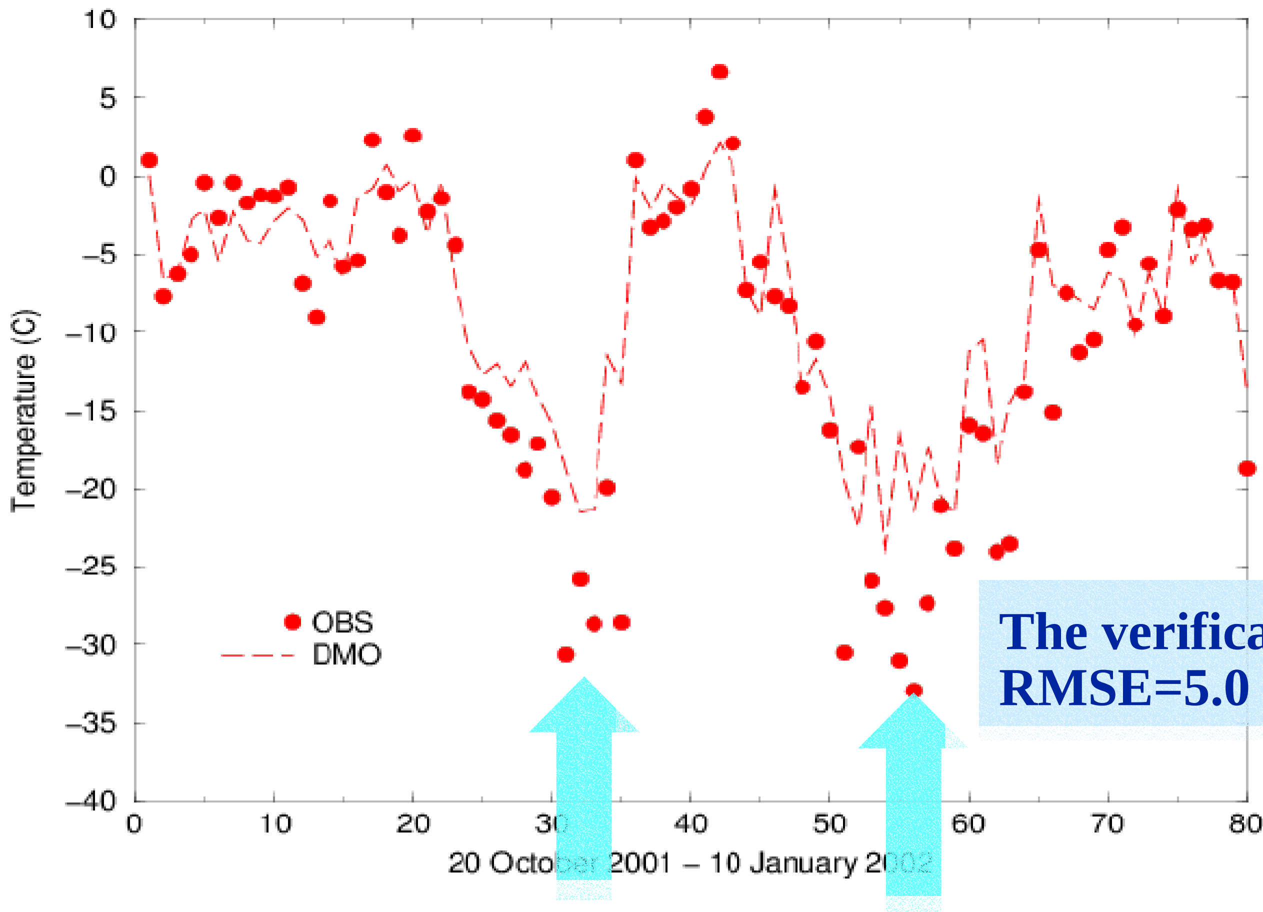
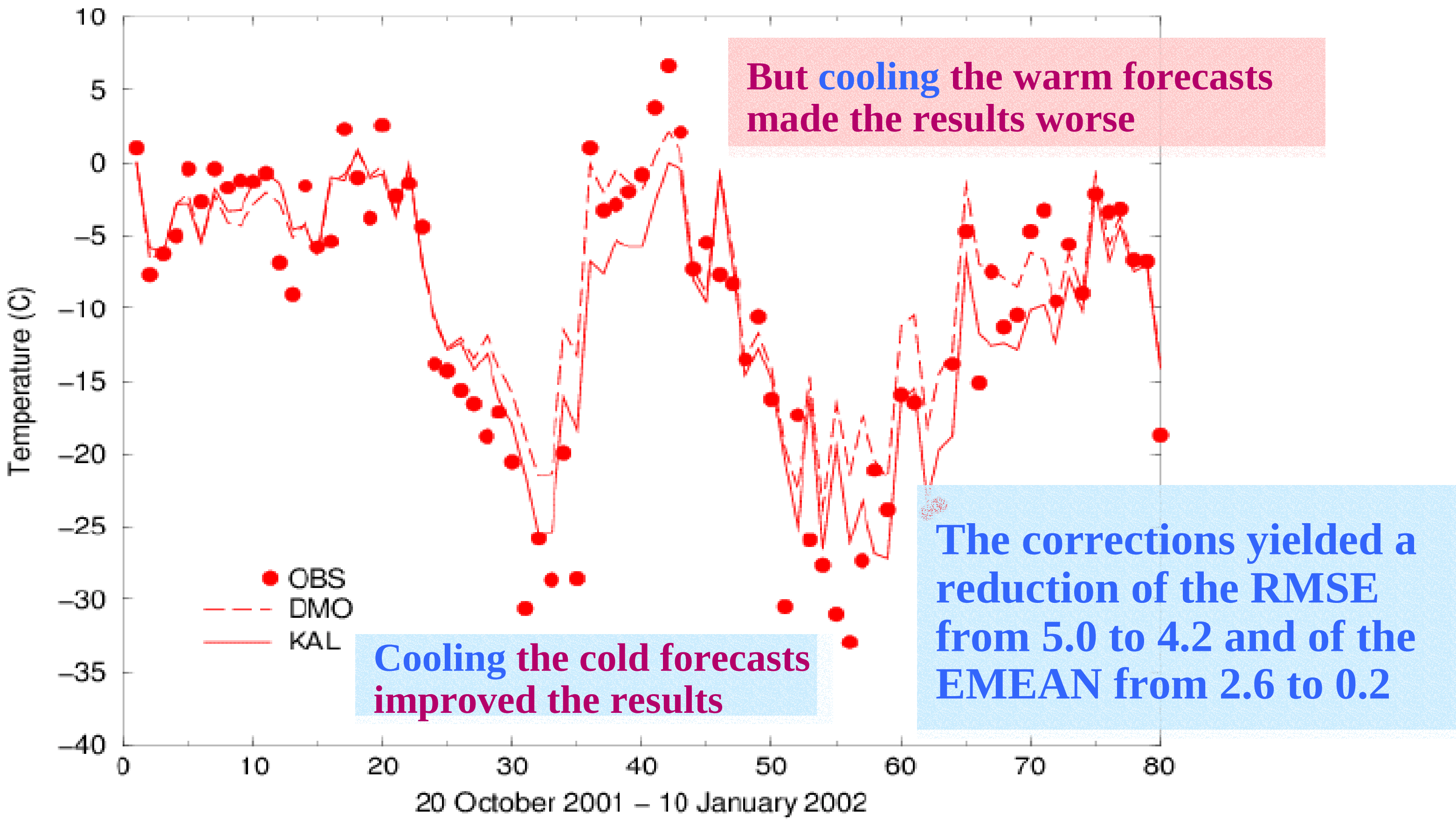
Understanding statistical interpretation and why better forecasts can look worse...

HIRLAM-44 24 hour 2 m temperature forecast for Kiruna in Lapland winter 2001-2002

