Intern rapport ; IR 2009-02

An exploratory study of the possibilities of analog postprocessing

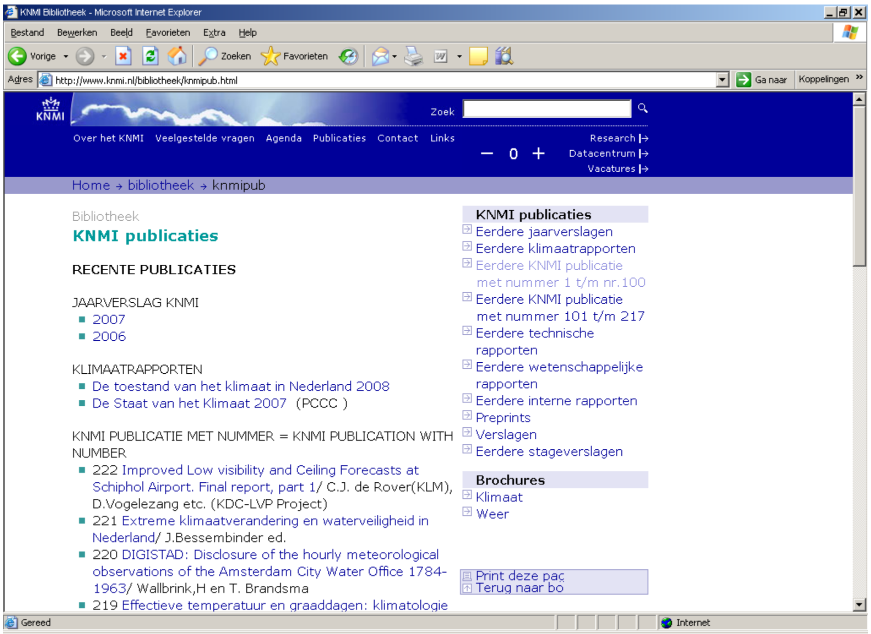
Dirk Wolters

De Bilt, 2009

De Bilt, 2009

PO Box 201
3730 AE De Bilt
Wilhelminalaan 10
De Bilt
The Netherlands
http://www.knmi.nl
Telephone +31(0)30-220 69 11
Telefax +31(0)30-221 04 07

Auteur: Wolters, D.



Please note:
All titles of KNMI-publications (and a full text PDF for the most recent ones) can be found on
http://www.knmi.nl/bibliotheek/knmipub.html

An exploratory study of the possibilities of analog postprocessing

Internship report by Dirk Wolters, Wageningen University
Supervisor: Kees Kok, Royal Netherlands Meteorological Institute

Summer 2008

Table of contents

I. Introduction

Analogs for weather forecasting

A weather situation from the past that is, according to certain fields of variables, much alike a present weather situation is called the present situation's *analog*. It is possible to derive valuable information about a present weather situation from its analogs. In short, the principle of using analogs in weather forecasting is that past observations, obtained from past analog situations, are used as a predictor for the observations in the present situation.

Traditionally, analogs have been used in weather forecasting in two ways. Firstly, analog situations have been used to directly obtain forecasts. In this approach, analogs are found for a present situation, and then observations some hours or days following these analog situations are taken as predictors for the observations some hours or days following the present situation. Research interest in this principle decreased after the emergence of physically-based numerical weather prediction (NWP) models. These models show much better skill than analog methods in forecasting weather evolution.

Secondly, another way of applying analogs to weather forecasting is to use analogs for postprocessing NWP model output. For this, forecast fields are taken from an NWP prediction, for some hours or days ahead. Then, historical analogs are identified for these predicted fields. Next, observations from these analog situations are taken as predictors for the observations in the situation forecast by the NWP model. In this way, analogs act as a link between one or more fields that have been forecast by the model, and local observations at the time of verification of these predicted fields. In this sense, analog postprocessing is a method of downscaling NWP model output. In the present study, such a system of postprocessing output from weather models with analogs will be considered.

With analog forecasting methods, it is possible to make deterministic as well as probabilistic downscaled forecasts. A deterministic predictand can be obtained by, for example, taking the observation of a variable in only the closest analog, or by taking the mean of the observations of a variable in a number of closest analogs. A probabilistic forecast can for example be obtained by regarding all observations in a number of best analogs as possible outcomes, and then constructing an ensemble of analogs.

The process of searching for analogs and using them for forecasting involves many degrees of freedom, especially in how analogs are defined. This makes it often difficult to optimize an analog forecasting method. Some of these degrees of freedom will be explored in this study, by performing sensitivity analyses.

History of analog methods

In the past, much research has been done on the application of analog methods to weather forecasting. The principle of analog forecasting has been used already before the adoption of fast computers, using the classification of large-scale generalized circulation types (as are called *Grosswetterlagen* by Baur, 1944 and Baur, 1948) to predict weather evolutions. Later, with increased computer power and increased weather archives, it became possible to numerically search weather archives for analogs for a given situation, as is also done in this study.

In the course of the nineteen seventies, when computer power had increased even more, the analog method as a candidate for predicting medium-range weather evolution became overtaken by physically-based NWP models (as described by Ross, 2005). Some other applications of the analog method however remained interesting, such as making short term probabilistic and deterministic predictions (as in the short-term fog predictions of Hansen and Riordan, 2001), long-term predictions (as in Toth, 1989), investigating atmospheric predictability (as in Toth, 1991a) and the postprocessing of NWP predictions as described above (as in Kok and Lammers, 1993, Lemcke and Kruizinga, 1988 and Kruizinga and Murphy, 1983).

At KNMI, a current application of analogs is in the MOS (Model Output Statistics) system, in which mean values of observations in series of 500 hPa geopotential height analogs work as a predictor for some variables, amongst other predictors. This procedure is applied for some deterministic as well as probabilistic predictands. See Lemcke and Kruizinga, 1988 for details on this application of analogs in MOS, and Kruizinga and Murphy, 1983 for details on the analog methodology used.

In the past, most research on analogs has been done on applications for medium- and long-range forecasts (e.g. Toth, 1989, Kerr, 1989, Ross, 2005). Therefore, analogy has been mostly defined on large (>500x500 km^2) horizontal areas. In almost all research, analogy has been defined on a single field, usually the 500hPa geopotential height field.

Some characteristics of analog postprocessing

As said before, in this study the principle of analog postprocessing of NWP model output is applied. The general routine of this analog postprocessing consists of 1) taking fields that are forecast by a numerical model, 2) searching the past and taking the closest historical analogs to these forecast fields, 3) taking past observations at one station in these analog situations as predictors for observations at the same station in the situation forecast by the numerical model.

Analog postprocessing makes three types of 'additions' to the information from a single (not ensemble) NWP model run. Firstly, the use of local observations makes this procedure a form of downscaling in space and time. The method identifies historical analogs for model output on a coarse spatial domain of grid points, and then uses past observations at a single sub-grid point as a predictor for observations at that point. Thus, model output on coarse scales is made representative for the specific observational location. By doing do, also some systematic errors in the model can be accounted for.

Secondly, analog ensembles give probabilistic information while the output of a single NWP model run does not. An NWP prediction for grid point x contains single (deterministic) values for temperature, humidity etc. An ensemble of observations from n of its best historical analogs though, contains n predictors for a future observation, that are all possible outcomes. With such an ensemble, a probability distribution can be made, which can be a useful addition to deterministic model output.

Thirdly, with analog postprocessing, information can be gained on some variables that are not, or not well represented in NWP models. For example, in present NWP models, there is no output on visibility, and forecasters must rely on existing postprocessing methods or on their own experience to predict the probability of low visibility. Applying analog postprocessing to this NWP output makes that every quantity that has an observational history, such as visibility, is a possible predictor/predictand variable.

In conventional MOS systems (without using analogs), a large number of past forecasts have been verified with observations. Empirical relations between the model output and local observations are then deduced, and used for postprocessing future forecasts. In principle, this conventional MOS postprocessing makes the same three additions to NWP output as have been mentioned above for analog postprocessing. A potential advantage of analog postprocessing, as compared to MOS, is its greater capability to cope with extremes. MOS systems are usually based on a limited number of past forecasts, typically in the order of 3 years. With analog postprocessing, analog situations are sought in an archive of 40 or 50 years, containing more extreme events. Also, the empirical relations in MOS are usually based on a bulk of situations and not on extremes, so that 'normal' situations are much better represented than extreme situations, even if a large archive of forecasts were used.

Problem definition

This research is concentrated on the possibilities of making short-term ensemble forecasts for the Netherlands, by postprocessing NWP output with an analog method. We have concentrated on short-term forecasts because at the moment, operational numerical ensemble forecast systems (like the ECMWF EPS) are designed to provide optimal probabilistic forecasts at medium range, and are less useful for short-range probability forecasts, so that other short-range ensemble systems (like analogs) might have additional value. Currently, short-range numerical EPS (SREPS) systems are being developed, and the analog postprocessing method used in this research could provide 'reference' or 'base line' ensembles in validating these new methods. We will take a special look on forecasting some extreme situations (more concretely: low visibility). This is because analog postprocessing might have an advantage over conventional MOS systems in forecasting extreme cases, as explained in the preceding section.

A straightforward application of analog postprocessing of NWP model predictions is considered. Analogs are identified for a predicted pattern, and observations from the analog situations act as predictors for the observations in this predicted situation. See the Methodology section for a more detailed description of the method. Our goal has been to make this concept into a working, verifiable system, and to use this system to investigate the possibilities of the concept. Therefore, the system has been kept as generic as possible, so that it was possible to vary the configurations of the method in several directions. For example, in our system, analogs can be taken for weighted combinations of different fields, from different parts of the year, and different numbers of best analogs can be taken (see section 'How to define analogs' in 'Methodology'). A sensitivity analysis will be carried out for some of these degrees of freedom.

As a second line of research, we have applied the method to investigate the effect of global warming on the climate in the Netherlands. Analogs have been used to study the historical change of temperature in circulation patterns that are similar. This has been done by, for many different situations, taking their best historical circulation pattern analogs, and looking at the temperature trend of these analogs throughout the archive. This temperature trend is an estimation of the temperature change of the climate that is not due to a change in the occurrence of circulation patterns. In recent research, some evidence has been raised that in the Netherlands, above the global warming of the mean, a part of the observed recent warming has been due to changes in the occurrence of 'warmer' circulation patterns (as van Oldenborgh and van Ulden (2003) made plausible by statistical means). This hypothesis makes investigating the temperature trend in circulation analogs an interesting approach to climate change.

Because it had to be completed in four months, the character of this research has been kept exploratory. Thus, the main goals were 1) to program a comprehensive system of analog postprocessing and 2) to perform some experiments with it, to explore its possibilities, both in forecasting and in investigating climate change. It was not intended to deliver a fully tuned system for analog postprocessing.

