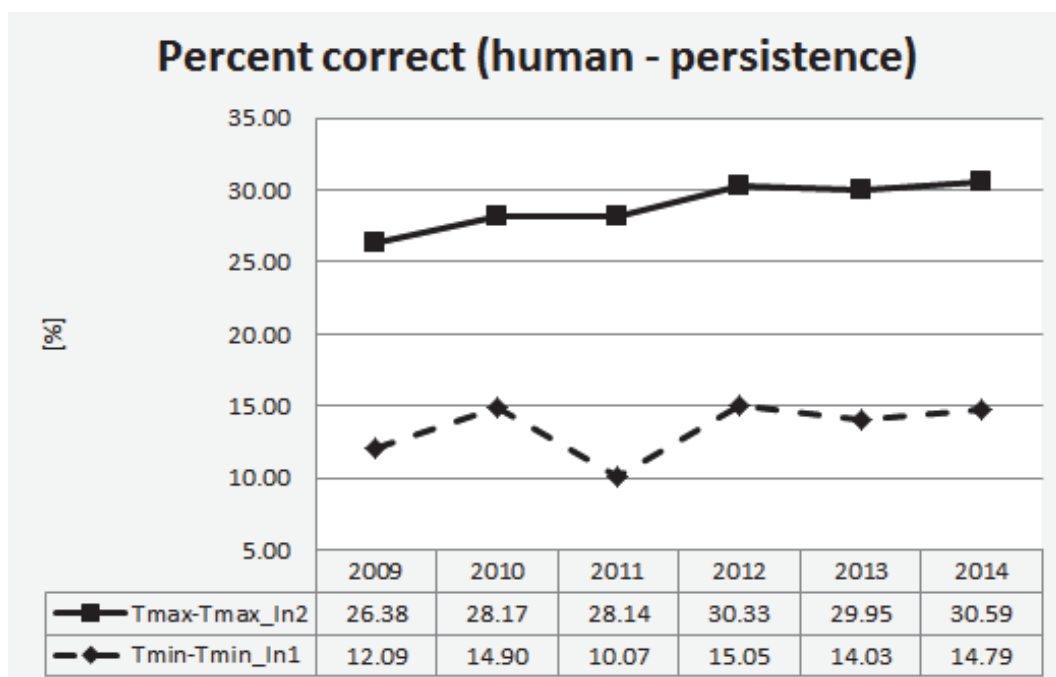


Fig. 8. Percent correct of the human forecast minus persistence forecast.

So far, we could only state (based on 15 years experiences as forecaster), that the impact of the persistence forecast on the human is really big. The reason is that every morning after the arriving of the synoptic data at 6 UTC, minimum temperatures are available, and the forecaster uses them to prepare the forecast for the minimum temperatures for the next day, making adjustments. This technique is very useful for a static atmosphere, but not applicable in case of

active dynamics (a cold front passage, for example). However, cases with more intensive dynamics occur relatively rarely compared to cases with more static atmosphere, that is why the *PC* of the persistence forecast (*Fig. 7 – Tmin_In1*) is about 60%, i.e., the big impact of the persistence forecast on the human is obvious.

At the maximum temperatures, this percentage (of NON-persistence forecast) is significantly higher, between 26 and 30%. Also, an increase of about 4% is noticed over the years. *Fig. 7* shows that the *PC* of the maximum temperatures (*Tmax_In2*) is almost constant over the years. It follows that these 4% of increase are due to other factors, as numerical models and forecaster’s experience. In addition, the *PC* is not influenced by the type of the year. For example, 2014 is one of the most humid years (more dynamics), but nevertheless, no drop in the difference is noticed (*Fig. 8*) in the drier years (when the influence of the persistence forecast should be greater). This is a proof that, in case of more dynamics (little influence of the persistence forecast), numerical models and forecaster’s experience compensate the smaller impact of the persistence forecast.

The difference between the *PC* in the maximum and minimum temperatures in *Fig. 8* is 15–20%. This difference results, on one hand, from the fact that for the maximum temperatures a persistence forecast from two days ago is used, and on the other hand, from the fact that at minimum temperatures fluctuations (day to day changes) are smaller.