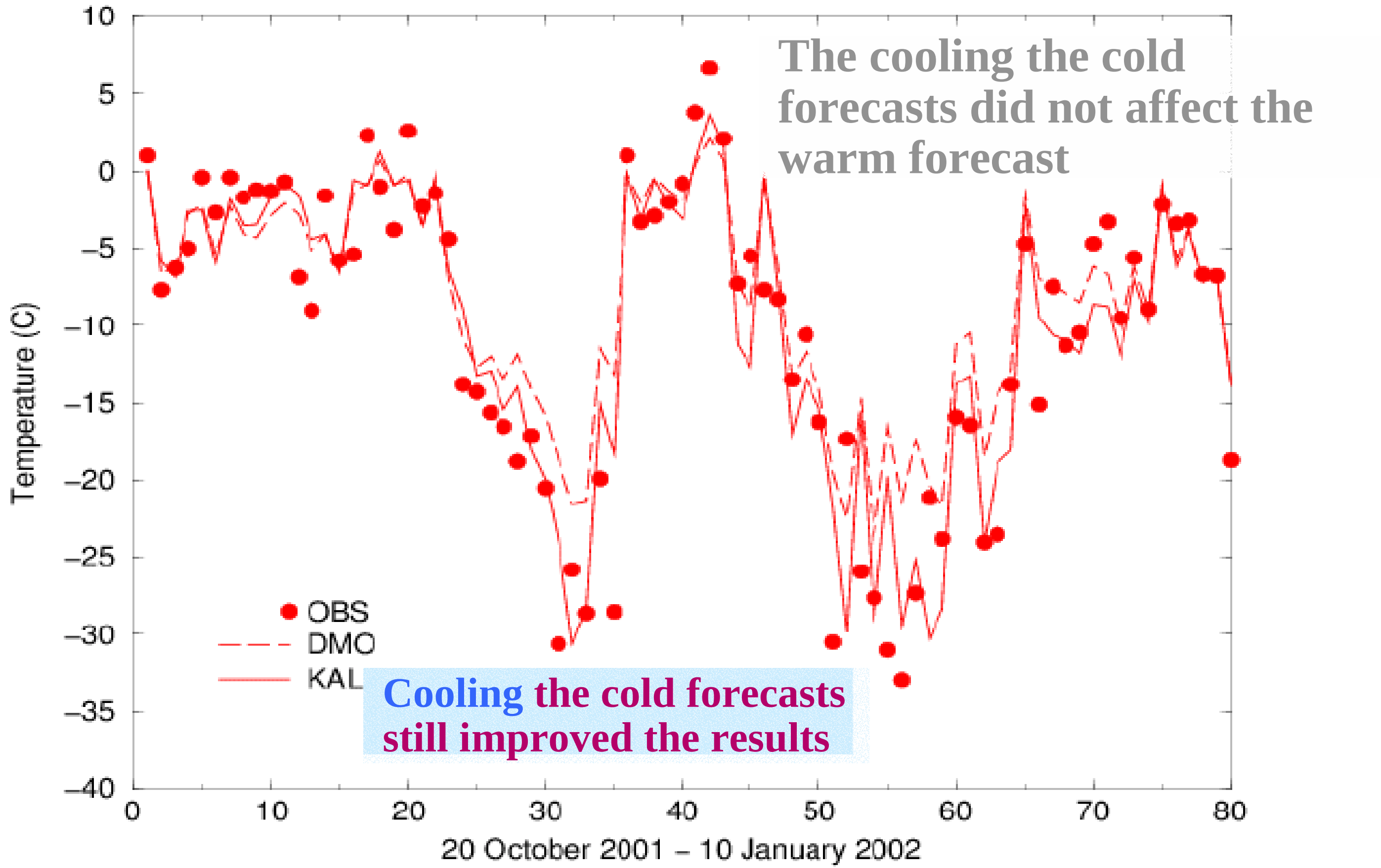


The verification yielded RMSE=5.0 Emean=2.6

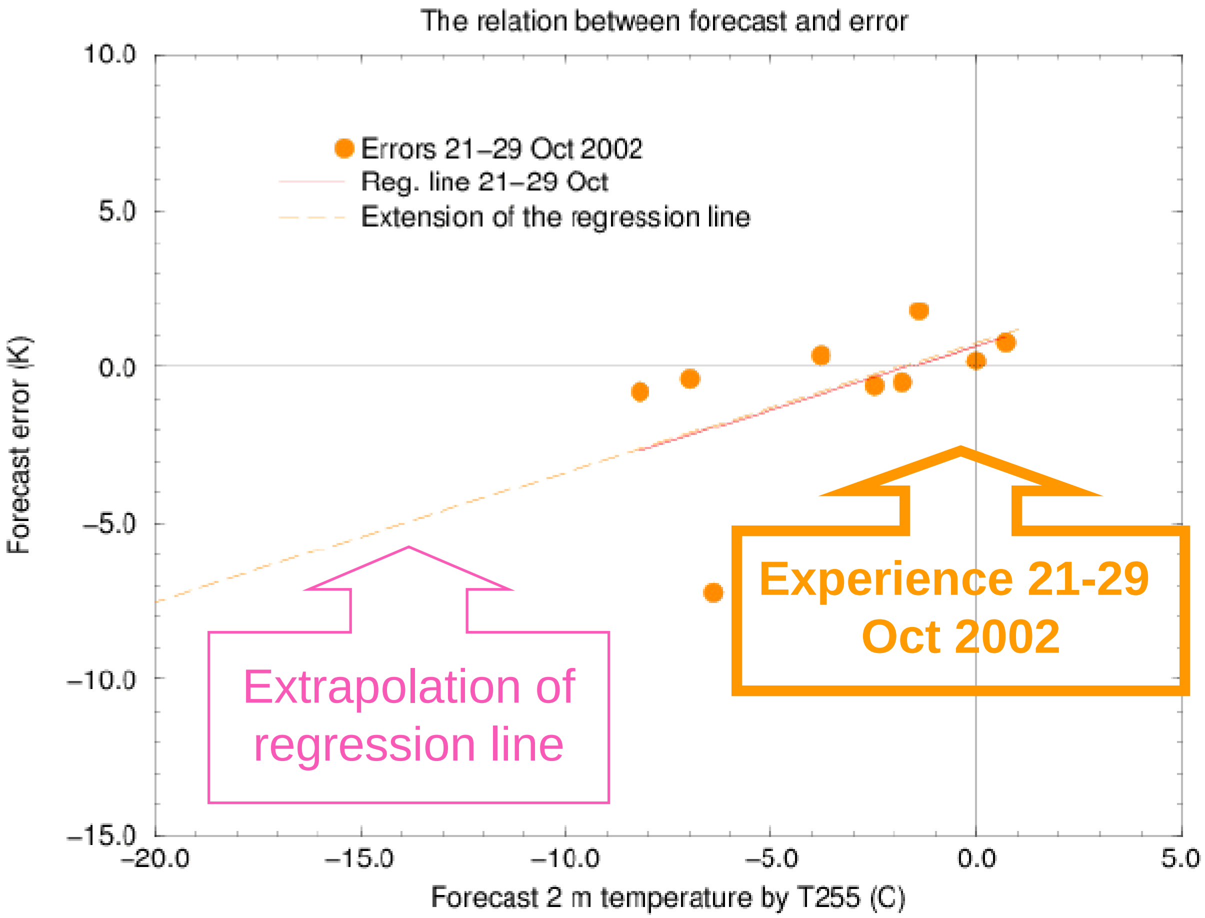
A 1-dimensional Kalman filter can reduce an overall bias



A 2-dimensional Kalman filter can provide different corrections to different regimes.

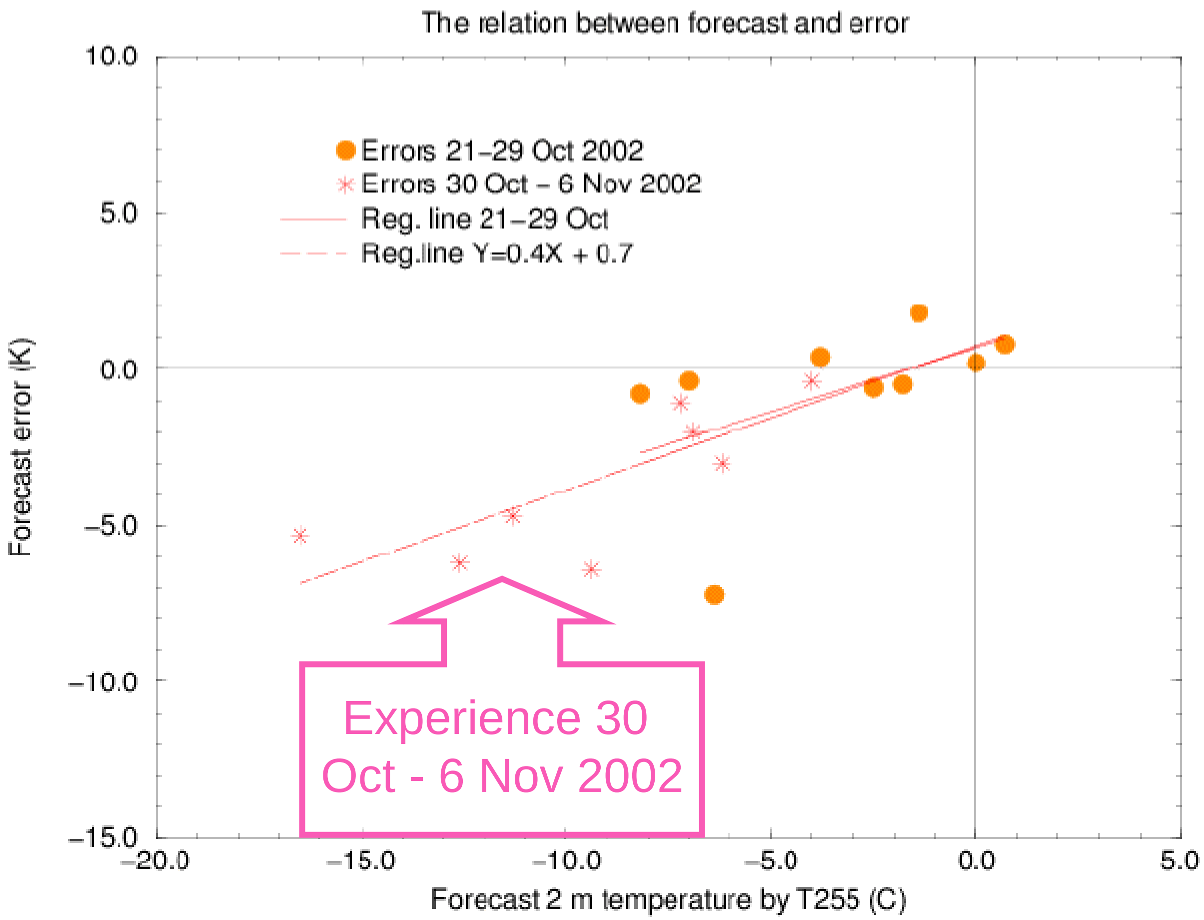


Principles of a two-dimensional error equation



Based on "experiences" in the range > -8 C the filter i capable of producing corrections for much lower temperatures for which it has no direct "experience"