II. Methodology

Structure of the analog postprocessing method

For this study, an analog postprocessing system has been constructed, based on ECMWF forecasts, the ECMWF ERA-40 reanalysis archive, and observational data from De Bilt, The Netherlands (see section 'Used data'). The system has been programmed in the programming language Fortran 90. The structure of the system will be explained by chronologically going through a postprocessing run, looking at the scheme of the system in Figure 1. See section 'How to define analogs' for a more detailed description of possible choices and settings in the system.

To perform a run of the analog postprocessing system, first some forecast fields from a numerical weather model are needed. Then, for these fields, a number of analogs (according to certain criteria for analogy) need to be taken from the analysis archive. This is done, by comparing scalar fields of atmospheric variables from the forecast with the same fields in all days in the analysis archive. For every analysis in the archive, the 'difference' between the forecast field and the archive field, according to root mean square (RMS) of their differences in all grid points, is calculated. Of course, it is also possible to use other metrics.

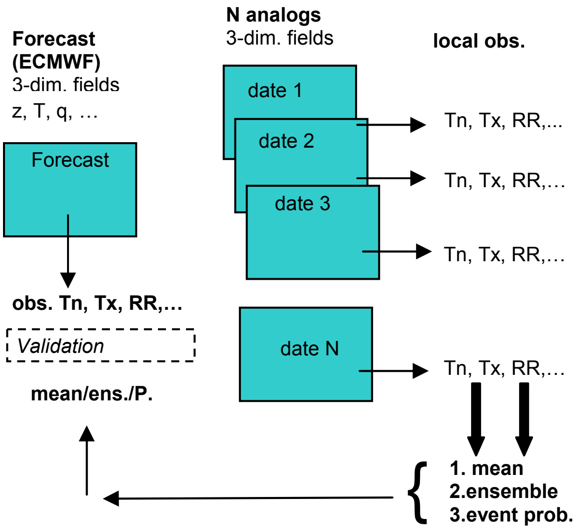


Figure 1: Schematic representation of the system

In the method used here, such differences can be calculated for up to three scalar fields separately (for example z_{1000}, T_{1000} and q_{1000}). Next, the difference values according to these three fields are combined using certain weights, see paragraph 'How to define analogs'. The situations in the archive that have the smallest weighed 'difference' value are considered best analogs.

In this way, a number of dates in the archive can be found that represent situations that are 'close' to the forecast situation under investigation, according to up to three atmospheric variables. Then, a number of closest situations are defined as the forecast situation's analogs.

Next, the observed value of the desired predictand variable is obtained from every analog situation, see the right column in Figure 1. These observational values are all predictors for the predictand in the forecast situation, together forming an ensemble. With this ensemble, deterministic and probabilistic predictions can be made.

Of course, the model forecast and archive fields that are used to find the analogs should be physically connected to the predictand. For example, for a maximum temperature analog ensemble, analogs could be taken for 850hPa temperature (the temperature above the boundary layer) and cloud cover (important for the radiation heat budget) at 12UTC. Of course, the T_x field forecast by the model, if available in the model output, is an even more obvious field of analogy to choose for these predictions.

From the ensemble of analog predictors, for example the mean can be taken as a deterministic predictand, or a probabilistic forecast can be obtained. In this study, two types of probabilistic forecasts are considered. Firstly, the ensemble of analogs can be kept as a whole, giving a complete probability distribution of the value of a certain predictand. Secondly, the occurrence of a certain event can be considered. This could for example be the occurrence of thunder, or the exceedance of 30mm of rain in 12 hours. Then, the probabilistic forecast is no longer a complete ensemble, but just the probability of a certain event occurring. The percentage of analogs in which the considered event did occur is then taken as the forecast probability.

In order to assess the skill of the analog postprocessing system, and to experiment with different configurations, of course a validation is needed as the next step. For a discription of the validation methods used here, see 'Validation methods used'.

How to define analogs

Essential to analog methods is the question of 'what is an analog?'. Analogy can be defined in many ways. There is the choice of which field(s) in analogs should resemble those in the forecast situation, the choice of how analogous a situation should be to be called an analog, how large the area of resemblance should be, and many other choices. This results in a large number of variables and settings, which need to have been specified before the computational routines can start. The degrees of freedom that are particularly relevant to the method used here will be listed and addressed below. In principle, the system that we have programmed for performing analog postprocessing contains all degrees of freedom described here.

Fields of analogy

Probably the most fundamental choice, is that of the fields used to base the analogs on. These are the fields that should resemble in order to call situations analogs, and will be called the 'fields of analogy' in the following. Of course, which field of analogy to choose, is most essentially dependent on the considered predictand. For some predictands, the choice is obvious. For example, for 2m temperature as a predictand, the 1000hPa temperature as an important field for analogs is obvious. The 1000hPa temperature contains relatively much information about the 2m temperature, compared to other fields such as wind speed, moisture etc. In most cases though, the most relevant information about the predictand outcome is divided over different fields. For example, fog can be predicted by a combination of fields such as moisture, wind speed and some others. In such cases, it is important to choose the right combination of fields to base analogy on. These fields should each contain as much information as possible about the predictand. At the same time, their mutual dependence should be small, to avoid a great overlap in information.

Spatial dimension of the fields of analogy

Once the most relevant fields of analogy are chosen, another choice is their dimension. Using a small field of analogy will result in analogs with the most location-specific information. In many cases though, the optimal domain of analogy is larger than one would expect. For example, with a very local pressure field as a field of analogy, analog fields can have strong resemblance with the forecast field, but might have only their local pressure and perhaps wind direction in common. With a larger field of analogy, more emphasis would be put on an analog synoptic situation, containing information that is not so location-specific, but relevant for the considered location as well. One must choose a spatial domain that is small enough to make predictions quite location-specific, but that still is large enough to find analogs that also match a more general pattern.

Choice of metric

In literature, most studies have used the RMS value of the difference in all points between two fields, as a metric for analogy. The smaller this value, the more resembling the analog is assumed. Of course, other metrics are also possible, such as a correlation coefficient or (as in Radinovic, 1975) the number of points in two anomaly fields that have the same sign. The choice of metric can play a significant role in forecast skill. In the study of Toth (1991b), who examined the use of different metrics for a general case of analog forecasting, the RMS worked better than the correlation coefficient. In this study, the RMS value is used in all experiments, but the use of other metrics is possible with the constructed system.

If more than one field of analogy are considered, another question is how the different fields should be relatively weighed. In one postprocessing run, the program used here constructs an array with *k* RMS values for all *k* situations in the analysis archive, for each of the three different analogy fields. Next, it standardizes these three arrays, by dividing all elements by the maximum RMS value in their array. Then, the values of the arrays 1, 2 and 3 are multiplied by a weight factor w_1, w_2 and w_3. With these weights, the relative importance of each of the three fields of analogy can be defined. Next, the program sums the three standardized and weighed RMS values for every situation in the archive, forming a new array, containing the definitive metric used to measure analogy (equation 1). Although the weights of the fields can be adjusted freely, in this study, where combined fields are used, relative weights have been kept equal.

$$metric_k = \frac{RMS1_k}{RMS1_{\max}} \cdot w_1 + \frac{RMS2_k}{RMS2_{\max}} \cdot w_2 + \frac{RMS3_k}{RMS3_{\max}} \cdot w_3 \qquad (1)$$

Archive dimensions

Intuitively, one would say that the larger the archive of potential analogs, the better. There is the important restriction though, that the archive should be more or less homogeneous. If the climate would change, analogs taken before and after the change could have a different predictive skill. Climate changes can put a limit on which parts of the archive can be used. In this study, we have briefly touched upon this, studying the effect of climate change on temperature in circulation analogs. Apart from this climate change experiment, the complete analysis archive has been used in our experiments, assuming homogeneity of the archive. In some cases this assumption is debatable.

Season

For the same reasons of homogeneity of the archive, it is an obvious choice to instruct the algorithm to not search for analogs all-year-round, but only in a part of the year. It can be expected that analogs from the same season as the forecast situation have a better predictive skill than analogs from a different season. In the analog postprocessing program used here, one has to define a 'centre' day of the year (for example January 15 for winter), and a 'season width'. The season width defines the number of days before and after the 'centre' day from which analogs may be selected.

Number of analogs

It is expected that the analog with the smallest metric value, being the closest to the forecast field, gives the best prediction. Prediction skill of analogs is expected to deteriorate with increasing metric value. Still, taking more than one analog is expected to reduce the forecast error of the analog mean, because the mean of a number of good analogs includes more information on possible outcomes, and thus is a more 'balanced' predictor than only the best analog. For probabilistic predictions, taking more than one analog is of course essential. The number of (best) analogs to use is an important parameter. Intuitively the best way of putting a limit to the number of analogs, is to only use analogs with a metric value below a certain threshold. In practice, it is often more convenient to use a limited, fixed number of analogs, allowing the metric value of the 'worst' analog to vary greatly sometimes. In the system used here, it is possible to either take the *n* analogs with the smallest metric value, or all analogs with a metric value below a fixed threshold.

Validation methods used

In order to validate the analog postprocessing method, it is needed to postprocess a large number of forecasts and examine the results. This is done by taking a large database of past ECMWF forecasts, and performing an analog postprocessing run for every forecast in this database. Then, all predictions of the analog postprocessing system are validated with observations, to examine the skill of the system. For such a cycle, the term 'validation run' will be used. One validation run consists of performing analog postprocessing on many numerical forecasts, and validation the resulting predictions. Our experimental strategy will be, that for different configurations of the analog postprocessing system, different validation runs will be carried out and their results compared.

Several statistical methods were available to use in these validations, which will be explained in the following. The two types of probabilistic forecasts described above (ensembles and probability forecasts for a certain event) of course need different approaches of validation. The validation of both types of probabilistic forecasts will be addressed separately in the following.

