FLEXe

Future Energy System

RESEARCH REPORT
HELSINKI 2016

Antti Mäkelä
Otto Hyvärinen
Jouni Kivirinne
Pentti Pirinen
Matti Kämäräinen
Hilppa Gregow

Report on design principles of the verification system of monthly to seasonal predictions in energy markets

CLIC

Solution Architect for Global
Bioeconomy & Cleantech Opportunities

CLIC INNOVATION OY
ETELÄRANTA 10
P.O. BOX 10
FI-00131 HELSINKI,
FINLAND
CLICINNOVATION.FI

CLIC Innovation
Research report no XX

Mäkelä Antti
Hyvärinen Otto
Kivirinne Jouni
Pirinen Pentti
Kämäräinen Matti
Gregow Hilppa

Report on the usefulness of long-range forecasts in the energy sector and ways forward

Name of the report: Report on the usefulness of long-range forecasts in the energy sector and ways forward

Keywords: Forecast, long-range

Summary

We examine the special needs of the energy sector in the long-range forecasting. The work is done jointly by Finnish Meteorological Institute (FMI) and Helsinki Energy (Helen Oy), the main Helsinki-metropolitan area energy provider.

FMI aims to provide long-range forecasts, in operational basis and tailored for various end users and sectors. FMI has for years monitored the usability and validity of long-range forecasts, from monthly to seasonal scales. The additional challenge for FMI is to provide usable long-range forecasts in the high-latitudes. The development and feasible operational usage of such long-range forecasts needs close cooperation with the specific end users as the weather parameters vary both in the accuracy of forecasts and the usability for a particular sector. For example, while precipitation is harder to predict and important for the hydropower production, temperature (and temperature anomaly) forecasts are more skillful and more important for energy consuming.

Based on the needs and requirements described by HELEN, we show the design principles of the operational usage of long-range forecasts, including the verification system for the products.

Helsinki, June 2016

CONTENTS

FLEX^e

Future Energy System

1 Introduction

Atmospheric conditions at present (i.e., weather) and their long term average (i.e., climate) are important factors in the optimized and secured integration and operation of future energy networks (Fig. 1). Meteorological data is nowadays available with high spatial and temporal accuracy, which enables their smart usage both in the energy production as well as in the customer side: energy producers may optimise their operations better by knowing the estimated demand within the next, say, 5 days; the consumers may estimate the need for, for example, heating or cooling by knowing the temperature range for the coming days. Furthermore, new emerging technology and automation will ease the consumers decision making by adapting e.g. the houses to the observed and forecast weather conditions automatically and by optimising from where and when the energy is acquired. Besides the smart control of energy, meteorological information is an essential parameter for