Principles of a two-dimensional error equation

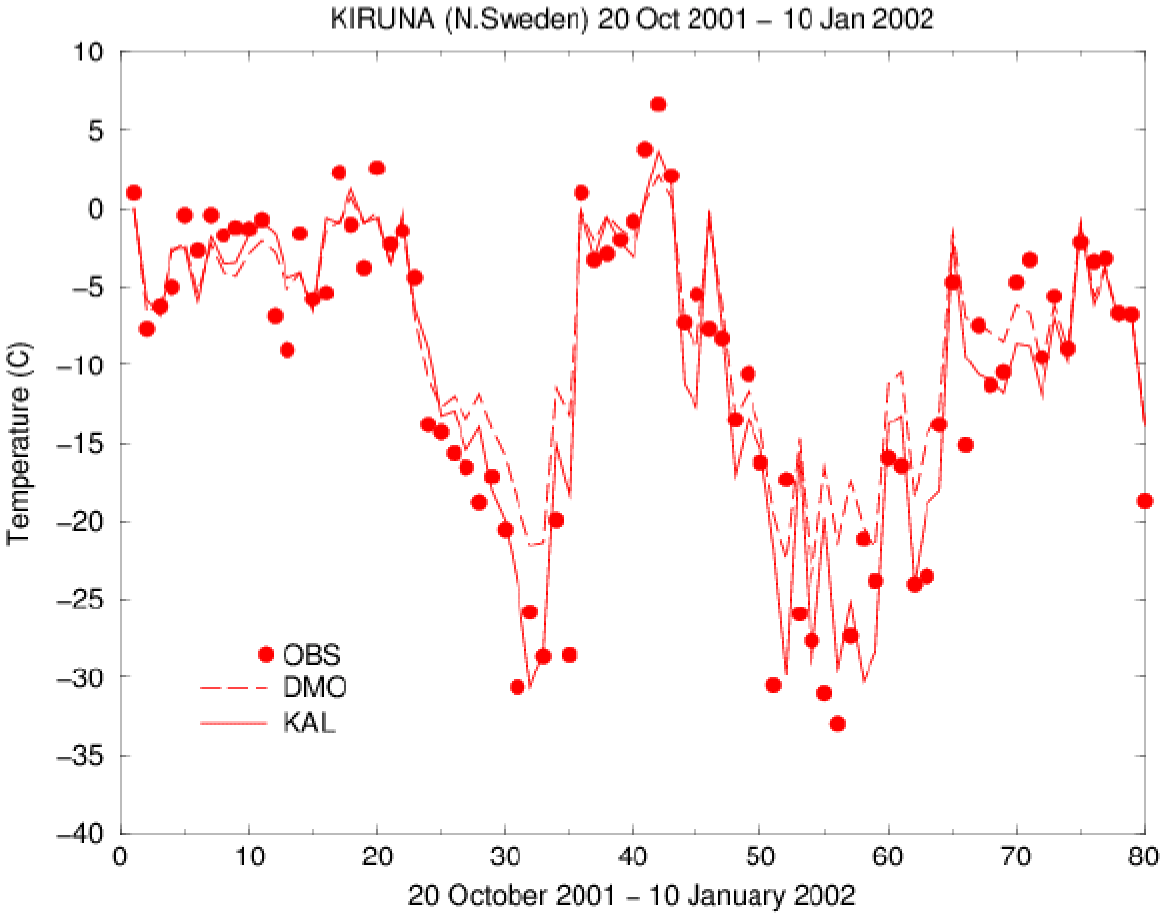


When the sub -8 C temperatures start to verify it can be seen that the extrapolation was realistic

The Kalman filtering has reduced two systematic errors: a positive mean error and an underestimation of the variability

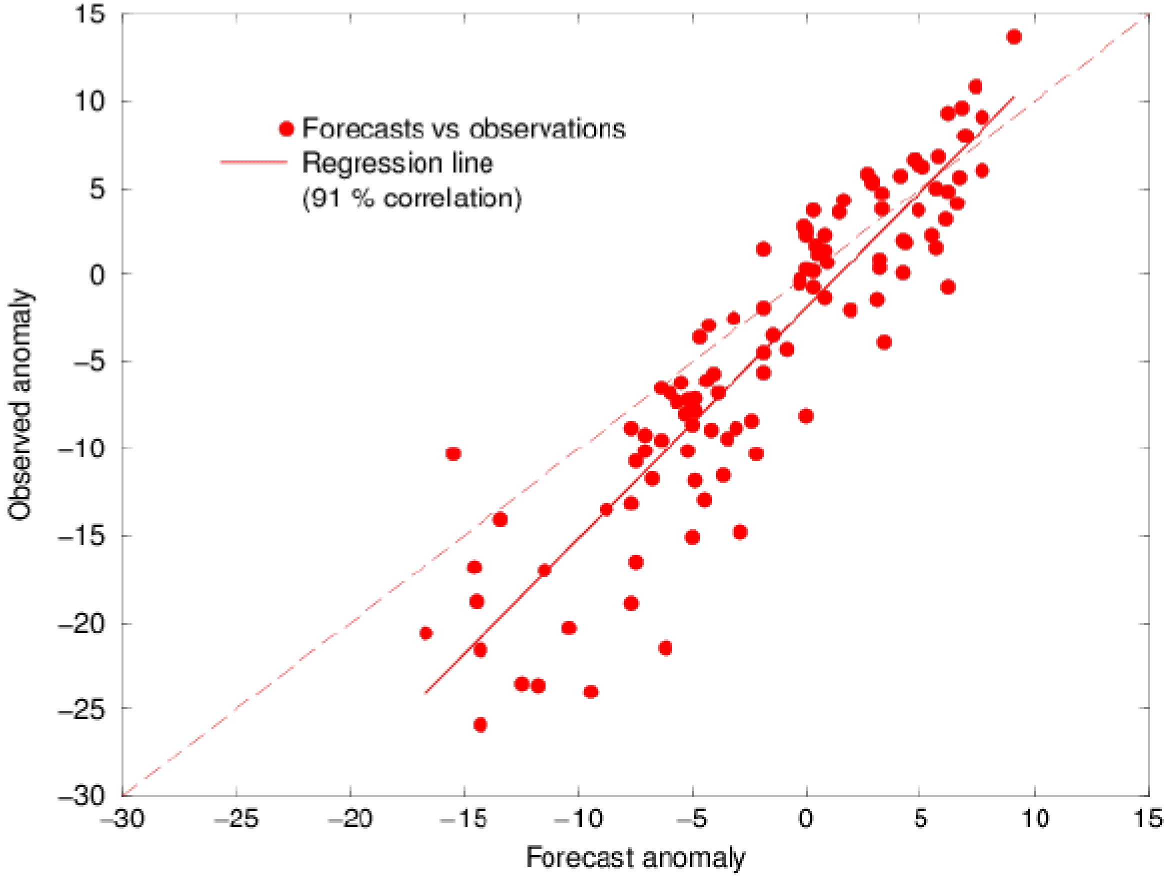
The mean error is reduced from 2.6 to 0.3!

…but the RMSE is only reduced from 5.0 to 4.6



To show that what looks like equal improvements are not quite "equal", we will make a simple manipulation of the data: *the observations and the numerical forecasts are "swopped"*

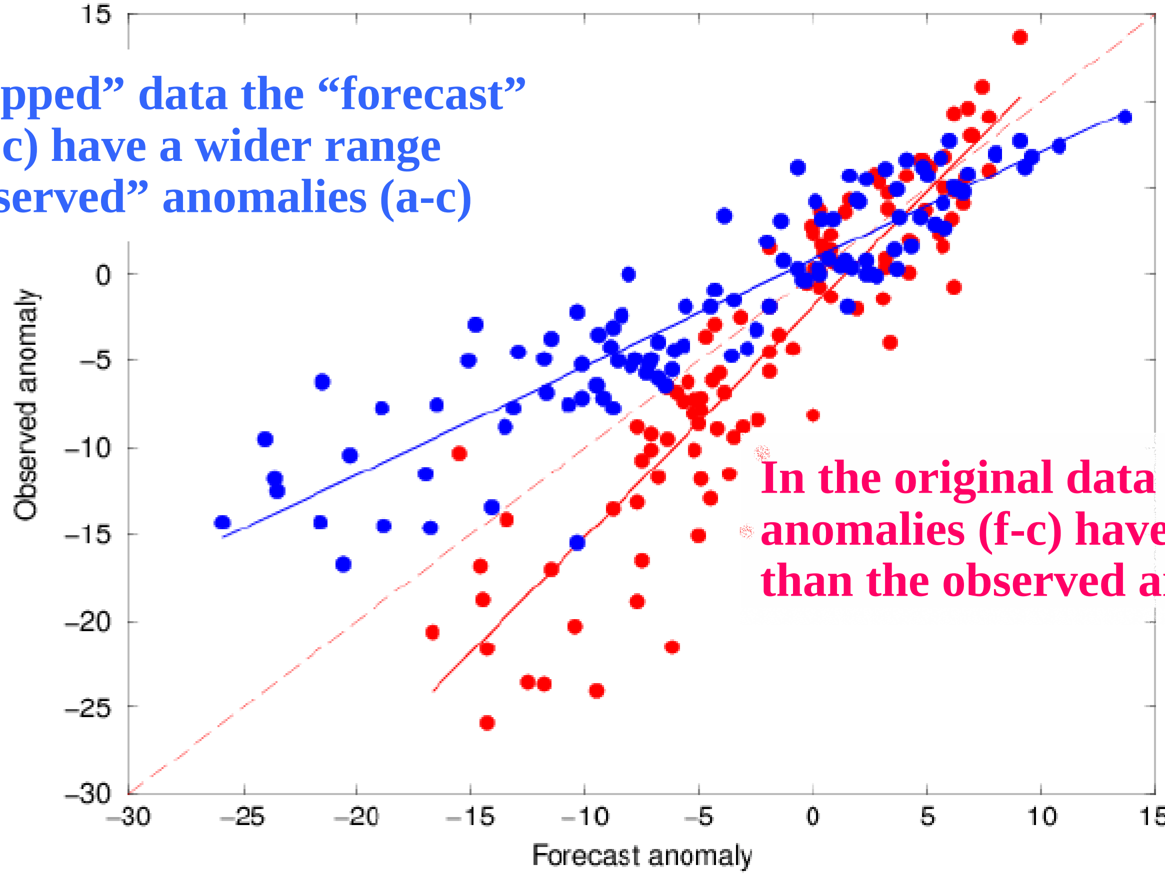
The relation between forecast and observed anomalies