Looking at the skill scores in *Fig. 9*, calculated using Eq. (6), a big difference is seen between minimum and maximum temperatures.

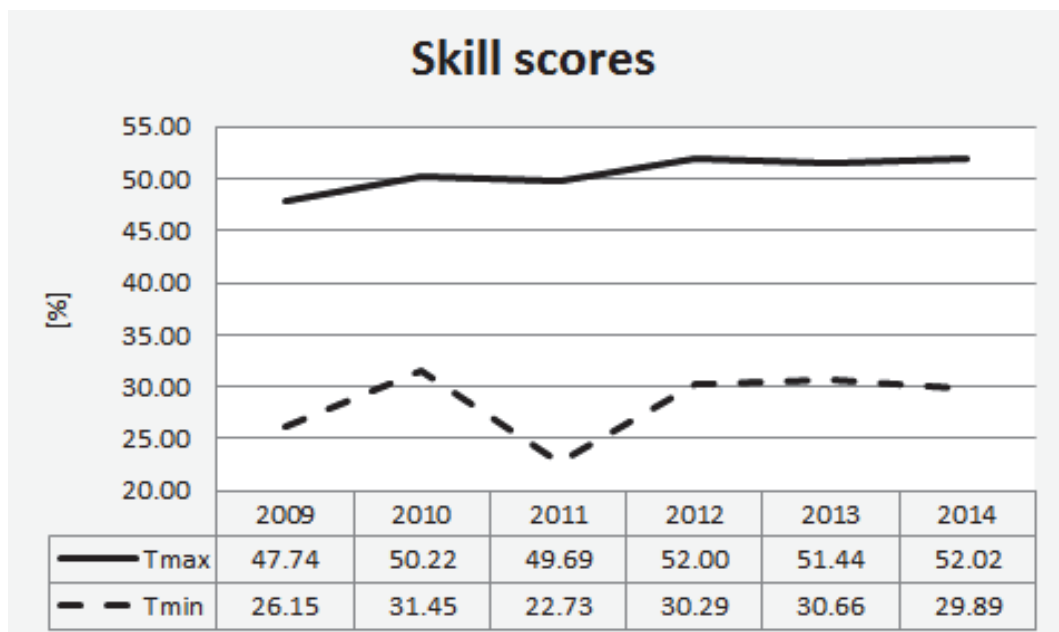


Fig. 9. Skill scores of the human forecast for the minimum and maximum temperatures.

Skill scores at minimum temperatures is 25–30% and at the maximum temperatures about 50%, i.e., the difference in between is 20–25%. This corresponds to the above statements, namely that at the maximum temperatures about 30% of the errors are due to the non-persistence forecast, i.e., to the numerical models and the **forecaster’s skill**; at the minimum temperatures they are 10–15%. These numbers are just obtained in a different way, not with the formula for the skill scores; that’s why they are different, but the end results and the conclusions are similar.

Now, let’s go back to the results of the skill scores. The reason for this big difference in the skills at minimum and maximum temperatures is that in calculating the skill scores of the maximum temperatures, a *MAE* of the persistence forecast from two days ago is used, and these results in nearly 2 °C difference between the human and the persistence forecasts (*Fig. 3*). At minimum temperatures, the *MAE* of persistence forecast is from just one day before, and thus, the difference between the human and persistence forecasts is only about 0.5 °C. Even if we use the *MAE* of the persistence forecast at the maximum temperature from one day before (*Fig. 3*), we would have a difference of 0.7–0.9 °C between the human and persistence forecasts, i.e., skills at maximum temperatures would be greater again.

Therefore, we think that the main reason for the smaller skill for the minimum temperatures is the smaller fluctuation (day to day changes) in the minimum temperatures, which leads to a smaller error in the persistence forecast, and on its side, this leads to less skills at minimum temperatures.

Let’s now have a look at the *PC* by months. We start with the averaged *PC* at the human forecasts (*Fig. 10*).