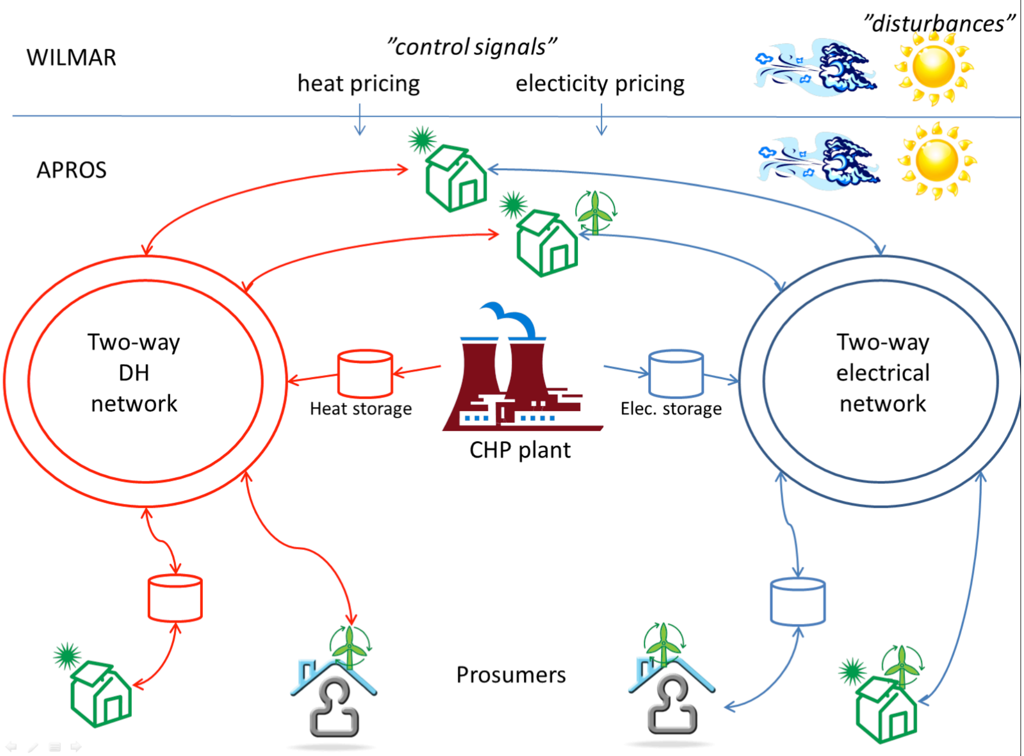

Fig. 1: Transition from (centralised) CHP plant towards a CHP system, where heat and electricity will be linked through multiple actors. Adapted from FLEXe project plan.

estimating the safety and outages of energy networks.

The basic or primary unit of meteorological information is a momentary and local observed property of the state of the atmosphere, an *observation* (Table 1). The observations can be considered as precise, although the accuracy of the measurement varies according to the weather parameter and the observations method (in situ vs. remote sensing). Meteorological *climate* is the 30-year average of the observations for any location (e.g. Helsinki, Finland etc).

Table 1: Definitions of meteorological information on various time scales. Adapted from World Meteorological Organization (WMO).

DEFINITIONS	RELIABILITY	SCALE&OTHER
OBSERVATION	~Precise (depends on the parameter)	Momentary in time and space
• **CLIMATE**	~Precise	30-year statistics of observations
FORECAST		
• **NOWCAST**	Excellent	Observed and modeled 0-2 hours ahead
• **VERY SHORT AND SHORT RANGE**	Good	Up to 72 hours
• **MEDIUM RANGE**	Good	72-240 hours
• **EXTENDED RANGE**	Moderate	10-30 days
• **LONG RANGE**	Moderate	30 days – 2 years
◦ **MONTHLY**	Moderate	Departure (deviation, variation, anomaly) of the averaged weather parameters from the climate values for that month.
◦ **3-MONTH OR 90-DAY**	Moderate	As above but for longer period.
◦ **SEASONAL**	Moderate	As above but for seasons.
• **CLIMATE PREDICTION**	Moderate	Beyond 2 years. Expected future climate including the effects of natural and human influences.

The purpose of weather forecast is to provide an estimation of the state and conditions of the atmosphere in the future. Weather forecasts include various temporal scales (Table 1), from few hours to decades. Generally, as the temporal scale gets longer, the more uncertainty the forecast has. The uncertainty (or the reliability) of the forecasts plays a key role in the usability of the forecasts.

At present, numerical forecasts have improved considerably. Still, weather forecasts cannot ever be 100% accurate. This is because of two main reasons:

- the chaotic nature of the atmosphere: the atmospheric motions are governed by equations which cannot be solved as such but need approximations. This leads to uncertainty which accumulates larger with time;
- accuracy and coverage of the meteorological observations: despite e.g. the FMI's modern and state-of-the art observation network, we cannot observe every point in the atmosphere, which induces a small uncertainty/error in the initial conditions of the model.

By looking at Table 1, we see that for the longer forecasts the uncertainty gets larger. However, it is surprising that for the longer scale forecasts the reliability can still be considered as moderate. This is because of the totally different nature and interpretation of the longer scale forecasts. For example, a seasonal forecast does not even try to model the exact temperature at a fixed location several months ahead, but predicts whether the temperature is *above or below the typical (mean) temperature of that month and how confident is the prediction*. This kind of information is still useful in many fields/sectors. However, for estimating the added value of a forecast of any scale, verification is needed. This means that the predicted parameter is compared against the actually occurred/observed value.