A. Validation of ensembles

Forecast ensembles can be seen as predicted probability density functions. In an ensemble, all members should have the same probability of being closest to the outcome. This property of ensembles can be investigated using *Talagrand diagrams*, also called *rank histograms*, as introduced by Talagrand et al. (1997). They are a commonly used tool in the investigation of ensemble forecasting techniques (e.g. Kok, 2001). A Talagrand diagram is a histogram that lists the frequencies of the predictand outcome occurring at different places within the ensemble. For every ensemble obtained by the system, the values of the 'ensemble members' are listed in a logical order (mostly from low to high values). In between the values of these ordered ensemble members, corresponding bins are defined. Next, for every ensemble, it is recorded in which bin the predictand outcome falls. Then, for every bin it can be summed and plotted how many times a predictand outcome fall into it. The resulting histogram is known as the Talagrand diagram. See Figure 2 for an example. To ease the interpretation, also a cumulative rank histogram (as introduced by Kok, 2001) can be plotted, see the blue line in Figure 2. The turquoise line in this figure represents the cumulative rank histogram in an ideal case, in which the observations have equal

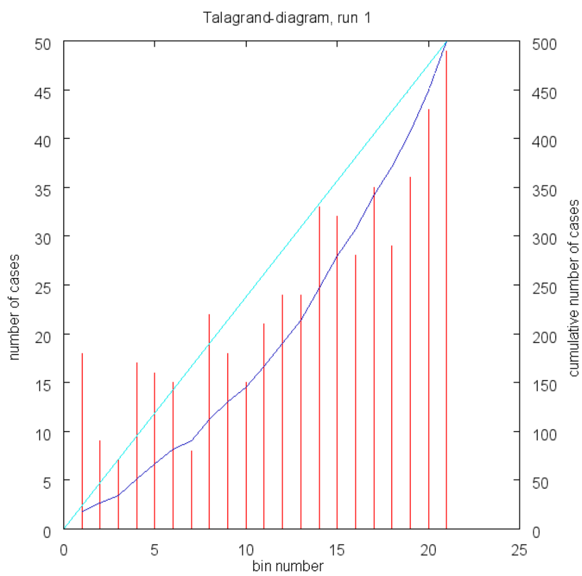


Figure 2: Example of a Talagrand diagram for 20 analogs (21 bins)

chance of occurring. In this case, all Talagrand frequencies would approach an equal value, and the cumulative rank histogram would approach a straight, diagonal line.

Various properties of the ensemble can be deduced from Talagrand diagrams. From the Talagrand diagram in Figure 2 for example, it becomes clear that many more observations fell in the high-numbered bins than in the low-numbered bins. This means that the ensemble systematically underestimated the predicted quantity.

A flat Talagrand diagram is not the only condition for a good ensemble. For example, if we take the climatological statistics of the date or season (mean, standard deviation etc.), and generate an ensemble of random forecasts from it, then this results in ensembles with a perfectly flat Talagrand diagram, that still are very poor predictors compared to modern NWP ensembles. Using random numbers with seasonal climatology gives a very wide ensemble, while an ideal ensemble is narrow.

A good measure for this 'wideness' of a forecast ensemble, is the *mean absolute error* (*MAE*), of the ensemble members and of the ensemble means. The MAE of ensemble members and ensemble means should be as small as possible. Another measure that provides useful information about the accuracy of a system, is the mean error, or bias (systematic error) of the ensemble members. Of course, it is important to have a bias that is close to zero, as a (strongly) biased predictor is not desirable. The suspicion of a bias also follows from a strongly non-balanced Talagrand diagram as shown in Figure 2.

B. Validation of event probability predictions

If we are investigating forecasts of probabilities of a certain event taking place, we do not have ensembles, but just a number of forecast probabilities to validate. Each forecast probability can be validated by a dichotomous occurrence, having the value of 0 (no occurrence) or 1 (occurrence).

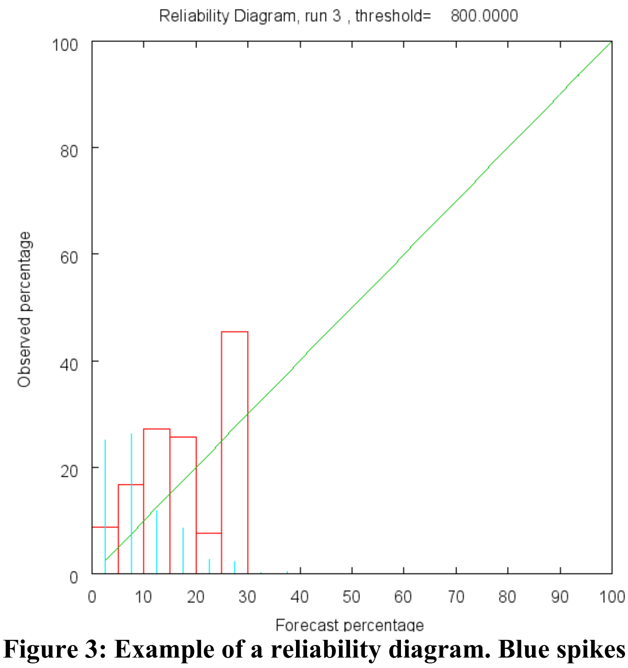


Figure 3: Example of a reliability diagram. Blue spikes indicate relative sample size

A method for the validation of this kind of forecasts, somewhat equivalent to Talagrand diagrams for ensembles, is the *reliability diagram*. See the example in Figure 3. A reliability diagram is a binned representation of all pairs of observed (0 or 1) and forecast (0-100%) values. Observations are binned with respect to their corresponding forecast probabilities, from low to high forecast probability, with a certain bin width (in Figure 3 the bin width is 5%). In every bin, the mean is taken of all (0 and 1) observed values, and converted to a percentage. This then results in mean observed percentages for a number of ranges of forecast percentages. These observed percentages are plotted in Figure 3 as red bars. Ideally, of course, for a selection of probability forecasts with around equal forecast probability, the outcome as an observed percentage is equal to the forecast probability. Thus, an ideal reliability diagram has a slope of 1, indicated by the green line in Figure 3. The relative number of cases per bin (the turquoise spikes in Figure 3) can act as a help to consider the significance of observed percentages.

In many cases very high or low forecast probabilities do not occur, because of the limited skill of the forecast method. This is particularly the case with analog methods. The best analog method produces lists with analogs in which the considered event always happened (forecast 100%) for all cases in which it actually happened (occurrence 1) and lists with analogs in which the event never happened (forecast 0%) for all cases in which the event did not happen (occurrence 0). These perfect forecasts would result in a reliability diagram with 0% observed in the 0-x% forecast bin, and 100% observed in the y-100% forecast bin, and nothing in between. In reality, lists with analogs will almost always be a mix of occurrence and non-occurrence. If a rare type of event is considered, it is expected that it will not occur in many of the analogs, so that it is hard for the system to forecast a high probability.

Reliability diagrams provide much information on how well the predicted probabilities of an event correspond to their observed frequencies. It is useful though, to use with these rather subjective diagrams an objective score for the forecast skill as well. For probabilistic forecasts for dichotomous events, as we are considering here, the most well-known score is the *Brier Score* (*BS*):

$$BS = \frac{1}{n}\sum_{i=1}^{n}\left(p_i - o_i\right)^2$$

The Brier score is the mean squared error of *n* forecasts, in which the forecast probability p_i is a value between 0 and 1, and the observed value o_i has value 1 for occurrence, and 0 for non-occurrence of the considered event. BS=0 means that the forecast is always perfect (only 100% forecast probabilities if the event occurred and 0% forecasts if not), BS=1 meaning the opposite.

A disadvantage of the Brier score is that it depends much on the climatological occurrence of the predicted phenomenon. For predictions of very rare events, only very small Brier scores will be seen, while for predictions of common events, Brier scores usually have larger values. To obtain score values that have a more absolute meaning, usually the *Brier skill score* (*BSS*) is used, which is a measure of the Brier score of the considered forecast, relative to the Brier score of a reference type of forecasts, usually forecasts that use only the climatology of the phenomenon (obtained by simply inserting the climatological probability for p_i in the *BS* formula above). The BSS:

$$BSS = 1 - \frac{BS}{BS_{ref}}$$

A positive value of *BSS* ($BS<BS_{ref}$) in general indicates that forecasts are better than the reference forecasts, and thus skilful. According to Mason (2004) the Brier skill score is a harsh score. Negative Brier skill scores can occur while the forecasting system does contain skilful information; Brier skill scores can ‘hide’ useful information present in the forecasts. Therefore, Mason suggests to not use it as a lone measure for forecast skill, to prevent forecasting methods with ‘bad’ Brier skill scores from being rejected unjustly.

Data

Three distinct data sources have been used in this investigation. The archive from which analogs are taken is the ECMWF ERA-40 reanalysis archive. Spatially, only part of this archive was available for this investigation, being the section between 20^o W and 20^o E, and between 40^o N and 60^0 N, on the levels of 1000, 925, 850, 700, 500 and 250 hPa. The time period from December 1, 1957 until December 29, 2007 has been used, with 0, 6, 12 and 18 UTC fields for each day. Considering no selection for season, in total 18292 archive days are available for an analog postprocessing run. The dataset has a horizontal resolution of 1^ox1^o. Available field variables are: geopotential height, temperature, moisture content, and zonal, meridional and vertical wind components. Additionally, a few derived sets of these fields (for example wind speed deduced from horizontal wind components) have been used.

Secondly, an ECMWF forecast archive has been used, to provide the forecasts to be fed into the analog postprocessing system. This forecast archive has the same resolution and field variables as the reanalysis archive, and covers the time period from September 1, 2002 December 31, 2007 (1944 days). It contains all forecasts initialized on 00UTC and on 12UTC. Both archive and forecast fields were available in binary GRIB format.

Lastly, historical observations from weather station De Bilt (WMO 6260) have been used, both to find predictor values for the analogs, and to validate their predictions. These data on maximum and minimum temperature, rainfall and visibility were supplied by the KNMI climatological department.

Software

A part of this investigation has been to develop the necessary software for applying the analog postprocessing method. All software has been written on a Unix workstation, in the programming language Fortran 90. It is not relevant to discuss the internal structure of the written programs here, but the approximate structure of the system consists of three main programs. The first program decodes the binary GRIB files using ECMWF Gribex libraries, and writes them into ASCII formatted files. The observational data from De Bilt are included in these ASCII files.

The second program, which is the actual analog postprocessing routine, uses this ASCII input, as it would be too slow if it had to decode the GRIB files by itself. This program first loads all data fields into the computer memory, and then searches for analogs, using a set of configurations, see under 'How to define analogs'. Analogs are sought for all forecasts, unless specified otherwise. Then, observational data at the time of the analog situations are linked to the analogs, acting as predictors. After this, prediction errors are calculated, and all results are written into output files.

The third program again loads these output files, and provides graphical output for validation of the predictions, automatically generating Talagrand diagrams, reliability diagrams, MAE and bias diagrams, metric vs. mean error scatter plots and error distribution plots. The program generates these plots using the external program GNUplot.

III. Description of the experiments

Three types of experiments are carried out, which will be explained successively in this section. Results of these experiments will be displayed and discussed in section IV. First, a sensitivity analyses on the most important degrees of freedom of the analog postprocessing method is done. An estimated optimal configuration of the system is derived from these experiments, and used in the other experiments. The next experiment investigates the effect of climate change on circulation analogs, attempting to estimate the magnitude of the part of the recent climate warming that is due to a warming of the global mean temperature. After this, a validation is done of an application of the analog postprocessing system to make probabilistic predictions of visibility. An overview of all configurations of the system for the different experiments can be found in Table 1, at the end of this section.

A. Sensitivity experiments

A comparison of validation results for different configurations of spatial analogy domain, season and number of analogs used will be made. These experiments must be seen as examples, without the objective to perform a complete analysis of what is the effective optimal combination of settings. Such a complete analysis was not achievable because so many degrees of freedom are involved that it would take too much computer power to optimize the system. Nevertheless, an estimated 'optimal' configuration will be derived, for use in the other two experiments. In this way, we can obtain a reasonable basis for the configuration of the system in the other two experiments, while still recognizing that too little is known to obtain a definitive optimal configuration.