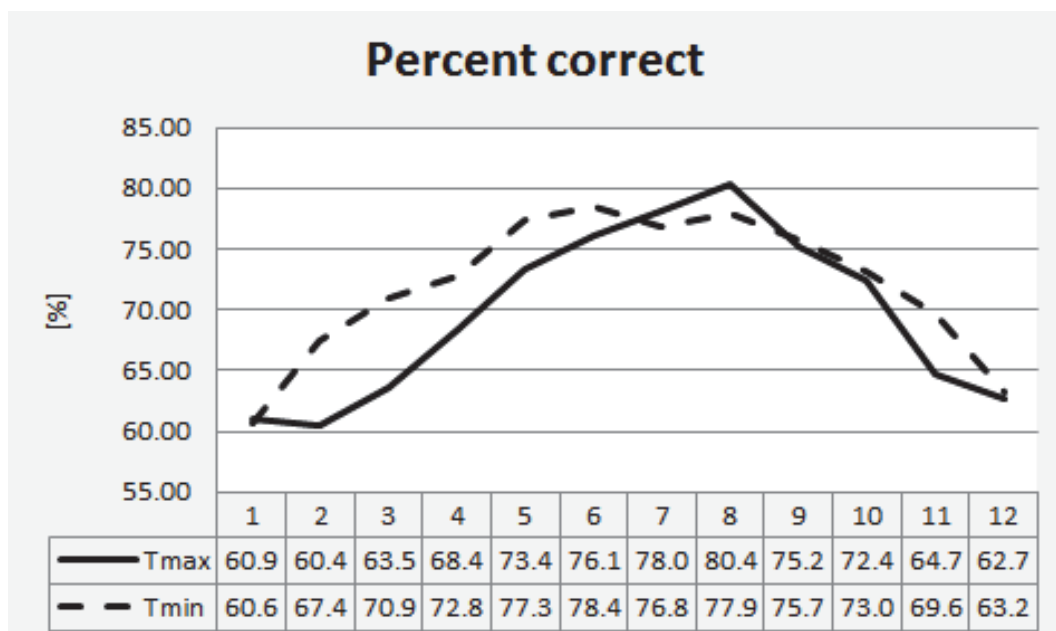


Fig. 10. Monthly averages of *PC* of the human forecast for the minimum and the maximum temperatures.

Looking at *Fig. 10*, it is noticed that the percent correct gradually grows from winter to summer; for the maximum temperatures it is the highest in August and the lowest in February. This totally corresponds (inversely proportional) with the *MAE* (*Fig. 4*), unlike the distribution by years. In addition, in most months, the *PC* of the human forecasts at the minimum temperatures is greater than at the maximum temperatures. At the minimum temperatures, there is a minimum in January, and the maximum is in June; this almost totally corresponds to the *MAE*. There is a difference only in May (the *MAE* in May and June is the same, while the *PC* in May is smaller than in June). This confirms the statement above, that the *MAE* and the *PC* are not necessarily always opposite (in inverse correlation).

In *Fig. 11*, the averaged *PC* of the persistence forecast (*Tmin_In1* and *Tmax_In2*) by months are added to the averaged *PC* of the human forecast (*Tmin* and *Tmax*). The Fig. shows that there is a good correlation between human and the persistence forecast.