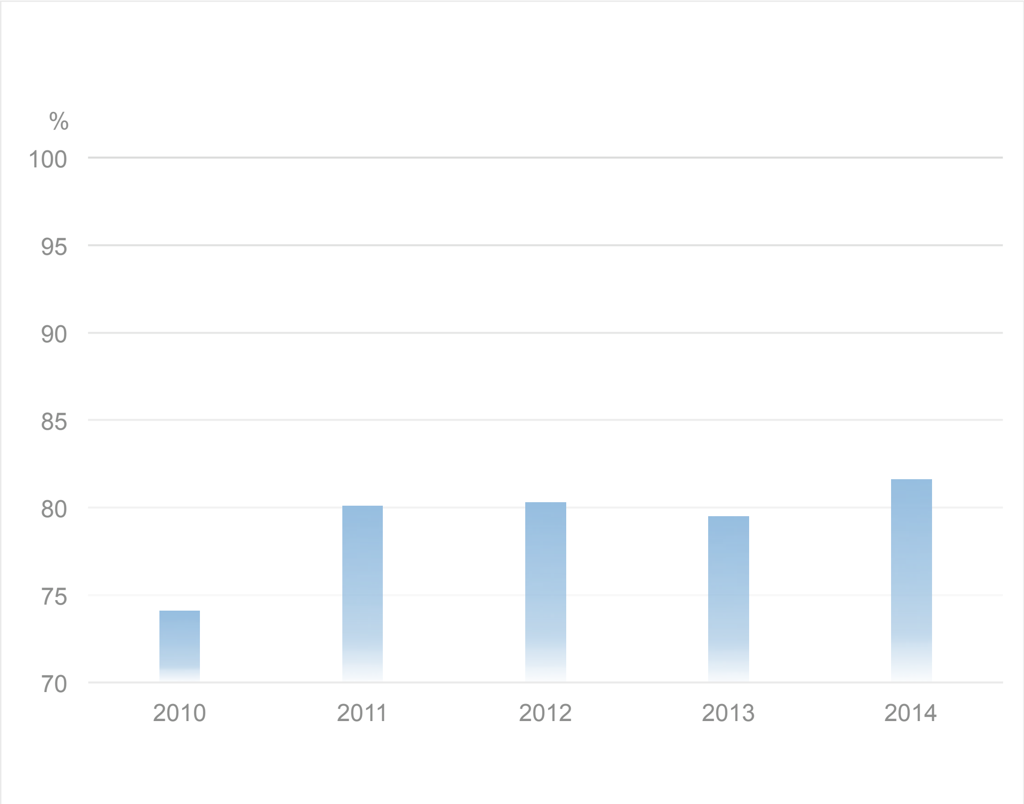

Fig. 2 Hit rate of temperature forecasts of 1-5 days in Finland.

In this report we concentrate to the usability of monthly to seasonal forecasts in the energy sector. The report is organized as follows: Section 2 focuses on the demands of the flexible energy management system in monthly to seasonal weather and climate scale according to a large Finnish energy company (Helen Oy). Section 3 illustrates the design principles of the verification system of monthly to seasonal predictions in energy markets. Section 4 summarizes the overall usefulness of long-range forecasts in the energy sector and how to implement them into the operational environment of energy providers in the future. Sections 5 and 6 contains conclusions and final discussion, respectively.

2 Demands of flexible energy management system in monthly to seasonal weather and climate scale

Helen Oy produces and provides electricity and district heating and cooling to Helsinki. We focus here on district heating. Regarding the demands of long-range weather forecasts for Helen can be summarized as follows:

- The production and consumption is largely governed by the temperatures.
- Spring and autumn are the most important times of the year: decisions have to be made regarding how much energy to be purchased and sold from/to stock.