While the forecast anomalies range between -18 to +10 C, the observations range between -22 C to +14 C

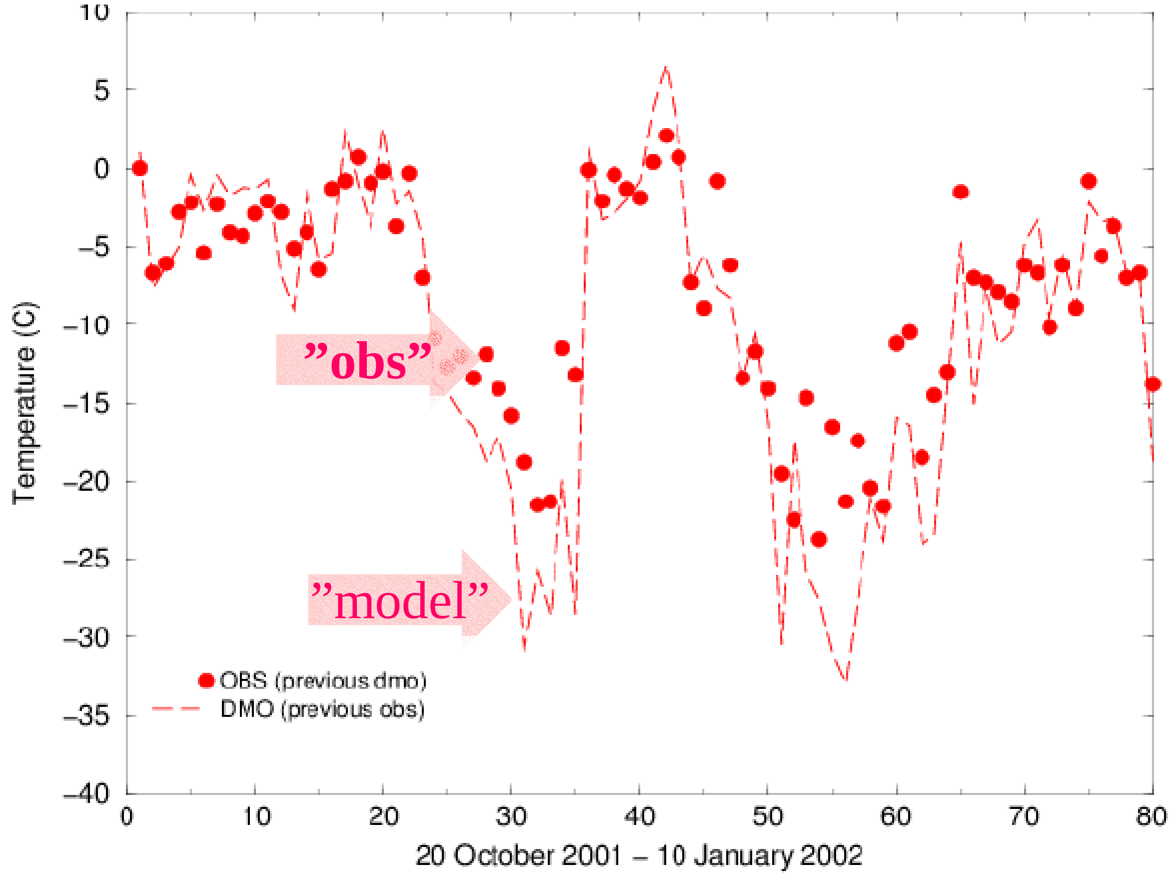
2 m temperature observed anomalies versus forecast anomalies

For the ”swopped” data the “forecast” anomalies (f-c) have a wider range than the “observed” anomalies (a-c)

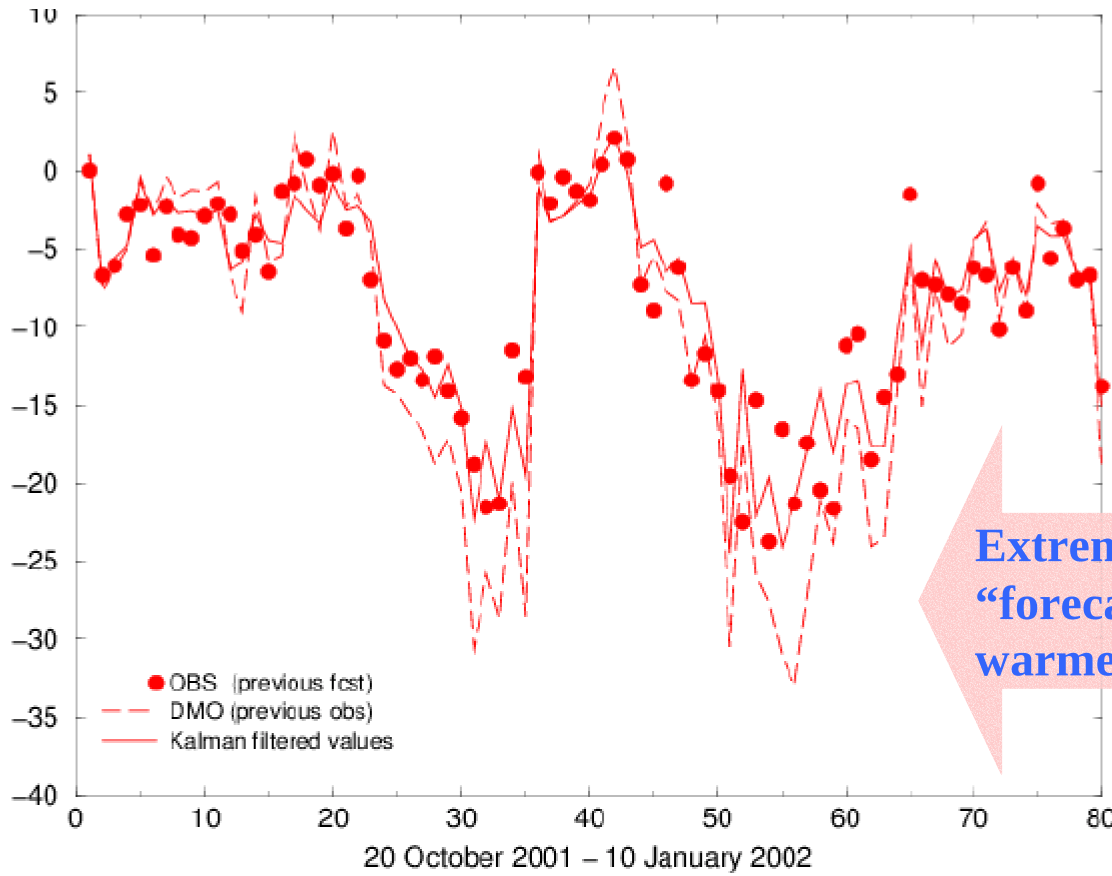


In the original data the forecast anomalies (f-c) have a smaller range than the observed anomalies (a-c)

24 hour 2 m temperature forecast for Kiruna in Lapland winter 2001-2002 - with observations and forecasts swopped



After Kalman filtering the EMEAN is reduced to zero and the RMSE is reduced from 5.0 to 2.9



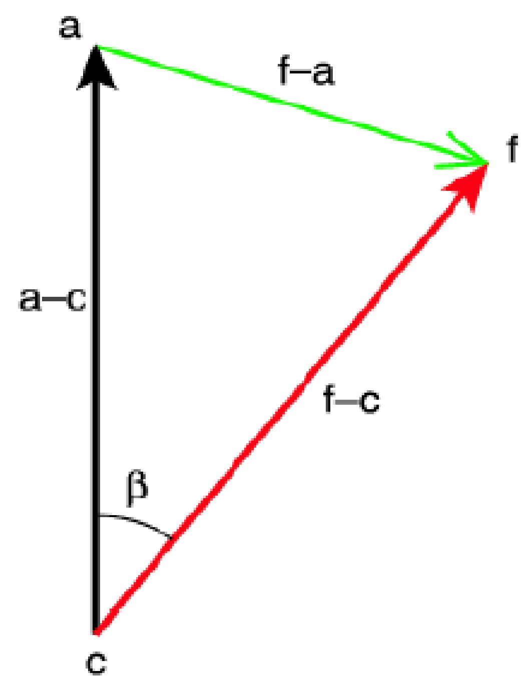
Extreme cold "forecasts" warmed

A simple vector-geometrical alternative

The previous mathematics can also be given a vector algebraic presentation where **a**, **f** and **c** represent states in some phase space

The length of the vectors represent A_a and A_f, and the difference **f-a** is proportional to the RMSE

With an underactive model **f-c** will become somewhat shorter. Also **f-a** will decrease and thus the RMSE