The subject of this validation will be analog ensemble predictions of maximum temperature (T_x), with analogs that will be defined on the 1000hPa temperature field. This type of analog postprocessing has been chosen because the dependence of T_x on T_{1000} fields is relatively simple, so that the results of these experiments can act as a general example. It is likely though, that other types of predictands than T_x would show a different sensitivity to the same configurations.

The criteria for validation will be the mean absolute error of forecasts made by all analogs, and the mean absolute error of all analog ensemble means.

First (experiment A1), an attempt will be made to investigate the sensitivity for the dimensions of the analogy domain. To start with, the latitude range will be been set constant at 50-54N, and the longitude range will be varied for different validation runs. From the results of this validation then, an 'optimal' longitude range for the latitude range of 50-54N can be selected. Next, the longitude range will be set constant at this 'optimal' range, and the latitude range will be varied. An estimated optimal latitude range can then be derived.

Then (experiment A2), the sensitivity to the metric value of the analogs will be considered. A validation run will be performed taking the 500 best analogs, and a scatter plot is shown of metric value of the analogs and their forecast error, for all analogs. Mean forecast errors are calculated for different metric values.

Next (experiment A3), the sensitivity for the season from which forecasts and analogs are taken is investigated. In 4 different validation runs, forecasts and analogs will be taken from winter, spring, summer and autumn separately.

In all sensitivity experiments one configuration is varied, while the others are kept fixed. The fixed configurations will be kept at their values found more or less ‘optimal’ in the other experiments. As mentioned before, the resulting combination of ‘optimal’ settings can not be seen as a real optimum, as far too little experiments have been carried out to know enough about all combinations of degrees of freedom. The goal has not been to find an optimal configuration of the system, but to gain information on its sensitivity, and to find a ‘reasonable’ configuration for the other experiments.

B. Investigating climate change in winter with z_{1000} spatial anomaly analogs

We have not only used the analog method for predictions, but also for investigating changes in climate. The goal of this experiment was to find out if throughout the archive, there has been a systematic change of temperature in situations that had the same circulation type. In this experiment, we have confined ourselves to only looking at winter situations.

In this experiment, we have taken analogs for circulation pattern, and have studied their temperature inhomogeneity. In this way, the temperature trend in weather situations that have a similar circulation pattern is studied. This is an interesting approach to climate change, as it gives the opportunity to study the temperature trend with much of the effects of circulation on temperature, like longer-term weather oscillations and (climatic) changes in mean circulation left out. An estimate can thus be made of the temperature change that is not due to circulation changes, which can then be compared to the total observed temperature change.

To do this, analogs have been taken for the 1000hPa geopotential height (z_{1000}) spatial anomaly field, representing the circulation pattern. This field is generated by taking the mean of a z_{1000} field, and then subtracting this mean of all values in the field. We have looked at the maximum (T_x) and minimum (T_n) daily temperature in these circulation pattern analogs. To have a better estimation, the circulation analogs have not been matched to forecast fields, as in the other experiments, but for all analyses in the last seven years of the archive. We have calculated the mean deviation of T_x and T_n in these analyses from their values in their circulation analogs. This mean deviation has been calculated for several validation runs, in which analogs are taken from different historical parts of the archive. The trend of this mean difference (bias) of T_x and T_n in circulation analogs throughout the archive then, is an approximation of the trend in mean temperature with similar circulation types (as a mean trend for all circulation types together). These differences have been calculated using different ‘runs’, the first run taking analogs from only the first years in the archive, the last run using only the most recent years, and the other runs using parts of the archive in between.

To ensure that all analogs are ‘good’ circulation analogs, a small number of best analogs are taken for every situation. Results are compared with results obtained by taking a tenfold larger number of analogs, to investigate if the choice for a small number of analogs did affect the results. The results of ‘warming-per-circulation-pattern’ are compared to the actual observed warming, to see if there are any differences. Also, a similar experiment is done taking so many analogs that every time almost all situations in the part of the archive considered are analogs. The results of this last experiment should show a warming close to the actual warming, as when taking all situations as analogs, no longer the ‘warming-per-circulation-pattern’ but the total observed warming would be considered.

C. Analog postprocessing for the prediction of visibility

A validation has been performed of an application of the analog postprocessing system to the prediction of visibility. Analog postprocessing can be applied to all predictands that have a long observational history, but it is expected that the skill of the method varies greatly between predictands. We have chosen visibility as a predictand, because it is not represented in direct model output, and because it is a difficult predictand for all present forecasting systems. Visibility is especially hard to predict in extreme cases, which makes it interesting to try the analog method as a system for predicting it. Almost all low visibility is caused by fog. Different factors are involved in causing dense fog, and thus, an analog postprocessing system for fog has to combine different fields of analogy. For these fields of analogy, we have chosen the square root of wind speed, temperature and absolute air humidity. All fields are taken at the 1000hPa level, which is nearest to the level where fog occurs. Wind speed is relevant because fog can only form in quiet weather with little vertical mixing, and we have taken the square root of it to get more sensitivity to low wind speeds. Temperature is relevant because fog in the Netherlands almost always forms in air that gets saturated by (radiative) cooling. Humidity is relevant because more water vapour means that less cooling is needed to bring air into a saturated state.

The choice of visibility as a predictand should be seen as arbitrary; it is an example of how a system for analog postprocessing could work. Visibility is expected to be a suitable predictand for this type of forecasting, but the forecasts could show better (or worse) skill for other predictands. Also, the analog postprocessing system for predicting visibility will not be optimized thoroughly, so that results might be better for other combinations of settings than the one used here.

The domain of analogy, the number of analogs and the season as configurations for these visibility predictions have been taken as found more or less optimal in the sensitivity experiments described earlier. First, a validation of visibility predictions performed by the system will be done, with results shown as reliability diagrams of probability predictions with different thresholds. In these predictions, the three used fields of analogy will be combined using equal weights. Then, the Brier skill score of these predictions will be compared to Brier skill scores of the system when analogs are taken using only one of the three fields of analogy.

See Table 1 for a schematic representation of different settings in all experiments explained in the above.

Table 1: Settings for all experiments

	A1.1	A1.2	A2	A3	B	C1	C2
Predictand	T_x	T_x	T_x	T_x	T_x,T_n	VV_{min} 00-06UTC	VV_{min} 00-06UTC
analogy field(s)	*T1000*	*T1000*	*T1000*	*T1000*	$z1000_{,anomaly}$	*T1000,q1000, FF1000*	***varied***
archive time	*12UTC*	*12UTC*	*12UTC*	*12UTC*	*12UTC*	*06UTC*	*06UTC*
Fc initialisation time	*12UTC*	*12UTC*	*12UTC*	*12UTC*	-	*12UTC*	*12UTC*
Fc lead time	*+24h*	*+24h*	*+24h*	*+24h*	-	*+18h*	*+18h*
Number of forecasts	*486*	*486*	*486*	*486*	*2555 analyses*	*486*	*486*
part of archive used	*1/12/1957-29/12/2007*	*1/12/1957-29/12/2007*	*1/12/1957-29/12/2007*	*1/12/1957-29/12/2007*	***Varying 7 years***	*1/12/1957-29/12/2007*	*1/12/1957-29/12/2007*
Number of best analogs taken	*20*	*20*	*500*	*20*	*2,20,500*	*30*	*30*
Season width	*2x45 days*	*2x45 days*	*2x45 days*	*2x45 days*	*2x45 days*	*2x100 days*	*2x100 days*
'centre' day of season	*Jan. 15*	*Jan. 15*	*Jan. 15*	***varied***	*Jan. 15*	*Jan. 15*	*Jan. 15*
Lat. domain of analogy	*50-54N*	***varied***	*50-54N*	*50-54N*	*50-54N*	*50-54N*	*50-54N*
Lon. domain of analogy	***varied***	*3-7E*	*3-7E*	*3-7E*	*3-7E*	*3-7E*	*3-7E*
metric	*RMS*	*RMS*	*RMS*	*RMS*	*RMS*	*RMS (eq.1)*	*RMS (eq.1)*

IV. Results and discussion

A. Sensitivity experiments

See the 'A' columns in Table 1 for the configurations of the system that were used in these experiments. The predictand, analogy field, archive time, forecast initialisation time, forecast lead time, number of forecasts, season width and the part of the archive used can be regarded as fixed 'experimental conditions' for these sensitivity experiments. In all sensitivity experiments, T_x is predicted based on T_{1000} analogs.

The configurations for number of best analogs taken, 'centre' day of the season and lat./lon. domain of analogy are taken as has been found in these experiments as more or less 'optimal'. In all sensitivity experiments, one of these will be varied, keeping the others constant.

A1.1. Sensitivity to longitude domain

In this experiment, the longitude range has been varied. See Table 2 for the different latitude/longitude domains used to investigate the sensitivity to longitude range. For the results with these different domains, see Figure 4. Shown in this figure are the MAEs of all separate analogs and of all analog ensemble means, in validation runs with the domains specified in Table 2.

Table 2: Latitude and longitude ranges for different validation runs

run #	1	2	3	4	5	6	7	8
lat. Range	*50-54 N*	*50-54 N*	*50-54 N*	*50-54 N*	*50-54 N*	*50-54 N*	*50-54 N*	*50-54 N*
lon. Range	*5 E*	*4-6 E*	*3-7 E*	*2-8 E*	*1-9 E*	*0-10 E*	*-5-15 E*	*-10-20 E*

For separate analogs, run #4 gives the smallest MAE, and for the analog ensemble means, run #3 performs best. We have taken run #3, using the longitude range of 3-7 E, as the 'best' run. Note that differences in MAEs between the different validation runs do not exceed 0.12K for separate analogs, and 0.055K for analog ensemble means. These differences are extremely small compared to mean error differences between other +24h temperature forecasting methods, so that these results might not be very significant. A possible cause for this is that these T_{1000} analogs might not be relevant enough for T_x forecasts, causing relatively bad skill of the method for all domains, without much difference in skill between domains. The choice of run #3 as the 'optimal' run is therefore rather arbitrary.