Seasonal forecasting of temperature would be highly useful for Helen Oy if the forecast skill were good enough (i.e., better than a value based on climate).

Example of a seasonal forecast of temperature is shown in Fig. 3. Figure shows how much the 2-meter temperature is predicted to differ from the climatological average within the next three months. For example, in Finland the temperature anomaly is predicted to be +0.5-1.0°C within the whole country. How useful this kind of information for is?

Clearly, to have an estimation of the temperature anomaly for three months ahead gives information on the demand of heating, which in this case would be less than on average. However, it is important to understand that the result of Fig. 3 does not indicate anything about shorter temperature variations (daily, weekly) within the predicted three months. For example, if the first two weeks of January are exceptionally cold and the following two weeks exceptionally warm, this gives the same (i.e., common) average monthly temperature, but of course the demand for heating within that month is not the same as if the temperature would be somewhat the same within the whole month.

For the practical usage of long-range forecasts for the heating purposes, the forecasts should provide not only monthly anomalies but also each month separated in at least 2-week periods; this increases the usability of the forecast substantially. The forecasts should also contain a measure for the uncertainty.

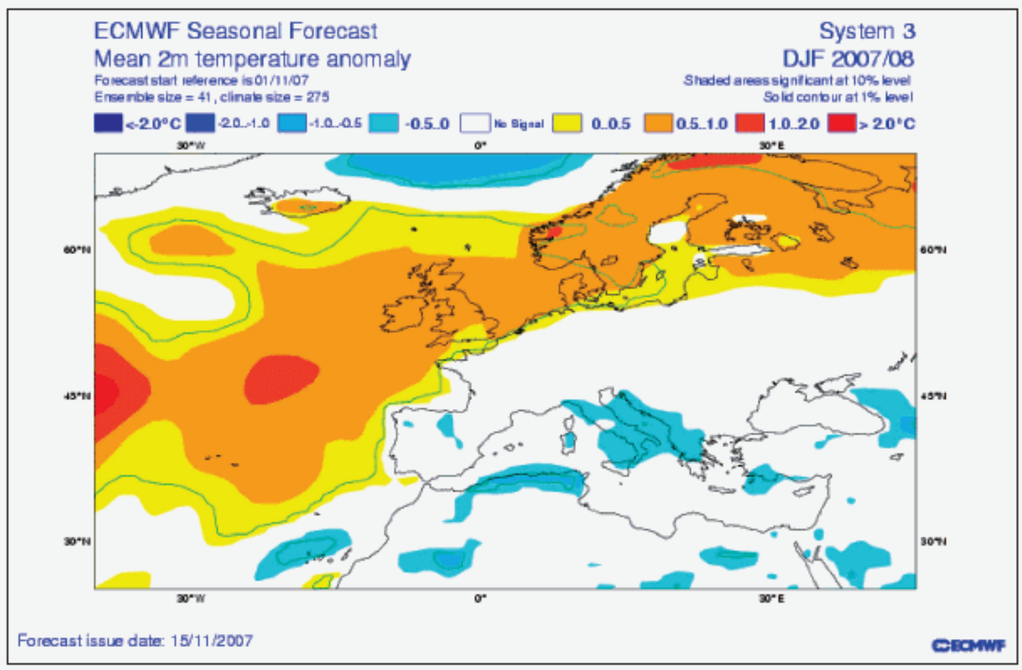

Fig. 3 Example of a long-range (seasonal) forecast of temperature anomaly for three months ahead from the launch date 15 Nov, 2007.

3 The design principles of the verification system of monthly to seasonal predictions in energy markets

Evaluation of forecast quality is usually called verification in meteorology. Verification can be divided into verification for administrative, diagnostic, and economic purposes. Here we concentrate on economic verification, which is concerned how forecasts support the end users in the decision making. If forecasts do not support users in the decision making, forecasts are useless to the end users. Economic verification is also most difficult part of the verification because in different situations different solutions are needed and usually no one-size-fits-all solutions exist.