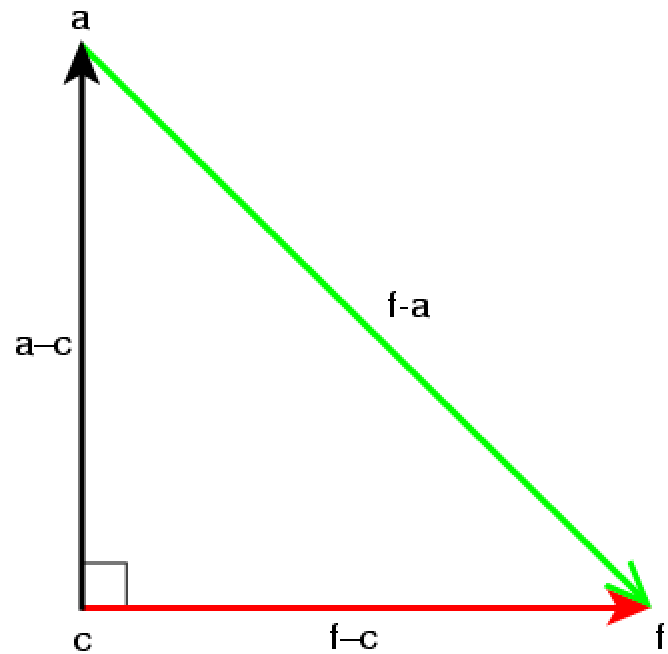


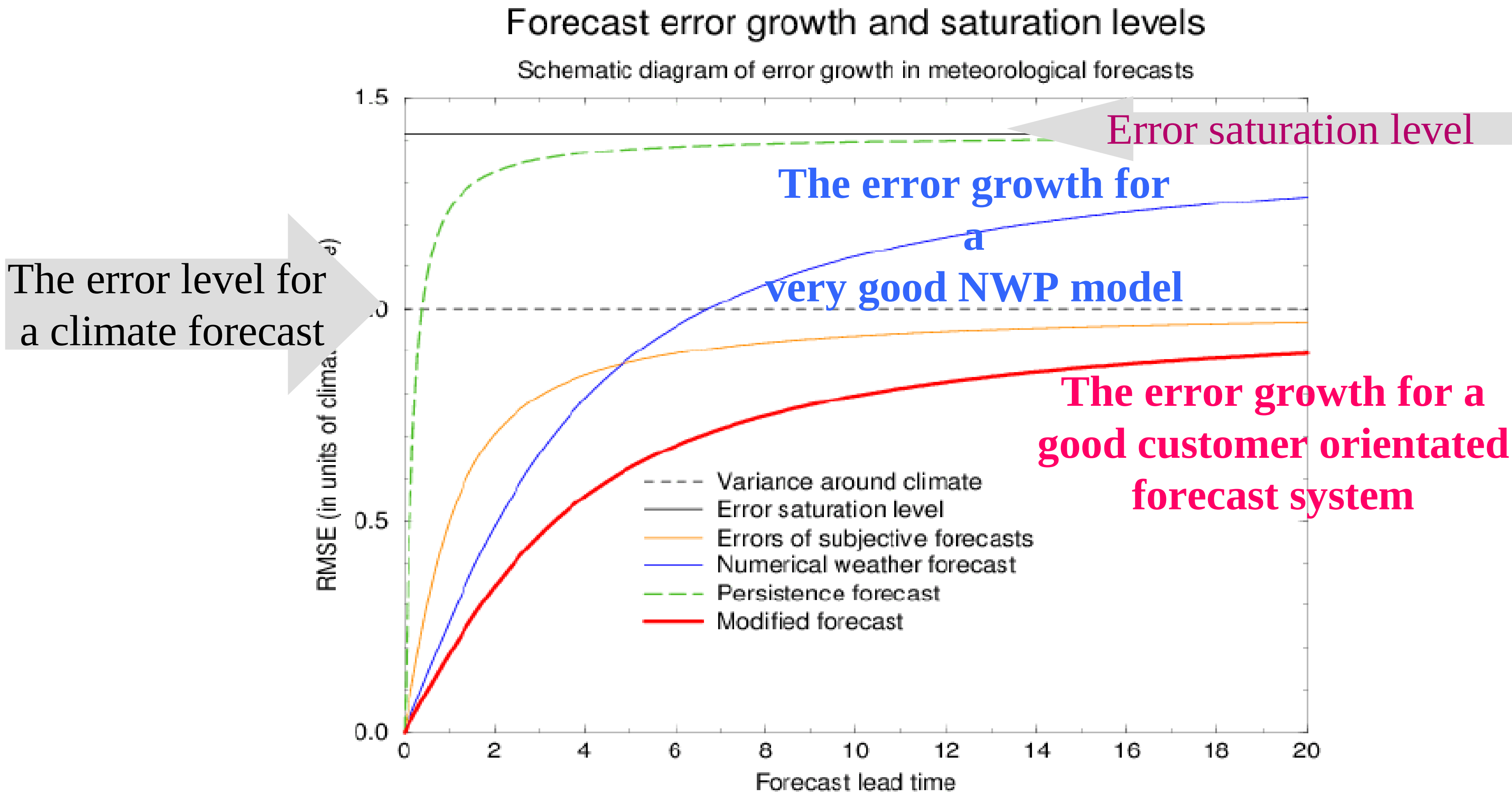
The RMS error saturation level

When the forecast is lacking skill the f-c vector is perpendicular to the verifying vector a-c

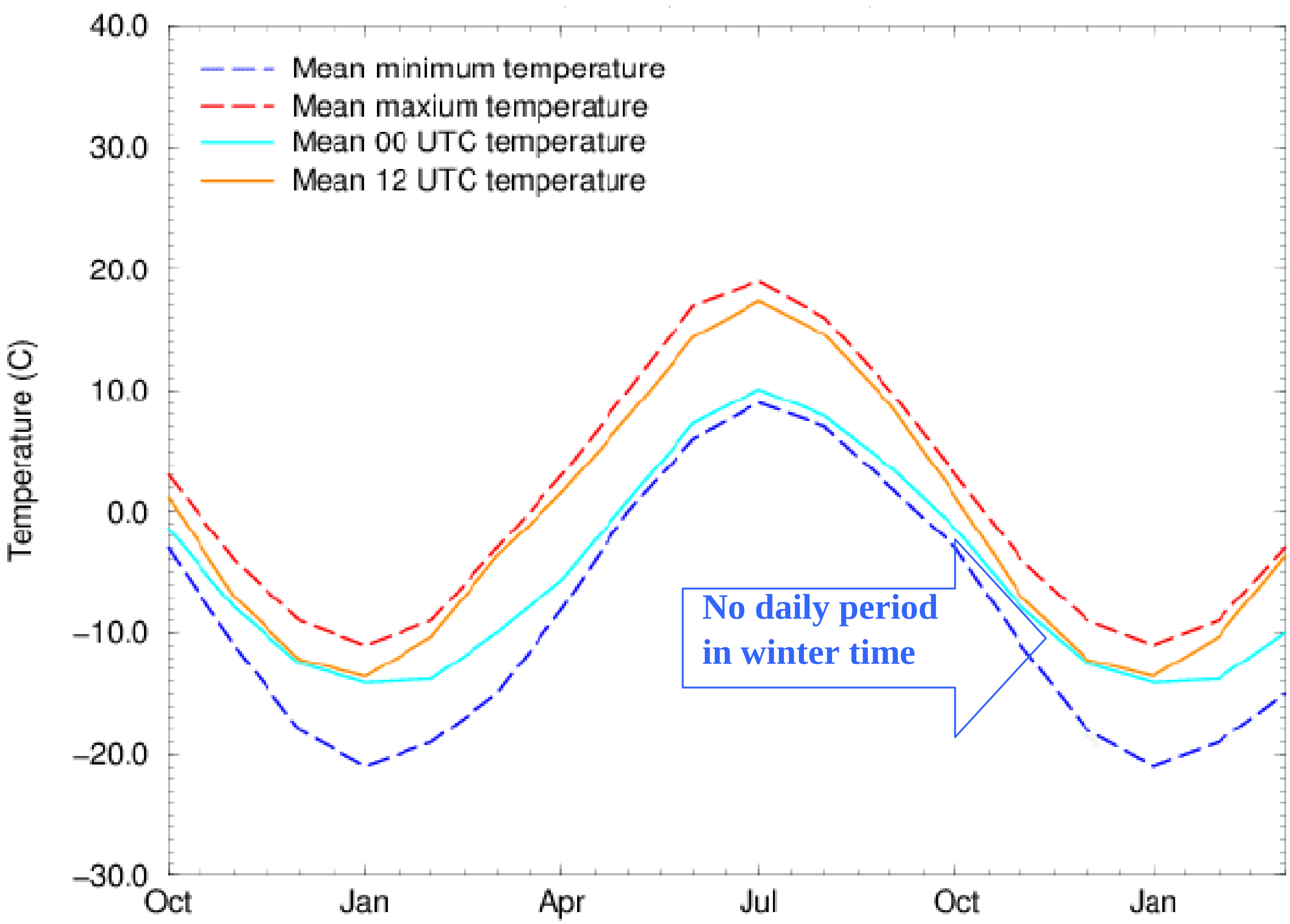
Cosine of the angle 90° is zero which is also the value of the ACC (Anomaly Correlation Coefficient)