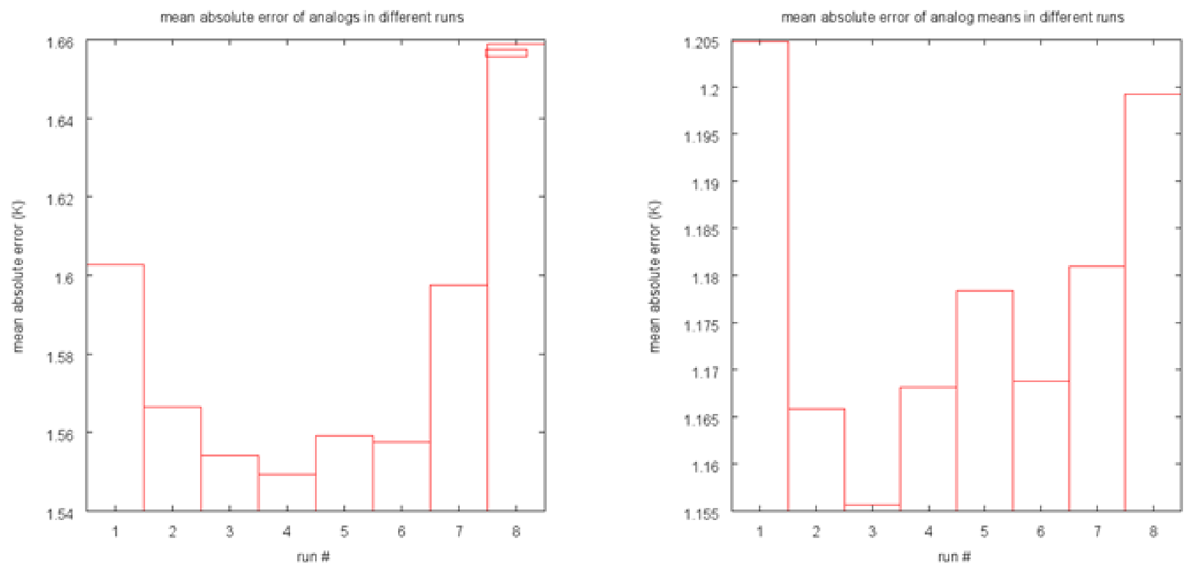


Figure 4: MAEs of analogs and of analog ensemble means, for the different longitude ranges given in Table 2

A1.2. Sensitivity to latitude domain

See Table 3 , for different latitude/longitude domains, used to investigate the sensitivity to latitude range. In this experiment, the longitude range has been kept constant at 3-7E, and the latitude range has been varied. For the mean absolute errors in this experiment, see Figure 5.

Table 3: Latitude and longitude ranges for different validation runs

run #	1	2	3	4	5	6	7	8	9
lat. Range	52-52 N	51-53 N	50-54 N	49-55 N	48-56 N	47-57 N	46-58 N	45-59 N	44-60N
lon. Range	3-7 E	3-7 E	3-7 E	3-7 E	3-7 E	3-7 E	3-7 E	3-7 E	3-7 E

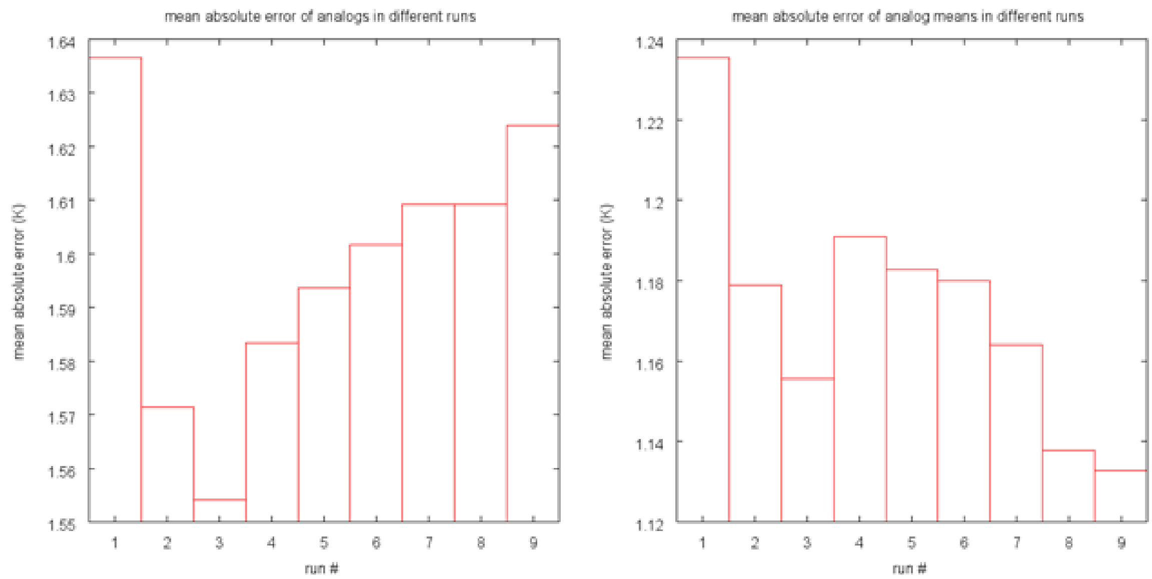


Figure 5: Mean absolute errors of analogs and of analog ensemble means, for the different latitude ranges given in Table 3

We can see that for separate analogs run #3 (latitude range 50-54 N) gives the smallest mean absolute error. For analog ensemble means, run #3 results in a 'local minimum' of mean absolute error, with the very wide latitude ranges in run #8 and run #9 giving a slightly smaller mean absolute error. Run #3 is taken as the 'best' run. Again, differences between mean absolute error are very small between runs, so that the significance of these results is presumably low.

Summarizing, a latitude range of 3-7 E combined with a longitude range of 50-54 N seems a more or less optimal choice for the situation and conditions in these experiments. The significance of this conclusion however remains doubtful, as relative differences between mean absolute errors between the validation runs performed were minimal. Intuitively though, an analogy domain of $4^o x 4^o$ (around 300km south-north and 450km west-east) seems a physically reasonable choice for the forecast range of +24h, giving analogs that have a local analogy as well as analogy on a regional pattern.

A2. Sensitivity to the metric value of analogs

In this experiment, a validation run is performed taking 500 analogs for every forecast. The total number of possible analogs in this experiment is 630, but for technical reasons, only a maximum of 500 could be taken. See Figure 6, for a scatter plot of metric value (RMS difference between mdoel forecast and analog) vs. absolute forecast error, for all 1944x500 analogs in this validation run. Because the cloud of points is too dense to interpret, bars have been plotted in the graph, giving mean absolute errors in 0.01 (-) wide bins of metric value. It is visible that there is a gradual increase in absolute forecast error with increasing metric value, from $\overline{\varepsilon} \approx 1.50$K between metric value of 0-0.01 , until $\overline{\varepsilon} \approx 3.77$ K for all metric values larger than 0.15. This is what we would expect, as analogs with a larger metric value are expected to give worse predictions, because their resemblance to the field forecast by the model is worse. From results not shown here, it appears that in order to obtain analogs that have a relatively 'skilful' metric value according to Figure 6, taking the 20 best analogs (0.44% of the whole archive) as we have done in the other sensitivity experiments, is a safe choice. Taking less than 20 analogs might hamper the reliability of probabilistic predictions.

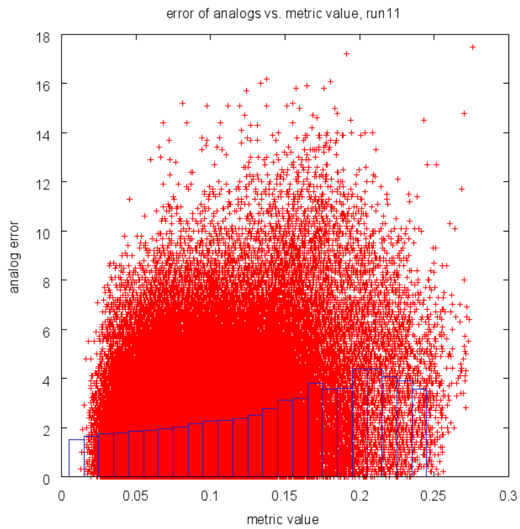


Figure 6: Metric value vs. absolute analog error, with binned means

A3. Sensitivity to the season

As described earlier, in the analog postprocessing program, a boundary can be put on the part of the year from which analogs are taken. In the previous experiments, we have looked at the winter season, taking analogs from the period between 45 days before and after January 15. We can also look at other seasons, and compare forecast skills for all four seasons. See Table 4 for the 'centre' day of the year numbers that have been used in this experiment. For a given 'centre' day number, all forecasts and all analogs are taken from the same season, from 45 days before until 45 days after the 'centre' day. See Figure 7 for graphs of the mean absolute errors for the different runs.

Table 4: Season 'centre' day numbers for different validation runs

centre day	15	105	195	285
run #	1	2	3	4

Large differences between seasons are apparent in Figure 7. It seems that winter is the season with the best results for T_x analog predictions based on T_{1000}. Spring predictions seem much worse, with MAEs roughly 0.5K larger than in winter. Summer and autumn results seem to lie in between. Differences between seasons are quite large and might well be significant, but we do not have an explanation for this.

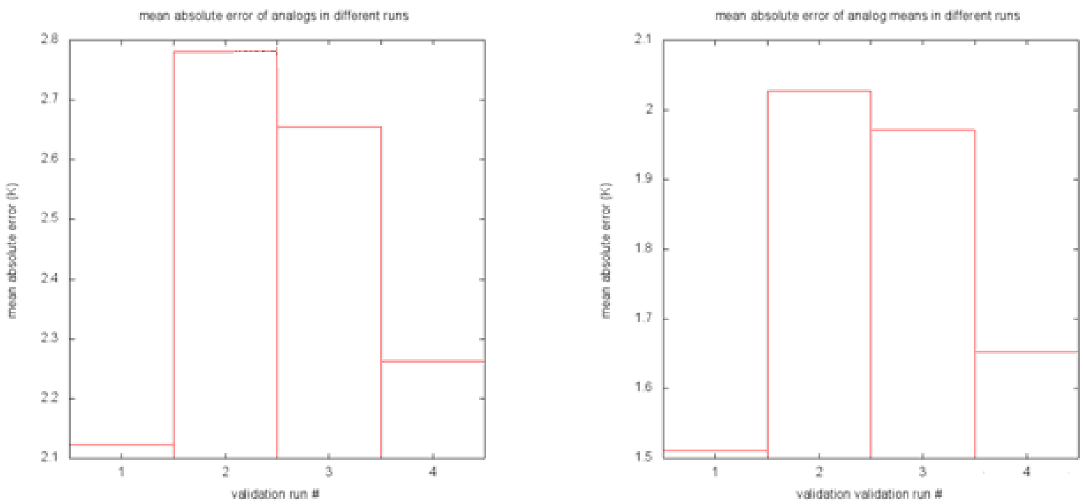


Figure 7: Mean absolute errors of analogs and of analog ensemble means, for different seasons

B. Investigating climate change with z_{1000} spatial anomaly analogs

To investigate the effect of climate change on temperature for similar circulation patterns, several validation runs have been performed. In each run, we have investigated the temperature in circulation analogs in a different temporal subset of the archive. We have looked at T_x and T_n in z_{1000} spatial anomaly analogs (see Table 1 for all other settings). In this experiment, to ensure that all analogs are good circulation analogs, for each 'prediction' only the 2 best analogs have been taken.