3.1 Ensemble forecasting

Because of the uncertainty at the start of the forecast, a single deterministic forecast is useful (has skill) only for, roughly, one week. For longer forecasts (and sometimes also for shorter forecasts!), only probabilistic forecasts are skillful. Probabilistic forecasts are made by running many forecasts with slightly different starting conditions, forming an ensemble of forecasts. From these ensemble members, the ensemble mean can be calculated. Most of time, this is more skillful than any deterministic forecast.

Other way of utilizing this ensemble is assume that the spread of ensemble values is similar to the probability distribution, then the probability of exceeding some temperature (or precipitation etc.) threshold can be calculated from the distribution of ensemble members.

3.2 The Cost-Loss Model

As forecast providers cannot know the needs of all possible end users, the quality of forecasts are often shown using diagnostic verification measures and very simple decision models. The often-used, very simple decision model is **the cost-loss model**, where there is only one possible action: *Protect* (or not) *against a binary event*. The "binary" means that the even either occurs or does not occur. The protection against this event costs C, but if the event occurs (with the probability of p) when there is no protection, there will be a loss L.

```
          Event Occurs  Does not occur    Average          cost
Protect                p  C          (1-p)  C                =  C
Do Not Protect  p L      (1-p) 0         = p L
```

To be useful, the cost of *Protect* has to be less than the cost of *Not Protect,* implying $C < pL$ or $p > C/L$. So, according to the cost-loss model, you should protect if the probability of the event is larger than the ratio of the cost and the loss. This ratio can, and will be, different for different end users. This is one reason why giving the "most-probable" forecasts is not useful, as most users cannot get the maximum benefit from these forecasts. Probabilities are therefore preferred, as they can then be tailored for different customers and different situations.

This cost-loss model, despite its limitations, has been successfully used in the number of studies. However, defining the cost/lost ratio can be difficult, and in the literature, not many numerical results are reported. When the results are reported, the values are usually rather low, for example, 0.01-0.12 to fuel-loading of the aircraft.

Next we show how common diagnostic verification measures can be interpreted in the cost/lost framework.

3.3 Skill Scores and the Cost-loss Model

The often-used measure of probability forecasts is the Brier score, the mean of squared difference of forecasted probability (y_k) and binary observations (o_k).

```
BS = 1/n sum( (y_k  - o_k)**2  )
```

The value of Brier Score itself is not very informative, lower values mean better skill, but it is somewhat hard to compare values from different verification cases. Therefore, verification measures (S) are often normalized with an unskilled reference value (S_ref) and the best value (S_best) of the measure. This is a so-called Skill Score, defined as

```
SS = (S-S_ref)/(S_best-S_ref).
```

Skill Scores are easy to interpret as the best value is one, and zero value or less means the forecasts do not give any useful information. Then for the Brier Score, the Brier Skill Score is

```
BSS = (BS-BS_ref)/(0-BS_ref) = 1 - BS/BS_ref.
```

Connecting this back to the cost-loss model discussion, it can be shown that the BSS value is equal to the value of forecasts if the distribution of cost-loss ratios has the uniform distribution from 0 to 1. But in real life, the distribution for a single user is much narrower (for example, concentrating around 0.01-0.12, as in the example above). So seasonal forecast that has the overall BSS of zero (or even slightly negative) can be useful for some users.

Another complication is that the weather phenomena are usually continuous, so a more or less arbitrary threshold has to be defined and then exceeded for the event to occur. Luckily, a more complicated measure, the continuous ranked probability skill score (CRPSS), can be interpreted as the average Brier Skill Score when this threshold changes and the average is weighted by the climatological probability of occurrence of the threshold.

Therefore CRPSS can be a be interpreted as normalized measure of the potential economic value of a forecast system for a family of users which span the possible range of cost-loss ratios and for weather events which span the range of possible thresholds. Unfortunately, not all forecast providers show CRPSS for their seasonal forecasts, while most show BSS.

Note that these scores measure the potential value of forecasts for a family of users, so the value of forecasts for a single user should be individually evaluated.