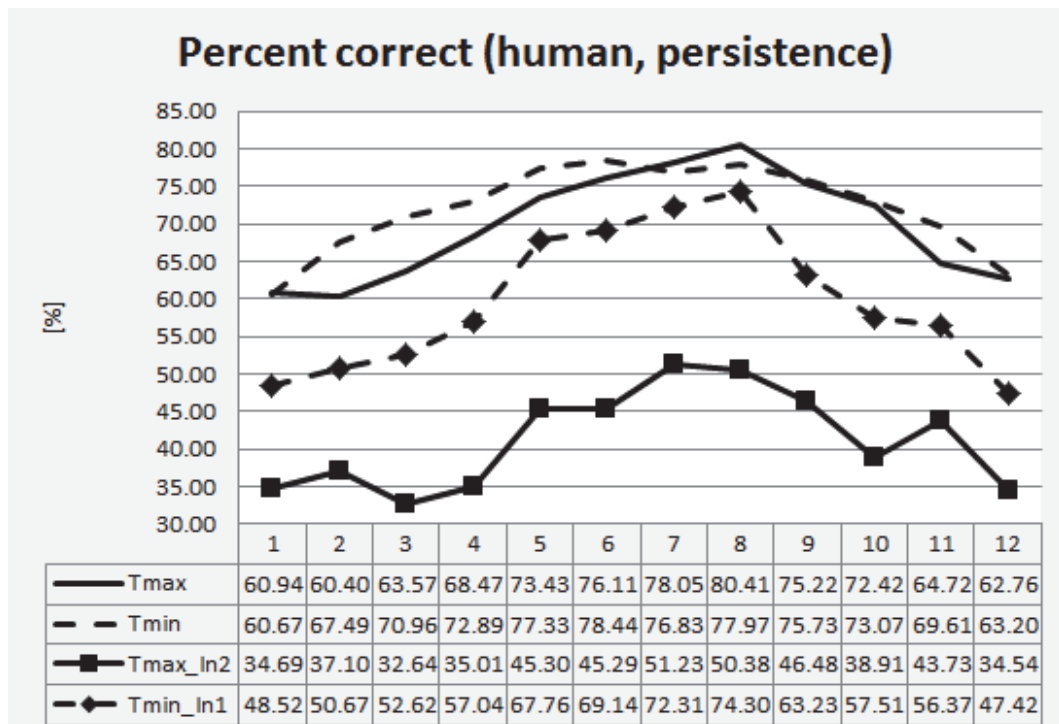


Fig. 11. Monthly averages of *PC* for the minimum and maximum temperatures of the human and persistence forecasts.

In the next figure (*Fig. 12*), we will show what percentage is due to the non-persistence forecast, i.e., we will make the difference: $Tmin - Tmin_In1$ and $Tmax - Tmax_In2$.

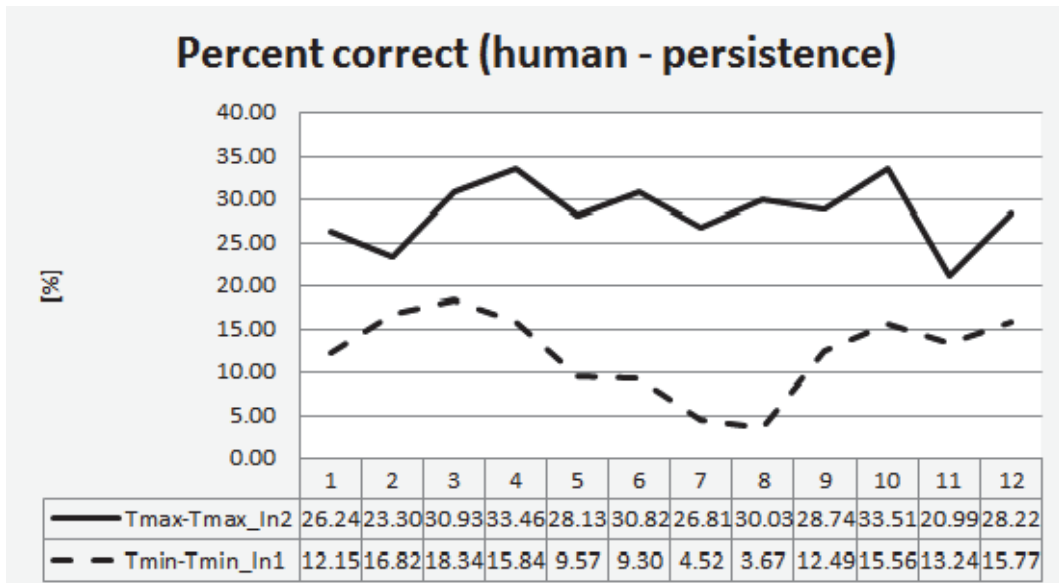


Fig. 12. Monthly averages of the PC of the human forecast minus the persistence forecast.