In every validation run, analogs have been taken for all analyses in the last 7 years of the archive, not for forecasts as in the previous experiments. Taking analogs for forecasts would include the forecast error of the model into the results, which is not desirable. In every validation run, analogs were taken from a different, 7 year-long part of the archive. See Table 5 for the different ranges of years that were used to take analogs from. In every run, only winter days have been used. To prevent the system from taking the considered analyses themselves as analogs in run #7, the analysis date itself was removed from the archive.

Table 5: Archive periods used in every validation run in the climate change experiment

run #	1	2	3	4	5	6	7
archive	*1959-1965*	*1966-1972*	*1973-1979*	*1980-1986*	*1987-1993*	*1994-2000*	*2001-2007*

See Figure 8 for the mean errors in the different validation runs for T_x and T_n. As expected, a warming trend is visible in both graphs, which is of a similar magnitude for T_x and T_n. Doing a rough estimation of the trend, we can say that there has been a warming in circulation analogs of around 1-1.5K in T_x and T_n.

We have also plotted the means of the measured T_x and T_n in the archive periods considered in the experiment, see Figure 9. These are the means of all T_x and T_n in the archive periods, as deviations from their value in the last archive period. Comparing Figure 8 and Figure 9, the first thing that becomes clear is their similarity in pattern, with a distinct warming in T_x and T_n. But also some differences appear. We can see that the total observed warming in Figure 9 seems larger (around 1.5-2K) than the warming of circulation analogs observed in Figure 8 (around 1-1.5K). It seems that especially in the period after 1986 the observed T_x and T_n are larger than T_x and T_n in circulation analogs. This 'surplus' warming observed can be seen as a part that is not 'explained' by a mean warming for all circulation patterns. A possible explanation for this 'surplus' warming is that there has been an increase in the occurrence of 'warmer' wind circulation patterns, for example as a (long-term) weather effects oscillation or as a side-effect of global warming. For The Netherlands, evidence for changes in mean circulation caused by global warming has been raised in earlier studies, for example in the statistical analysis by van Oldenborgh and Van Ulden (2003).

In Figure 10, results are shown from the same experiment for T_x, now taking 20 respectively 500 analogs for each 'prediction' (the total number of possible analogs in this experiment is 630, but for technical reasons, only a maximum of 500 could be taken). Taking the 20 best analogs, tenfold more than in the experiment above, results in a warming pattern in the analogs that is very similar to that in Figure 8. This indicates that the choice of 2 analogs per forecast gives significant results, and also, that the 20^{th} analog might still be a good circulation analog. Taking 500 analogs (the right graph in Figure 10) shows a warming that seems stronger than with 2 or 20 analogs, especially after 1986. This pattern looks much like that of the observed warming in Figure 9. This can be explained by the notion that this experiment shows the T_x deviation in almost all situations in the archive fractions, not just in the close circulation analogs. The resulting pattern is then indeed likely to look much like that of the observed mean temperatures in Figure 9.

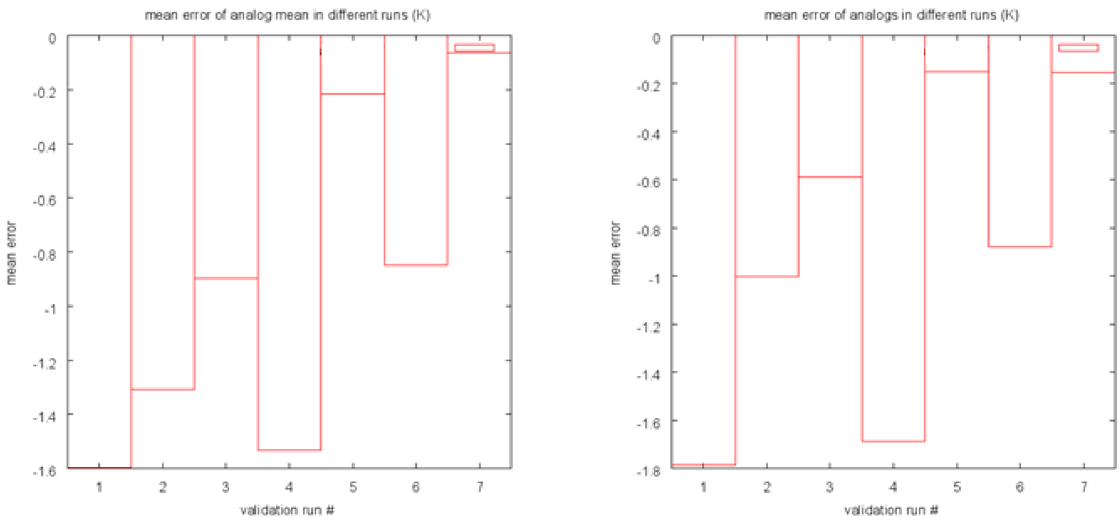


Figure 8: Mean errors of circulation analogs from the different archive periods given in Table 5, for T_x(left) and T_n (right)

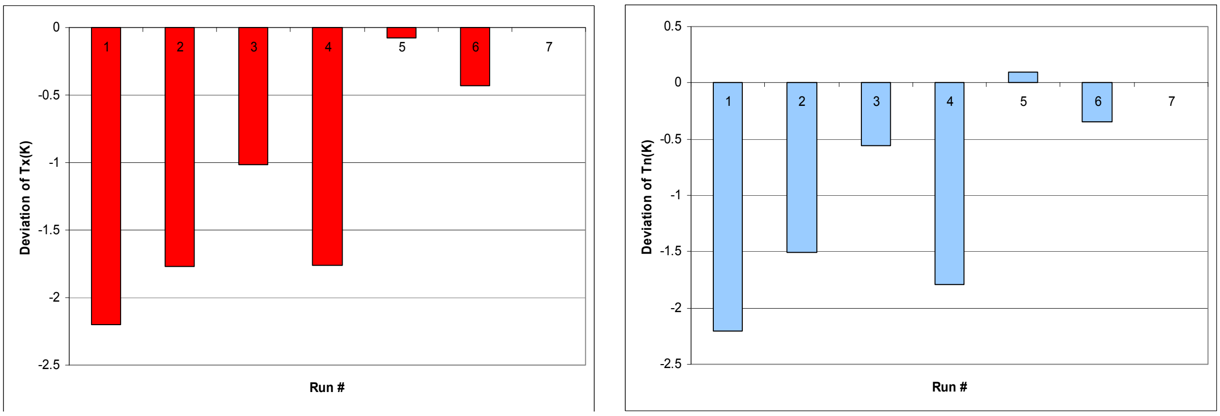


Figure 9: Mean winter T_x and T_n in the archive periods given in Table 5, as deviations from the mean of the last archive fraction

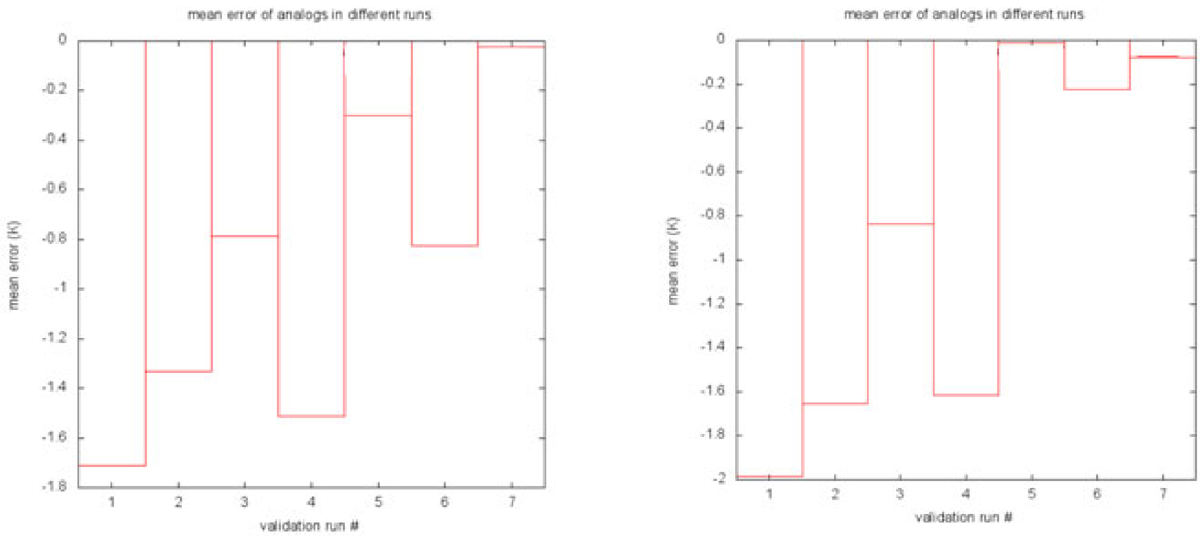


Figure 10: Mean errors of circulation analogs from the different archive periods given in Table 5, for T_x, with 20 analogs (left) and 500 analogs (right)

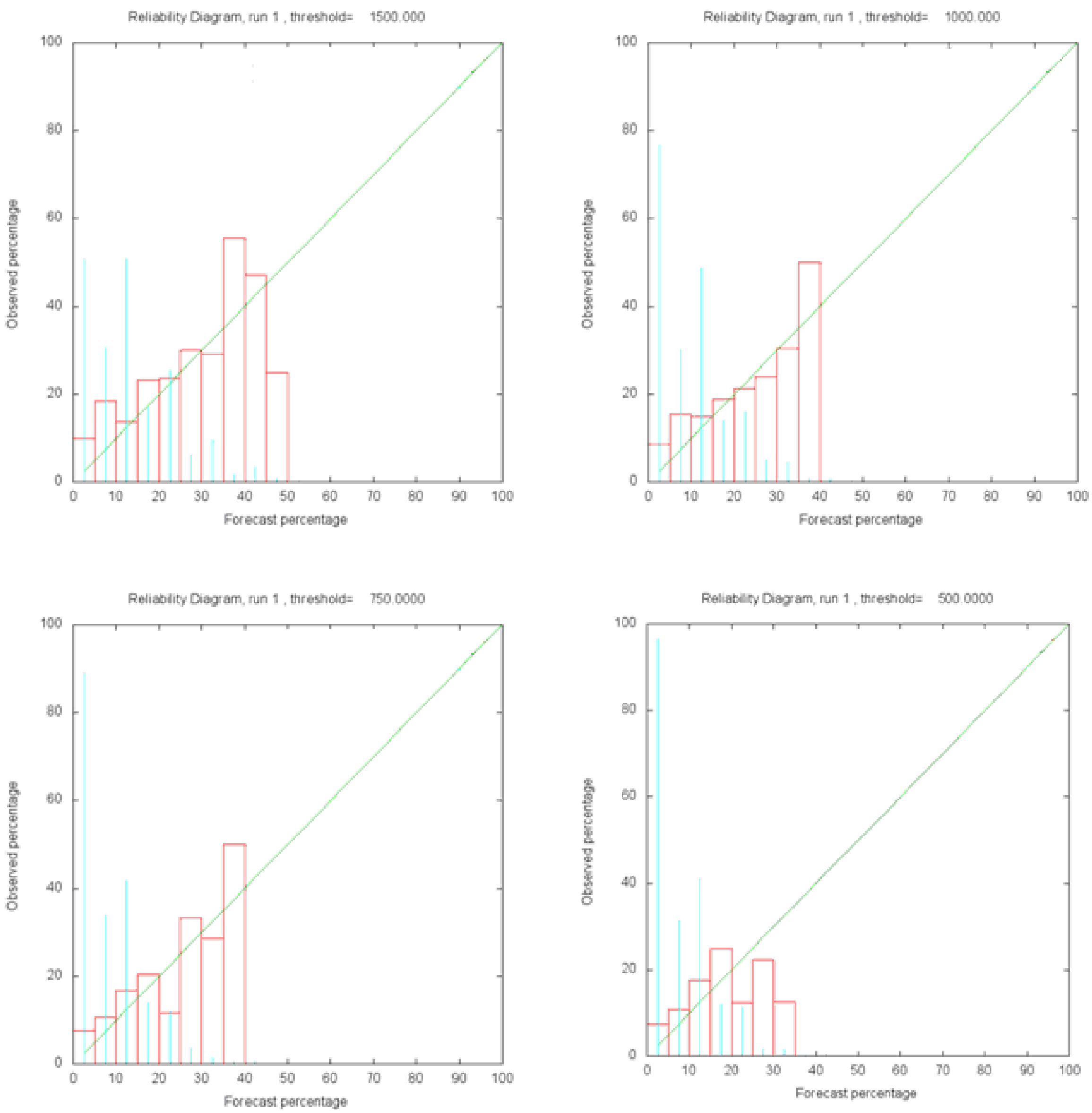
C. Analog postprocessing for the prediction of visibility