It is easy to see that the maximum RMSE equals the variability times √2

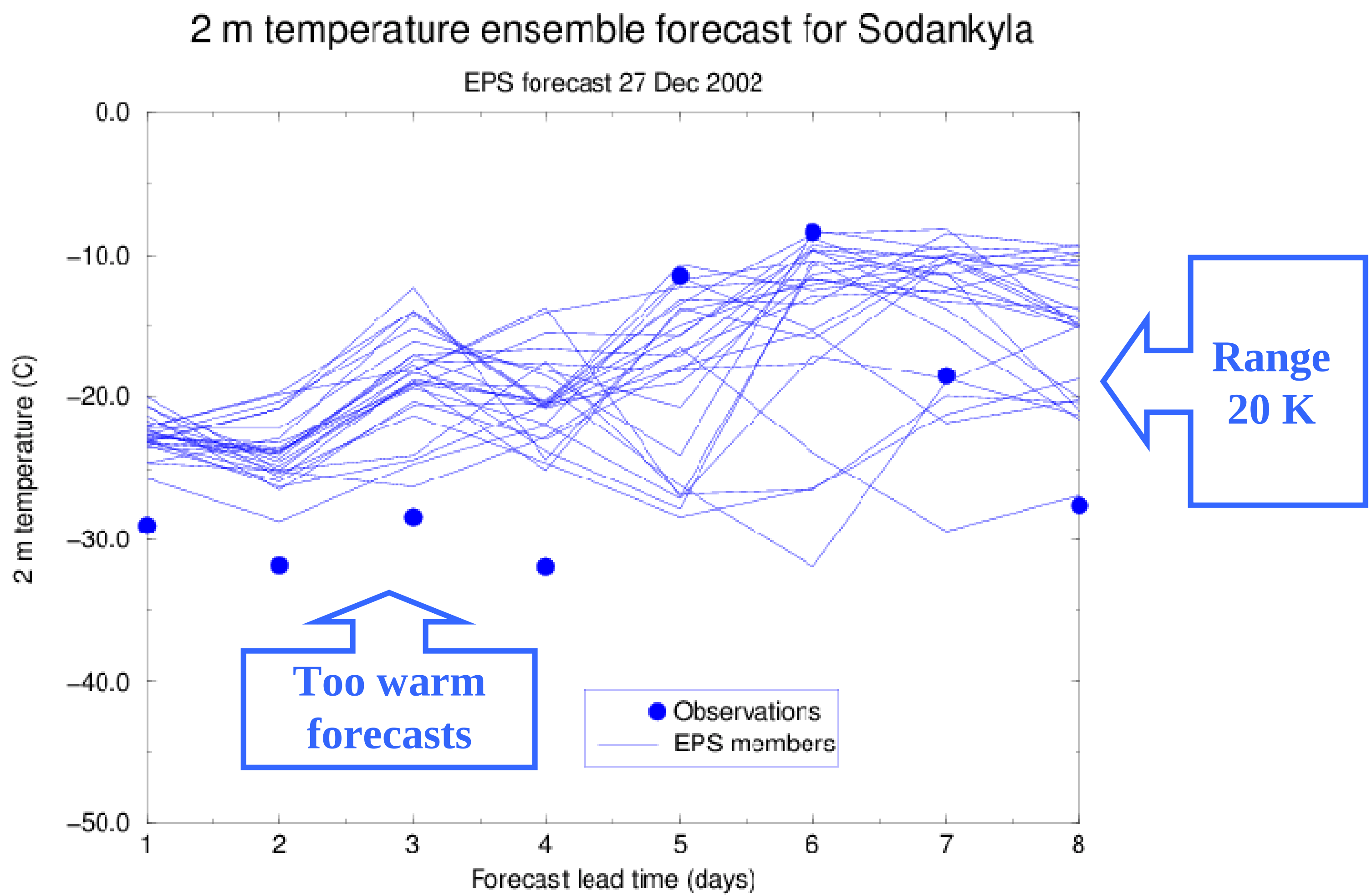


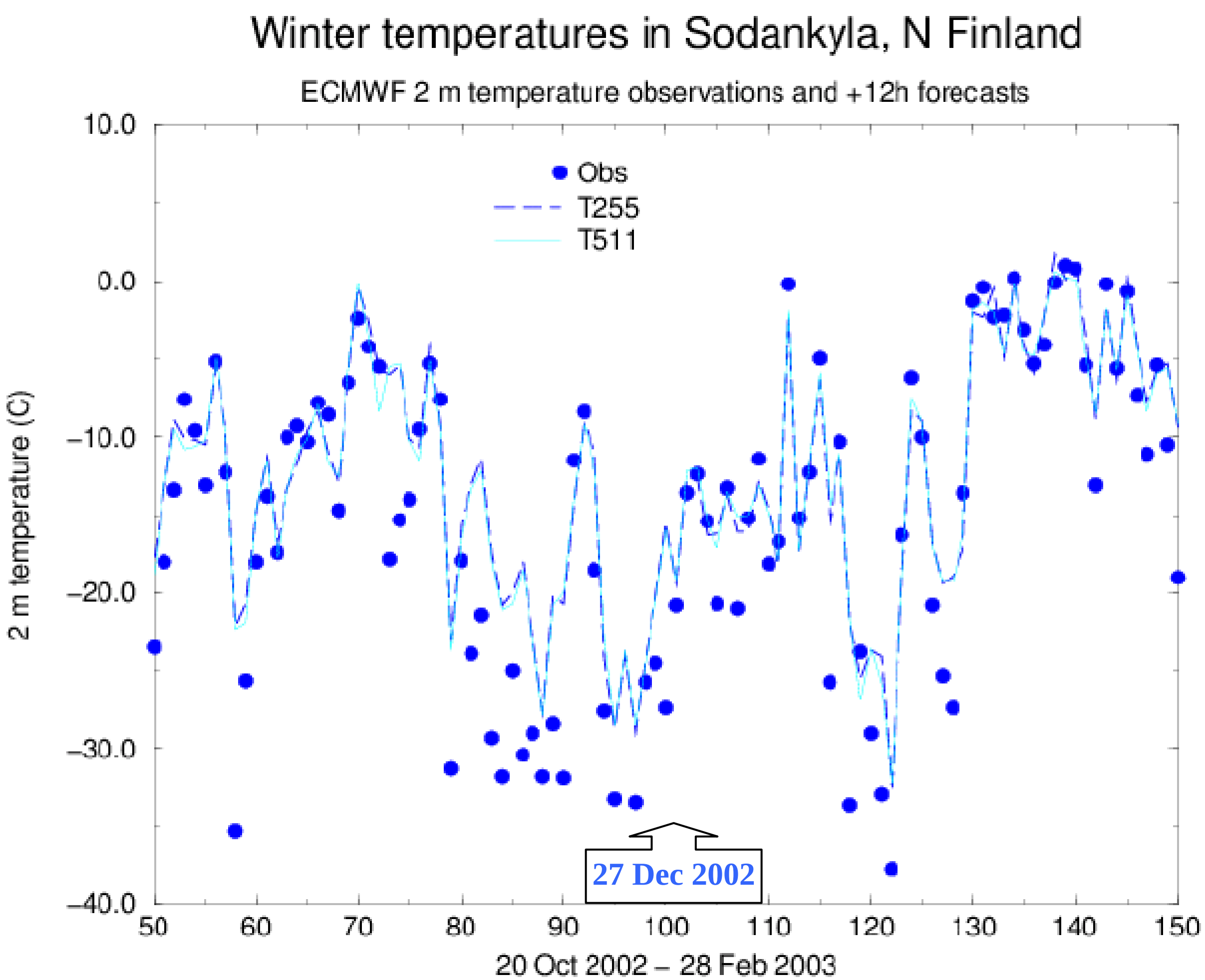

Forecast error growth and saturation levels
Schematic diagram of error growth in meteorological forecasts
RMSE (in units of clima
1.5
0.5
0.0
0
2
4
6
8
10
12
14
16
18
20
Forecast lead time
Variance around climate
Error saturation level
Errors of subjective forecasts
Numerical weather forecast
Persistence forecast
Modified forecast
Error saturation level
The error growth for a very good NWP model
The error level for a climate forecast
The error growth for a good customer orientated forecast system

Mean(Max/min) and mean(00/12z) temperatures for Sodankyla (02836) in Finnish Lapland