Fig. 12 indicates that at the minimum temperatures, the non-persistence forecast changes between 4% and 18%, the minimum of 4% is in July and August, and this is the proof for the big impact of the persistence forecast on the human forecast in summer, when dynamics in the atmosphere are weak. Looking at the situation over the years (Fig. 13), it is seen that in August 2010 and 2013, there are even negative values, which means that the persistence forecast is better than the human.

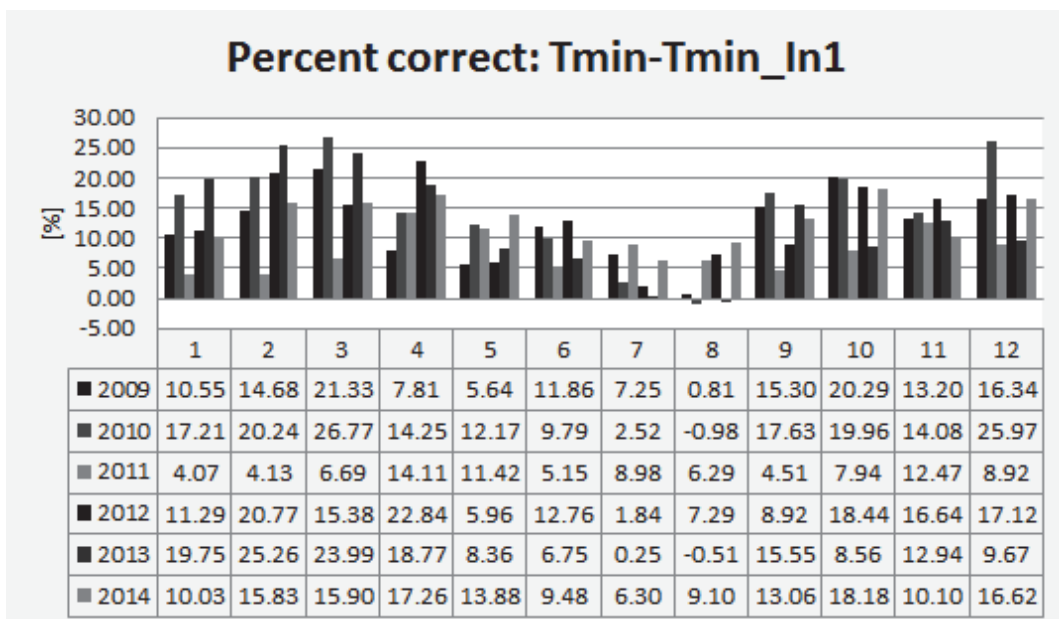


Fig. 13. Monthly and yearly distribution of PC for the minimum temperatures of the human forecast minus the persistence forecast.

For the maximum temperatures (*Fig. 12*), the difference between the human and persistence forecasts (with some variations) is also constant, i.e., about 30% of the human are due to the non-persistence forecast (as it was when looking by years). There are two big minima in February and in November – the persistence forecast had higher values (*Fig. 11*). In fact, there is a minimum at the minimum temperatures as well, but it is less pronounced.

3.3. Mean error (ME)

Finally for this study, we will discuss a very important error - the mean error. First, we shall examine it by years (*Fig. 14*), however, here we shall not consider the change over the years, but the concrete values only.

The mean error of the persistence forecast is also included in *Fig. 14*. Looking at the errors (human forecast) of the minimum and maximum temperatures, it is seen that during all of the six years, they have negative values, i.e., there is a systematic underestimate in the minimum and maximum temperatures forecast.