In the following experiments we have looked at a practical application of analog postprocessing, and investigated the skill of visibility predictions. We have made predictions of the probability of minimum visibility between 00UTC and 06UTC dropping below a certain threshold. Analogs have been taken for a combination of temperature, air moisture content and the square root of the wind speed, all at 1000hPa. Analog postprocessing has been applied to +18h forecast fields from forecasts initialized on 12UTC on the previous day. See Table 1 for all other configurations of this experiment. The system has been validated using reliability diagrams, and Brier skill score. As reference forecasts in the calculation of the Brier skill scores, we have used the climatological probability of visibility dropping below the considered threshold in the considered part of the year (a 'seasonal climatological probability').

An 'optimal' configuration has been derived earlier, from results of the sensitivity analyses with z_{1000} analogs for T_x forecasts. The same spatial domain of analogy has been used, and only situations in the winter part of the year are considered. For visibility, taking a time domain centered in winter is important, since in the Netherlands dense fog almost only occurs in the winter half-year. A season width has been taken of 100 days before and after January 15. The number of analogs has been put at 30, as predicted probabilities with 30 analogs can be binned in 0-5%, 5-10% etc. bins more properly than with 20 analogs. The relative weights of the fields of analogy (see 'How to define analogs') have been kept equal for the three fields T, q and √ff.

C1. Reliability diagrams

In Figure 11 reliability diagrams are shown of the analog postprocessing system for visibility. Reliability diagrams are shown for probability forecasts for different visibility thresholds.



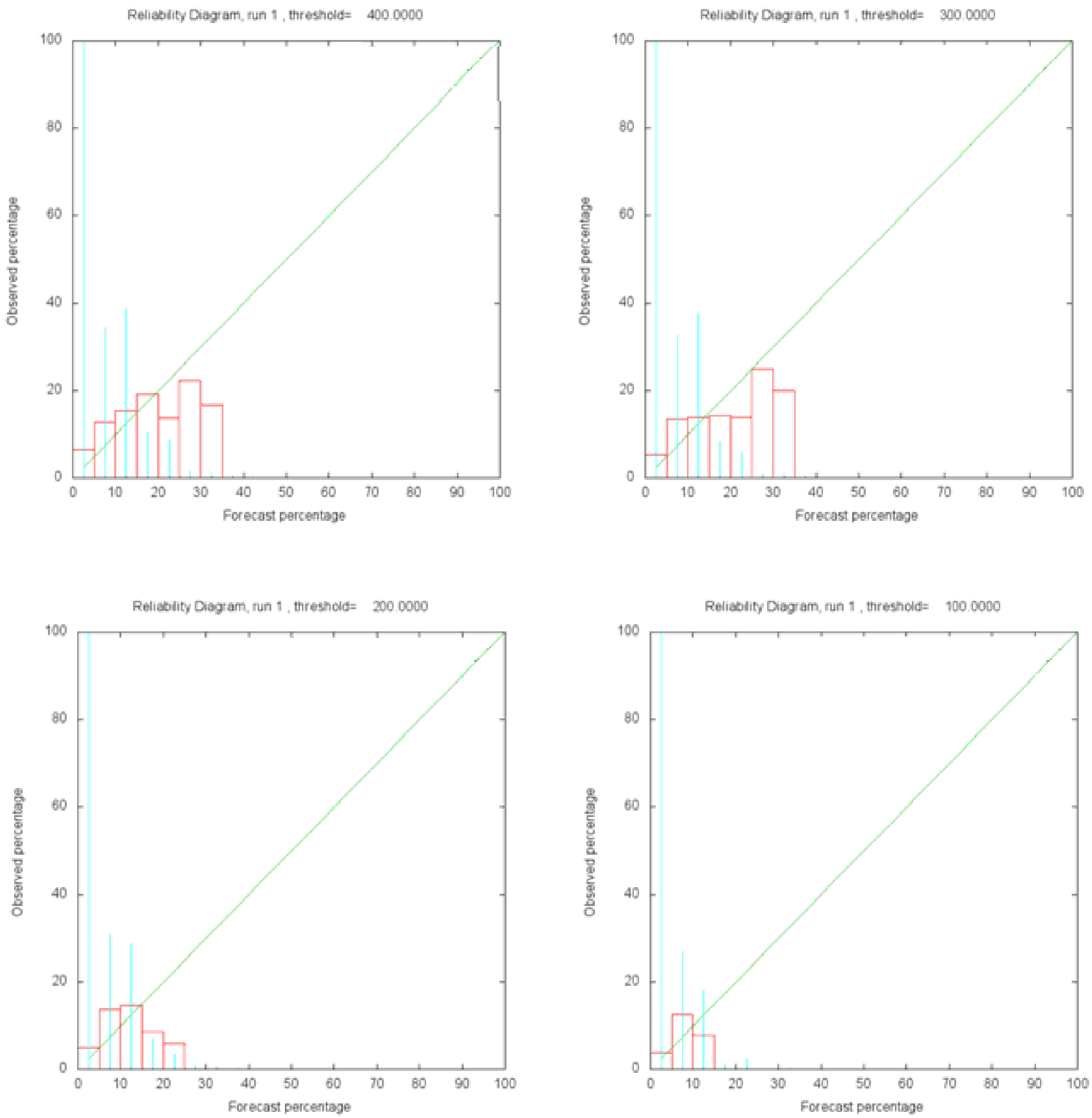


Figure 11: Reliability diagrams for forecasts for visibility below 1500, 1000, 750, 500, 400, 300, 200, 100m

Firstly, from the diagrams in Figure 11 it becomes clear that for all thresholds, forecasts are more or less reliable: the observed percentage grows with climbing forecast probability. The observed percentages follow the diagonal line of equal forecast probabilities quite closely, and the largest deviations in these diagrams occur in the bins with smaller sample sizes. This means that, in principle, the system is reliable and produces informative visibility forecasts.

Secondly, we can see in these reliability diagrams that with decreasing visibility threshold, the maximum forecast probabilities drop, from 50-55% for visibility below 1500m until 10-15% for visibility below 100m. This is a drawback of these forecasts, because one would very much like to be able to forecast extreme events with high probability. It is logical though, because the lower the threshold, the more rare the event, and more rare events are less likely to be forecast with high probability.

Another reason for these low forecast probabilities could be that low visibilities often have a very heterogeneous horizontal distribution. Fog often comes in banks and not as a horizontally uniform layer. A realistic situation could be that there is fog, but only a more or less stochastical 20% of the area is covered by dense <100m visibility fog. It is likely that of all good analogs of that situation, only around 20% corresponds to an observation of <100m visibility just at the observational site. As a result, no probabilities much higher than 20% can be forecast, while the analogs could still be relevant and show almost perfect analogy. The temporal integration of considering 00-06UTC minimum visibility is expected to reduce part of the spatial variation already, because the advection of fog in the 6h-period will partly remove the effect of the spatial variability. A spatial integration, like taking the minimum visibility of not one, but of a number of stations in the same region, could further account for the effect of more or less random spatial heterogeneity of low visibility events.

C2. Visibility predictions using one field of analogy

We have also investigated the importance of the procedure of combining different fields of analogy, for doing analog postprocessing for visibility. For each of the fields used in the previous experiment (T, q and √ff), we have performed a validation run in which analogs have been taken for only one of the three fields. We have plotted the Brier skill scores of these runs together with that of a validation run as in the previous experiment, using all three fields combined with equal weights. We have looked at predictions of visibility under the thresholds of 1000m (fog according to WMO) and 200m (very dense fog). See Figure 12 for the results.