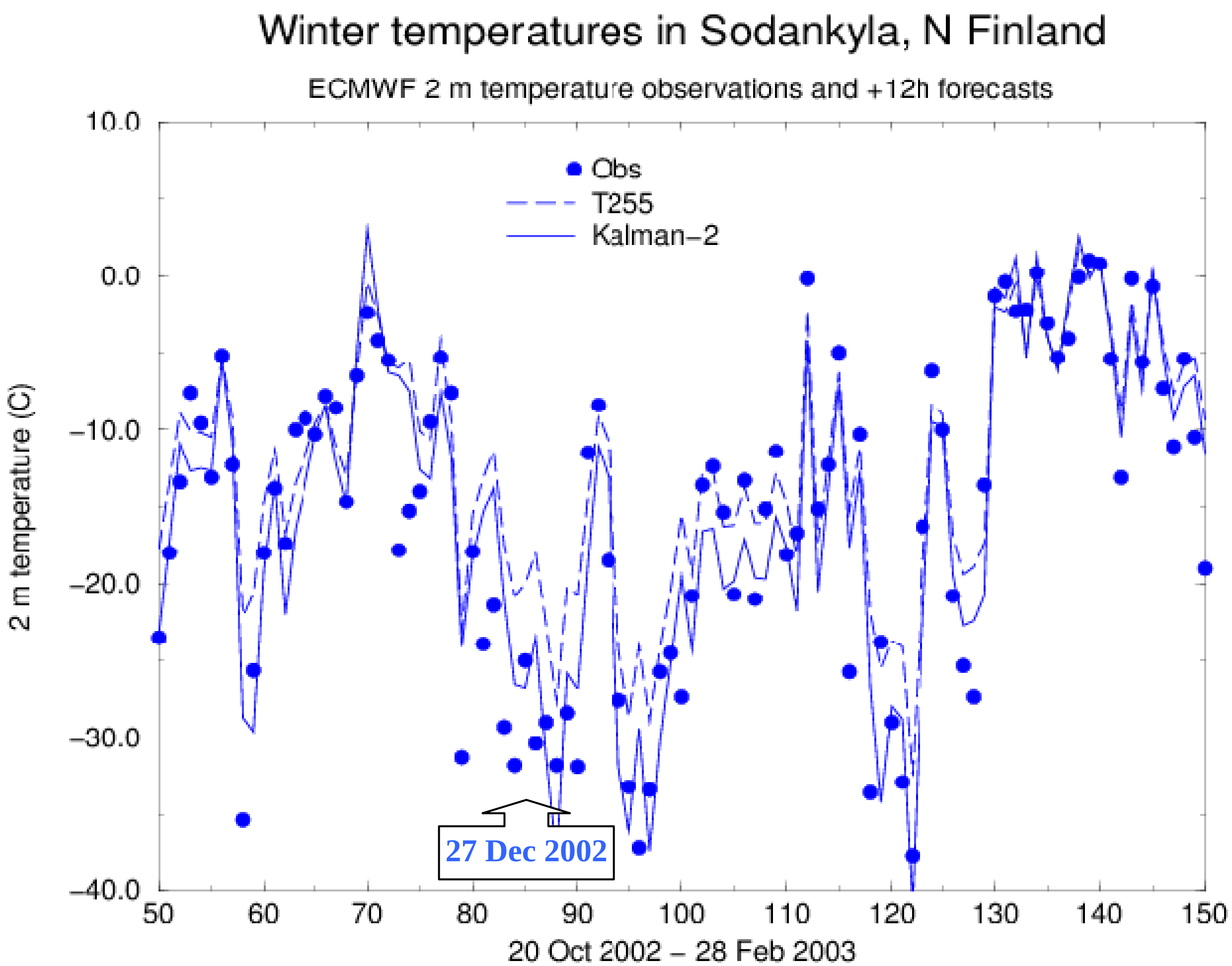


The T255 (and T511) have problems with temperatures below -25 C



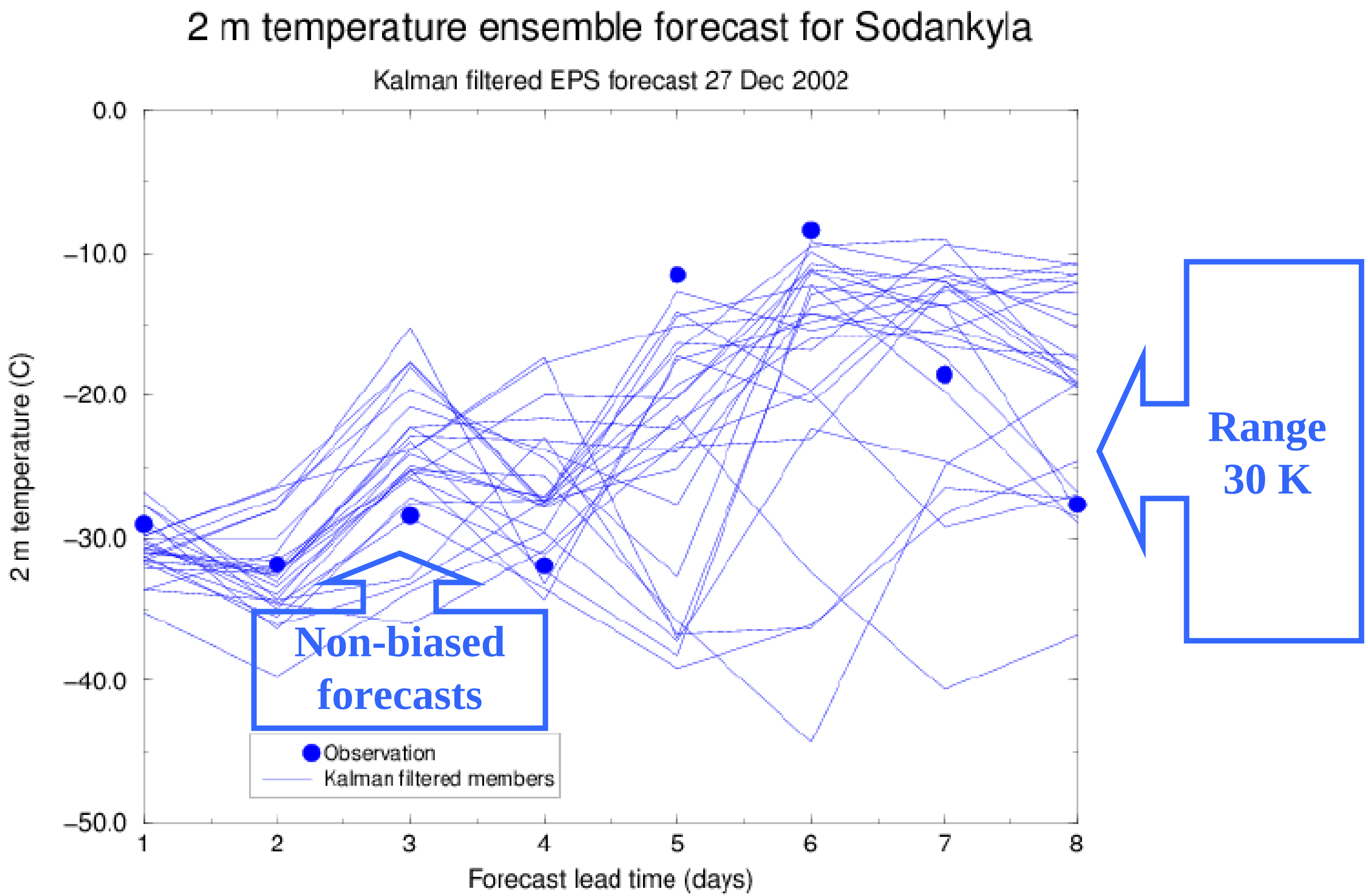


Winter temperatures in Sodankyla, N Finland
ECMWF 2 m temperature observations and +12h forecasts
Obs
T255
T511
2 m temperature (C)
10.0
0.0
–10.0
–20.0
–30.0
–40.0
50
60
70
80
90
100
110
120
130
140
150
20 Oct 2002 – 28 Feb 2003
27 Dec 2002