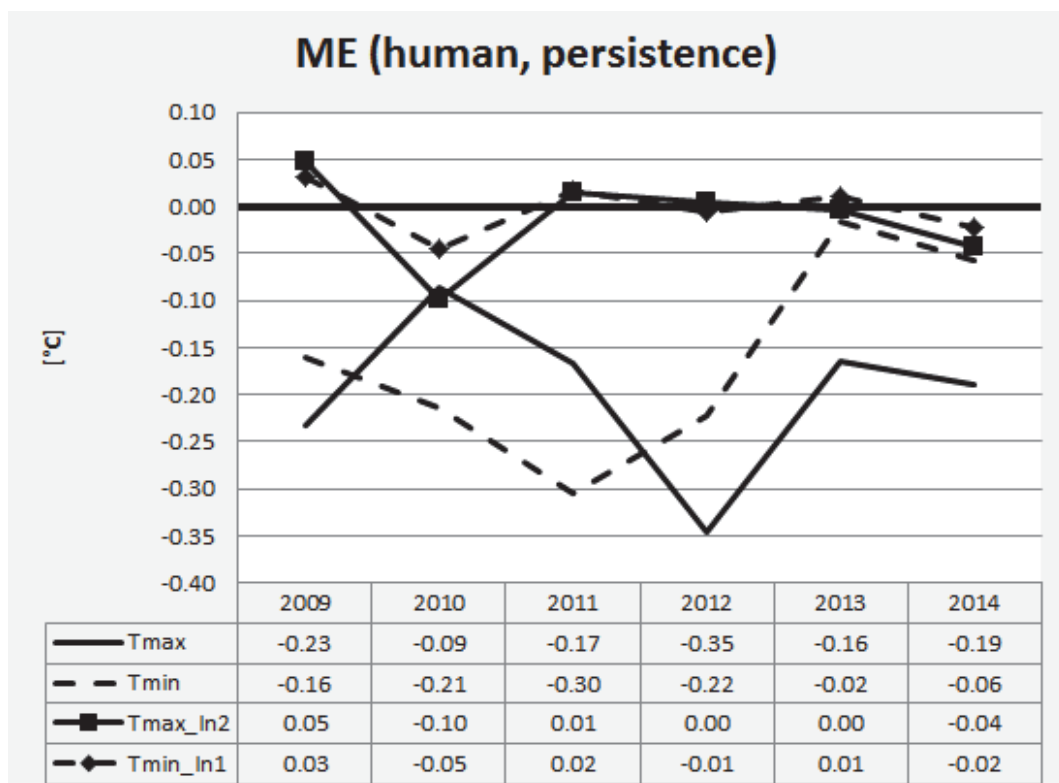


Fig. 14. Mean error of the minimum and the maximum temperatures of the human and persistence forecasts through the different years. *Tmin_In1* is the persistence forecast of the minimum temperatures from one day ago, and *Tmax_In2* is the persistence forecast for the maximum temperature from two days ago.

It is also seen that at the minimum temperatures, the biggest systematic error is in 2011, while that at the maximum temperatures is in 2012. It has to be mentioned as well, that over the past two years, the systematic error at the minimum temperatures is very small. Looking at the mean error of the persistence forecast, it is seen that it is very small, almost there is no systematic error. As mentioned above, when elaborating the forecast, the forecaster uses the persistence forecast, the numerical models, and his own experience. Based on personal experience, the numerical models underestimate the maximum temperatures, i.e., they have a negative systematic error (we shall discuss this error in a further work). From everything said till now, a conclusion can be drawn that the systematic error is mainly due to the numerical forecasts and the weather forecasters.

Let us now discuss the situation by months (*Fig. 15*). What impresses is that there is an underestimate (systematic error) from September till April, while during the warm months (from May to August) the deviation is minimal and positive. In other words, during the cold months, the temperatures are underestimated, and during the warm months they are overestimated.

One can say that if the systematic error could be reduced, then the forecast accuracy would be improved. This is, namely, the purpose of this study. To a great extent, this purpose can be achieved through the results in the second panel part, where assessment by stations will be made.