This motivates local verification, because we have shown that most studies have been done on a global scale, and in especially seasonal forecasting the needs of the users have to be incorporated in the verification scheme to avoid both unnecessary pessimistic and sometimes too eager conclusions.

4 Usefulness of long-range forecasts in the energy sector and ways forward

In this Section we show example of the skill of long-range temperature forecasts in Helsinki in the period of March 2015 – August 2016. For the temperature, we estimate the heating degree days (HDD) for each month. HDD is the need for energy for heating a building and it depends directly on the outside temperature of the air; the need for heating is assumed to start when a certain outside temperature is observed (in Finland this baseline values is +17°C). For the long-range forecast data we use the UK MetOffice forecasts available at the Finnish Meteorological Institute.

The results are as follows (Fig. 4 and 5):

- The forecast has overall skill which is better than compared to the climatological average.
- Skill has monthly variation: in this example, May had the lowest skill, but with a simple bias-correction, the skill was improved.
- The Jan 2016 very short and cold period was difficult to forecast; however, it was out of ordinary based on the climate as well. Also, the operational short-range forecasts had also difficulties in capturing the start of the cold period.

To conclude, the forecast skill found in this study for HDD is surprisingly good, and it suggests to investigate the method further. Also, it would be important to widen the study to other available long-range data as well to examine the differences. As our study contained only about one year of data, the study should be rerun for a longer test period. However, even with the present findings, it would be interesting to test the method in operational environment for example at Helen Oy and estimate the methods operational usability.