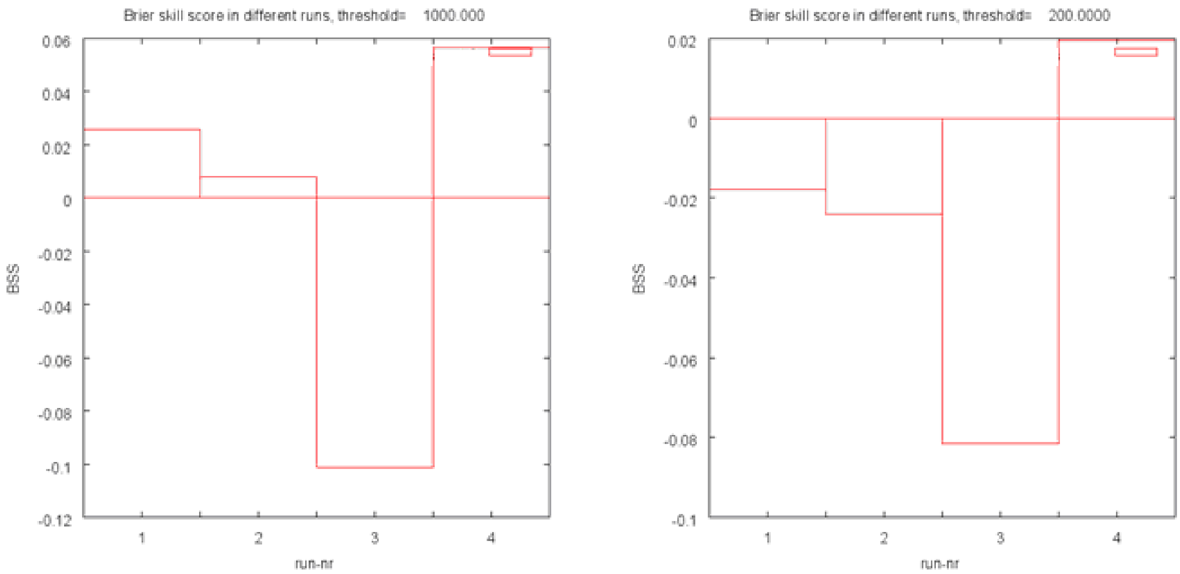


Figure 12: Brier skill scores of forecasts using 1) only T, 2) only q, 3) only √ff and 4) all three fields for analogy, for thresholds of 1000m (left panel) and 200m (right panel)

We can see that for probabilistic forecasts of visibility less than 1000m, temperature and moisture have a positive Brier skill score compared to climatology, whereas √ff has a negative BSS. Combining the three fields results in a positive BSS. For visibility less than 200m, all three fields have a negative BSS when used separately, whereas combining the three fields results in a positive BSS. Apparently, it is doubtful if any of the fields of analogy can separately work as a basis for skilful prediction of fog, while a combination of the fields seems skilful. Considering the harshness of the Brier skill score (Mason, 2004), it can not be said that the negative scores in these figures mean that the associated forecasts do not contain valuable information.

V. Conclusions

A. Sensitivity experiments

A number of sensitivity analyses have been performed, to learn about the sensitivity of the analog postprocessing system for different parameters, and to find out which configuration of the system could be appropriate for use in the other experiments. In these analyses, a 'generic' type of analog postprocessing is considered, predicting T_x on the basis of the T_{1000} analogs of 1 day (+24h) forecasts. A more or less 'optimal' configuration of 3-7E/50-54N as the domain of analogy, and 20 (0.44% of the used archive) as the number of analogs has been found, to be used in the other experiments in this study. Also, it has been found that the validation results of T_x predictions based on T_{1000} analogs vary between different seasons, and presumably are best in the winter season. Therefore, all other experiments in this study have considered predictions in the winter season based on analogs from the same season, unless mentioned otherwise.

Although these sensitivity experiments showed different validation results for all different settings, differences in validation results were sometimes extremely small. Thus, the significance of the derived choices for settings of the other experiments can be doubted. The choices for 'optimal' settings made by the use of these sensitivity experiments do seem reasonable from an intuitive point of view.

B. Investigating climate change with z_{1000} spatial anomaly analogs

The analog postprocessing system has been used to investigate some of the effects of climate change in the Netherlands. For this, the temperature trend in situations with a similar circulation pattern is considered. As a measure of circulation pattern, the z_{1000} spatial anomaly field is used. For all situations in the 7 most recent years of the archive, the best circulation pattern analogs have been taken from the whole archive, and their mean maximum and minimum temperatures in different subsets of 7 years of the archive is calculated. The trend in these temperatures is seen as a temperature trend without the effect of changes in the occurrence of circulation patterns. This experiment has only considered the winter season.

A significant warming in z_{1000} spatial anomaly analogs was found. This warming shows a pattern that is similar to that of the observed warming trend. It seems though, that the warming in z_{1000} anomaly analogs has a smaller amplitude than the observed warming, especially in the parts of the archive after 1986. It could be that some of the observed warming is not due to a mean warming of the air regardless of wind direction, but to a change in the occurrence of relatively warm or cold circulation types. It has not been investigated whether indeed there is a trend in the occurrence of relatively warm or cold circulation types in the archive, but this hypothesis agrees to conclusions of the statistical analyses by Van Oldenborgh and van Ulden (2003).

C. Analog postprocessing for the prediction of visibility

The technique of analog postprocessing has also been used for making and verifying short-term weather predictions. For this, +18 hour probabilistic predictions of 00-06UTC minimum visibility are made. In every prediction, the 30 best analogs with a combined analogy for T, q and $\sqrt{ff}$ were taken, the predicted probability being equal to the percentage of analogs in which 00-06UTC visibility dropped below the considered threshold.

For all thresholds (which were varied between 1500m and 100m visibility), reliability diagrams showed that the observed occurrence was higher with higher forecast probability. It appeared that the lower the threshold, the smaller the maximum probability that was forecast by the system. This can be caused by lack of forecasting skill, but it can also be related to the spatially heterogeneous character of most low visibility situations. Considering this heterogeneity, maximum forecast probabilities of visibility could be increased by considering not the minimum 00-06UTC visibility at one site, but at a range of sites in a certain domain. This would probably improve the method for predicting visibility.

We have also looked at the results of visibility forecasts that use only one of the fields of analogy T, q and $\sqrt{ff}$. It appeared, as would be expected, that taking analogs that are only analogous to one of these fields results in worse predictions than taking analogs that are analogous to all three fields with equal weights.

VI. Recommendations

The results of this study leave many possibilities for further research. Regarding the climate change experiment, the significance of the results found here can be explored further, by doing the same experiment with different settings. Also, the possibility of not using a fixed number of analogs but a *fixed threshold* for the metric value of analogs could improve the results, because this would in principle guarantee that all analogs are good circulation analogs. Furthermore, the results of the experiment could be quantitatively linked to the results of other research. Most importantly, the amount of warming observed that is not explained by a warming of the mean per circulation type, could be related to the estimated warming caused by changes in the occurrence of circulation types found in other research (e.g. in the statistical analysis by Van Oldenborgh and Van Ulden, 2003, as mentioned earlier).

Considering the part about making predictions with analog postprocessing, a first recommendation would of course be to investigate the possibilities of predicting other variables than visibility as considered in this study. Visibility was just an arbitrary choice here, and other predictands that are not well represented in present NWP output could be tried as well. Examples are the occurrence of thunderstorms and freezing rain.

Also, many things can be thought of that could improve the visibility forecasts considered in this study. A much more thorough sensitivity analysis could be carried out, exploring the degrees of freedom of the method multidimensionally, so that a better configuration of the system can be found. Mathematically elaborated techniques can be used, such as in Fraedrich and Rückert (1998) who adapted metric weights of fields by iteration on forecast error.

Also, the visibility forecasting method could be improved by not forecasting visibility below a certain threshold at a particular site, but forecasting the probability of that visibility in a particular area, using data from several stations. This would presumably make that higher probabilities will be forecast, also for (dangerously) low visibilities, and it would make forecasts more practically useful, as weather forecasts are generally issued for an area and not for a single point.

Furthermore, for analog postprocessing in general, a coupling can be made with mathematical methods such as Empirical Orthogonal Functions (EOFs). These functions provide information on spatial and temporal teleconnections between fields. Relations found by using EOFs could be relevant for the choice of fields of analogy, their spatial dimensions and their mutual weights.

One can also think of using a different type of metric for analogy, for improving the system used here. Using standardized and weighted RMS values as we have done is a very simple approach. More elaborate techniques of determining the similarity between situations, like modern computational techniques of pattern recognition (as fuzzy logic, as in Hansen and Riordan (2001) and Hagen (2003)) are promising.

VII. Acknowledgements

I have done this work as an internship at KNMI, as a part of my Meteorology and Air Quality MSc education at Wageningen University. Firstly, I would like to thank my supervisor Kees Kok. For bringing this work to a good end (or even starting it), his experience with analog methods, statistical knowledge and programming insight have been essential. Others at KNMI who's support has been vital were Daan Vogelezang, who helped me accessing and decoding the ECMWF data needed, Janet Wijngaard, who worked on acquiring the necessary observational data, and Seijo Kruizinga, who provided some valuable final comments. I also have to thank KNMI itself, for letting me do this internship. I have learned a lot in the past half-year, about programming, statistics and forecasting, and about how to do research and write a report.

VIII. References

Baur, F.; Hess, P.; Nagel, H., 1944: Kalender der Großwetterlagen Europas 1881-1939. (Bad Homburg v. d. H.)

Baur, F., 1948: Einführung in die Großwetterkunde. (Dietrich, Wiesbaden)

Fraedrich, K. and Rückert, R., 1997: Metric adaptation for analog forecasting. Physica A, 253, 379-393

Hagen, A., 2003: Fuzzy set approach to assessing similarity of categorical maps. Int. J. Geographical Information Science, 17, 3, 235-249

Hansen, B. and Riordan, D., 2001: Weather Prediction Using Case-Based Reasoning and Fuzzy Set Theory. Workshop on Soft Computing and Case-Based Reasoning, International Conference on Case-Based Reasoning

Kerr, R., 1989: A new way to forecast next season's climate. Science, Volume 244, Issue 4900, pp. 30-31

Kok, C.J.. and Lammers, I., 1993: Using analogs to obtain weather information from the ensemble prediction system. Report of the expert meeting on ensemble prediction systems, 6-7 july 1997, p. 93-101

Kok, C.J., 2001: Calibration of eps derived probabilities. KNMI Scientific Report WR-2001-04

Kruizinga, S. and Murphy, A.H., 1983: Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. Monthly Weather Review, 111, 2244-2254

Lemcke, C. and Kruizinga, S., 1988: Model Output Statistics forecasts: three years of operational experience in the Netherlands. Monthly Weather Review, 116, 1077-1090

Mason, S. J., 2004: On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. Monthly Weather Review, 132, 7, 1891-1895

Radinovic, D., 1975: An Analogue Method for Weather Forecasting Using the 500/1000 mb Relative Topography. Monthly Weather Review, 103, 7, 639-649

Ross, J. D., 2005: Reexamining analog forecasting using NCEP/NCAR reanalysis. Master of Meteorology thesis, Pennsylvania State University

Talagrand, O., Vautard, R. and Strauss, B., 1997: Evaluation of probabilistic prediction systems. Proceedings of the ECMWF workshop on predictability. Reading, England 20-22 October 1997, 1-25.

Timbal, B. and McAvaney, B.J., 2001: An analogue-based method to downscale surface air temperature: application for Australia. Climate Dynamics, 17, 947-963

Toth, Z.,1989: Long-range weather forecasting using an analog approach. Journal of Climate, 2, 6, 594-607

Toth, Z., 1991a: Estimation of Atmospheric Predictability by Circulation Analogs, Monthly Weather Review, 119, 1, 65-72

Toth, Z., 1991b: Intercomparison of Circulation Similarity Measures. Monthly Weather Review, 119, 1, 55-64

Van den Dool, H.M., 1989: A new look at weather forecasting through analogues. Monthly Weather Review, 117, 10, 2230-2247

Van Oldenborgh, G.J. and van Ulden, A., 2003: On the relationship between global warming, local warming in the Netherlands and changes in circulation in the 20th century. International Journal of Climatology, 24, 1711-1723