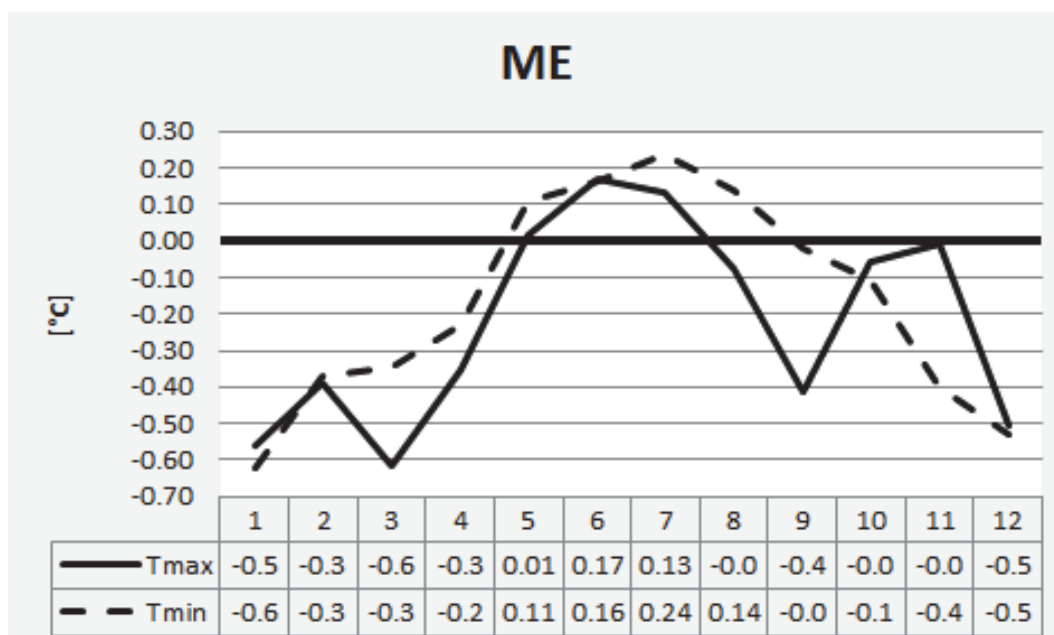


Fig. 15. Mean error for the minimum and maximum temperatures of the human forecast through the different months.

Next figure (*Fig. 16*) comprises the mean error of the persistence forecast. *Fig. 16* indicates the conclusion, that during the transition months (spring and autumn), there is a bigger systematic error of the persistence forecast, especially at the maximum temperatures.

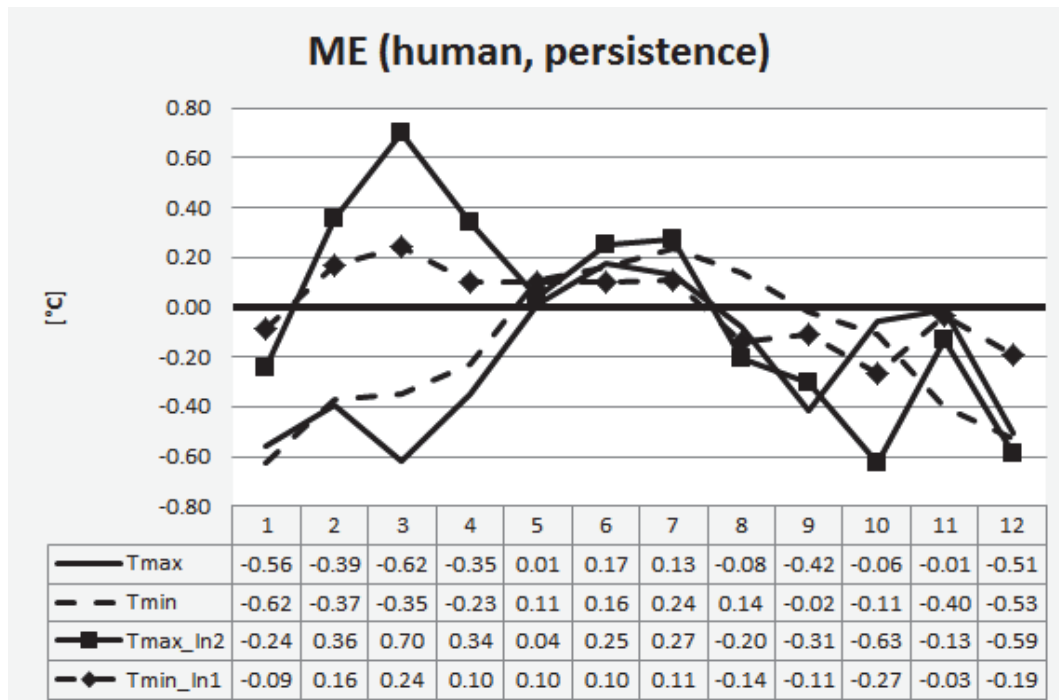


Fig. 16. Mean error for the minimum and maximum temperatures of the human and persistence forecasts through the different months.