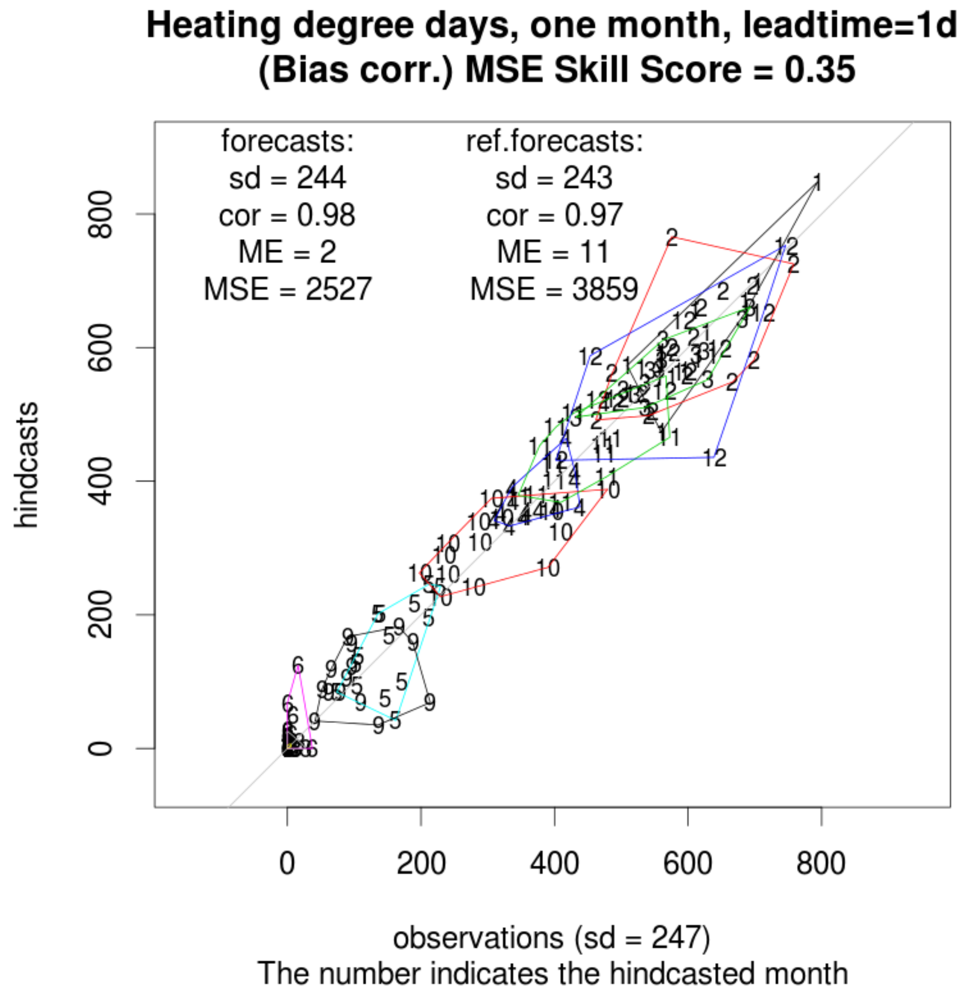

Fig. 4 The overall skill of the monthly forecasts for HDD in the test period.

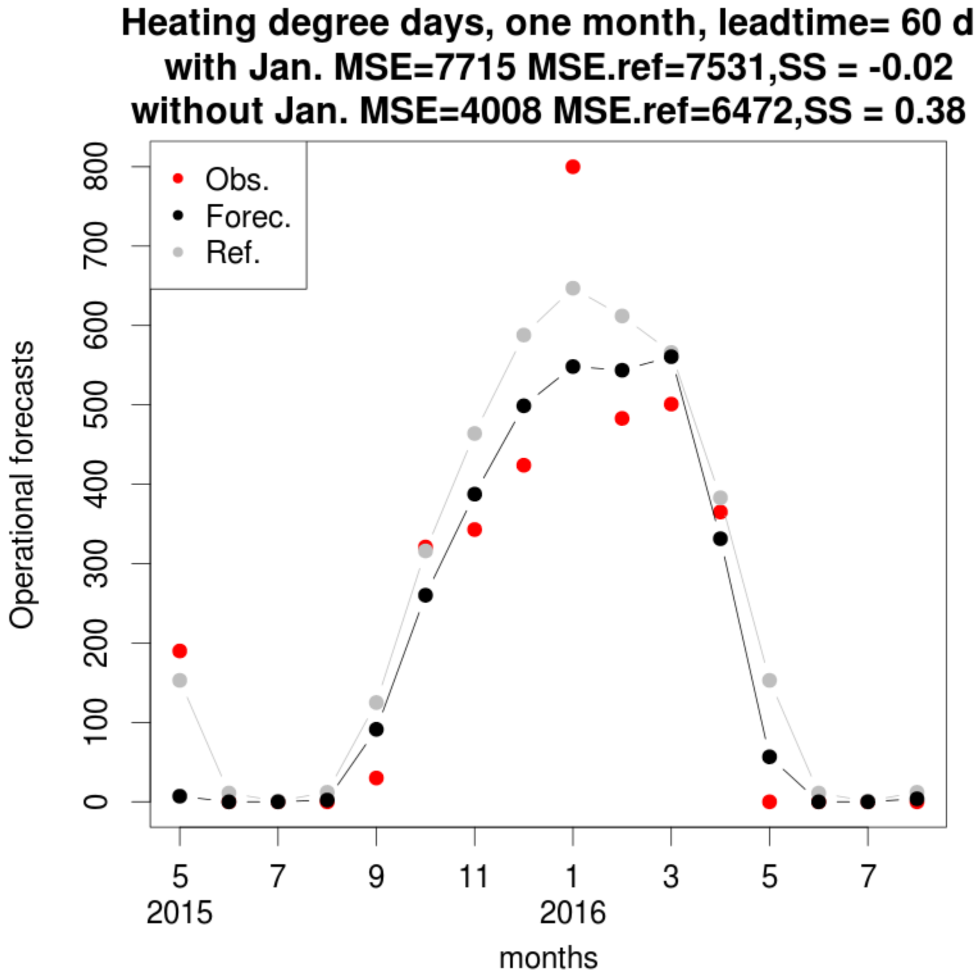

Fig. 5 The skill of the monthly forecasts of HDD in the test period.