Despite the bigger positive bias of the persistence forecast in spring (February, March, and April), which is more than 0.3 °C, it does not compensate the significant negative mean error in the human forecast. There is an overlap only in September and December, when there is a negative bias of more than 0.3 °C in the maximum temperatures. Thus, it can be stated that the persistence forecast, even by months, does not import a big systematic error in the human forecast. Or, we can say, that by months and by years, the persistence forecast does not bring a serious defect (systematic error) in the human forecast. Therefore, the systematic error in the human forecast is primarily due to the numerical models and the weather forecasters.

4. Conclusion

1. Over the years (2009–2014), the accuracy of the forecasts increases and the *PC* rises in the minimum temperatures with nearly 6%, while in the maximum temperatures with a little more than 4%.
2. The accuracy of the forecast is higher during the warm months (when dynamics of the atmosphere are lower); expressed in the *PC* this difference is nearly 20% (*Fig. 10*).
3. The accuracy of the minimum temperature forecast is greater, which is caused by smaller fluctuations (day to day), compared to the maximum temperatures (*Fig. 3*).
4. The influence of the persistence forecast on the human is great, and it is greater at the minimum temperatures.
5. Human forecast during the cold months underestimates the temperatures, while in the warm months slightly overestimates them.
6. Summarized by years, there is a negative systematic error: an underestimation of the temperatures (*Fig. 14*).
7. The systematic error in the human forecast is mainly due to the numerical models and to the forecasters.

References

- Bogatchev, A., 1988: A five-day forecast method of mean daily temperatures and daily precipitation totals. *Probl. Meteorol. Hydrol.* 4, 17–22.
- Bogatchev, A., 1994: Statistical schemes for medium-range weather forecast. Abstract of a PHD thesis.
- Brier, G.W., 1950: Verification of forecasts expressed in terms of probability. *Month. Weather Rev.* 78, 1–3.
- Brier, G.W. and Allen, R.A., 1951: Verification of weather forecasts. *Compendium of meteorology*, 841–848.
- Fajman, P., 2011: Improved verification and analysis of national weather service point forecast matrices. For the Degree of Master of Science in the Graduate College at the University of Nebraska.
- Lashley, S.L., Fisher, L., Simpson, B., Taylor, J., Weisser, S., Logsdon, J., and Lammers, A., 2008: Observing verification trends and applying a methodology to probabilistic precipitation forecasts at a national weather service forecast office. 19th Conf. on Probability and Statistics, New Orleans, LA, Amer. Meteor. Soc., 9.4. <http://ams.confex.com/ams/pdfpapers/134204.pdf>
- Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather Forecast.* 8, 281–293.
- Ripley, E. and Archibold, O., 2002: Accuracy of Canadian short-and medium-range weather forecasts, *Weather* 57, 448–457. <https://doi.org/10.1256/wea.245.01>
- Ruth, D.P., Glahn, B., Dagostaro, V., and Gilbert, K., 2009: The performance of MOS in the digital age. *Weather Forecast.* 24, 504–519. <https://doi.org/10.1175/2008WAF2222158.1>
- Spiridonov, V., 1987: Skill measures of daily extreme temperature forecast. *Probl. Meteorol. Hydrol.* 2, 19–26.

- Wilk, G.E.*, 2005: The Temperature and Precipitation Verification Program at WFO Corpus Christi Texas, United States Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service.
- Wilk, G.E.*, 2007: Temperature and Precipitation Verification Results and Interpretation at WFO Corpus Christi and Other WFOS January 2004 Through June 2006, United States Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service.
- Joint Working Group on Forecast Verification Research: <http://cawcr.gov.au/projects/verification/>