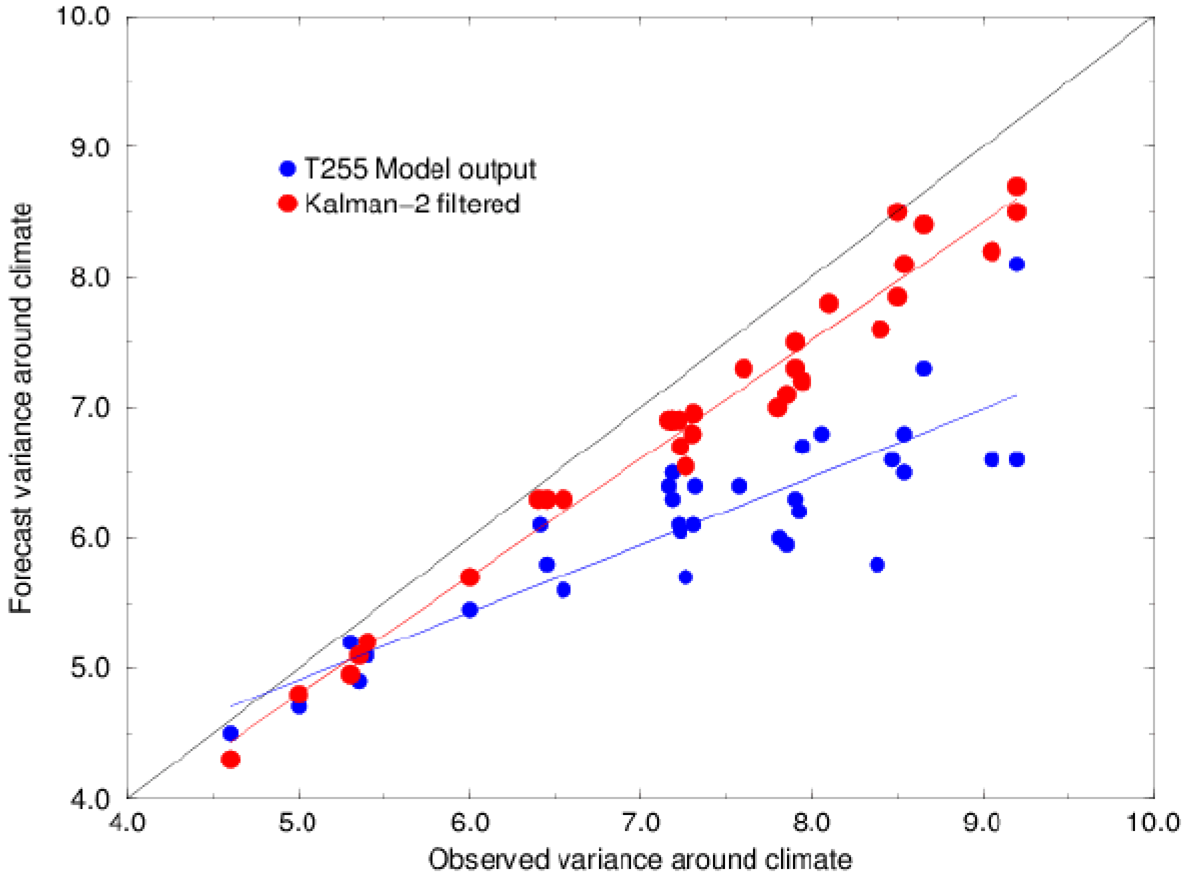


The mean error (ME) decreased by 1-2 K but the RMSE only by 0.5 K

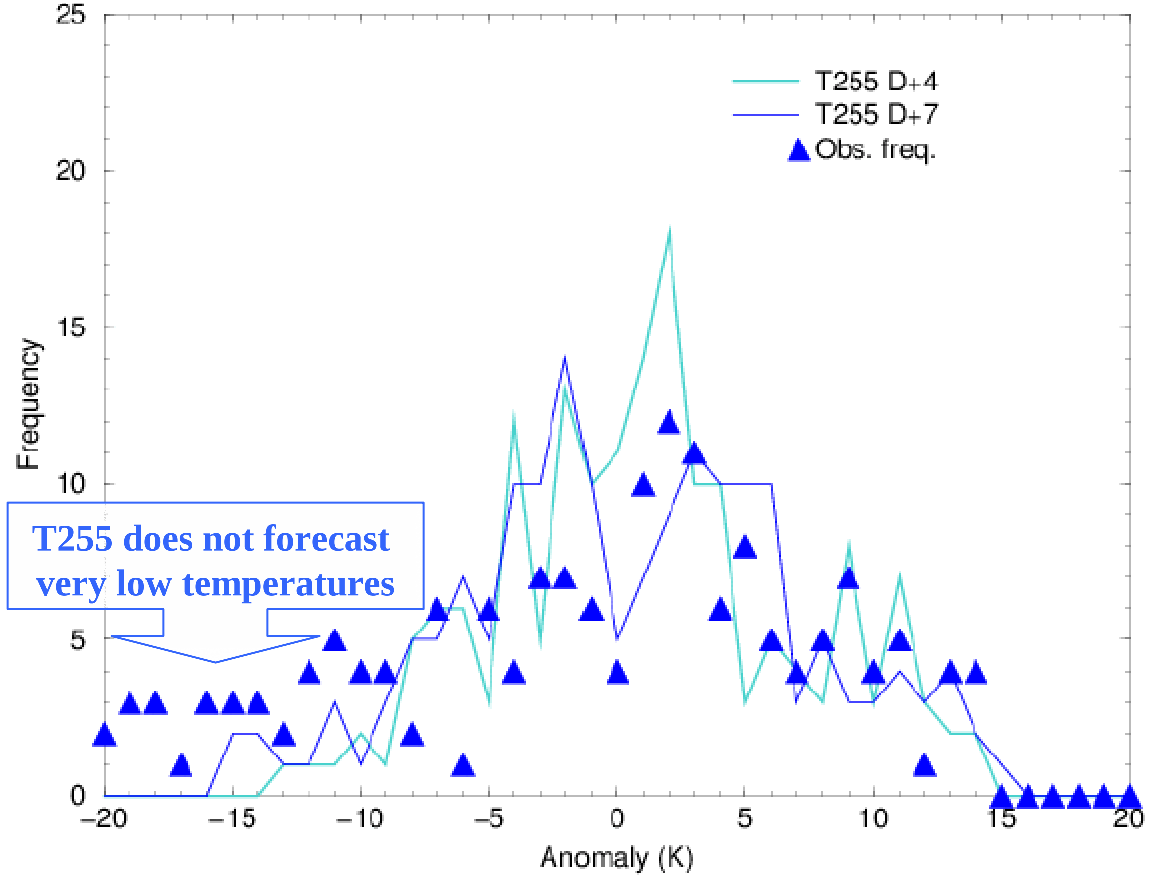
The solution lies in the ensemble approach...


2 m temperature ensemble forecast for Sodankyla
Kalman filtered EPS forecast 27 Dec 2002
0.0
–10.0
–20.0
–30.0
–40.0
–50.0
2 m temperature (C)
1
2
3
4
5
6
7
8
Forecast lead time (days)
Observation
Kalman filtered members
Non-biased forecasts
Range 30 K

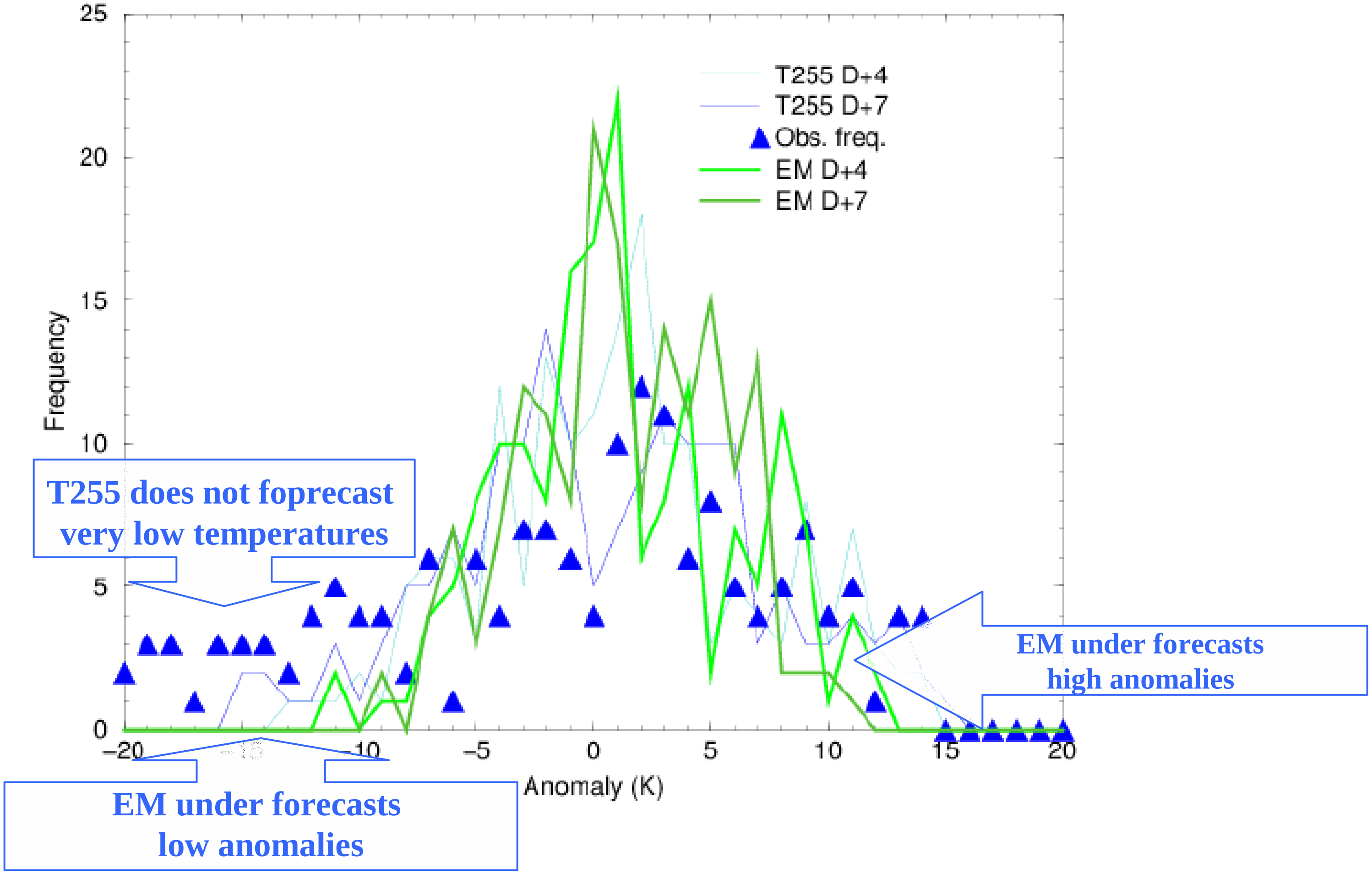
Variance around climate before and after Kalman filtering (00 UTC) for Finnish and Swedish stations in Lapland winter 2002-2003



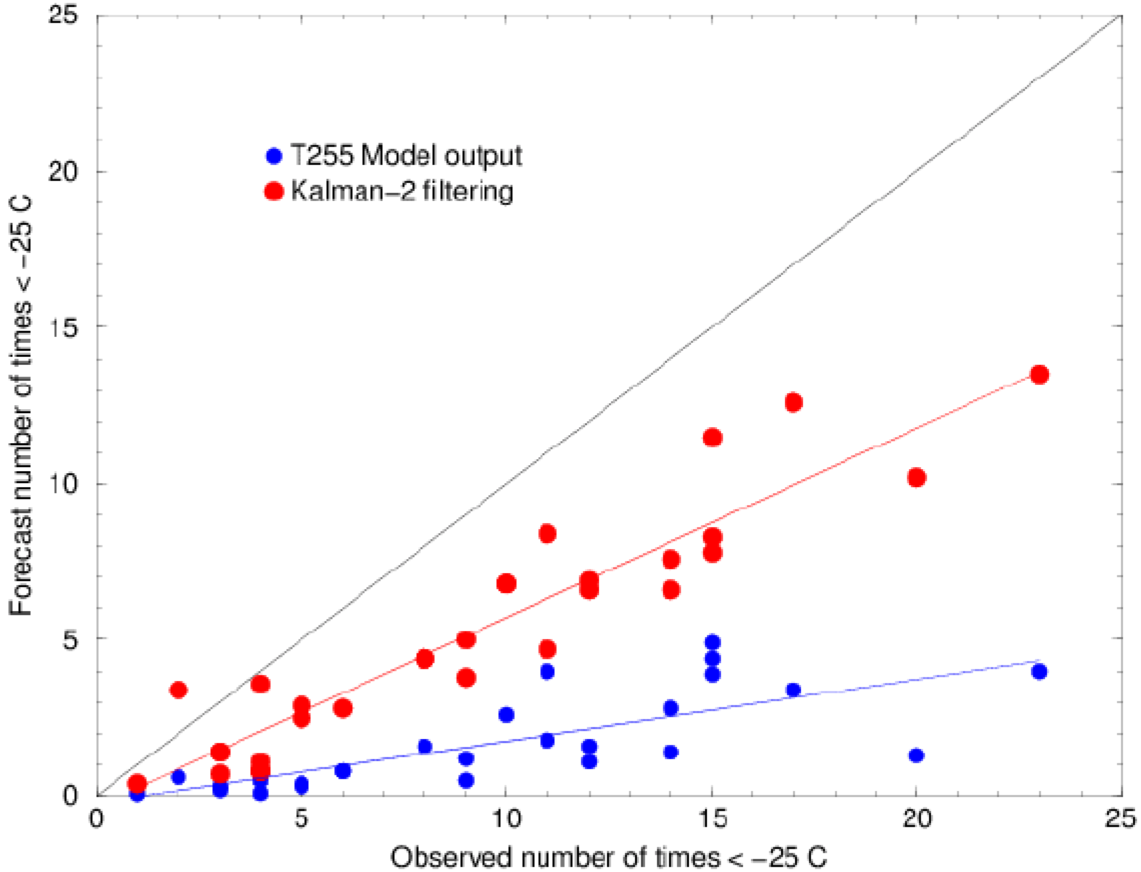
Range of observed and 255 (T511) forecast anomalies in Sodankyla winter 2002-2003



Range of observed and T255 EM forecast anomalies in Sodankyla winter 2002-2003



The range of the forecasts before and after Kalmanfiltering for Finnish and Swedish stations in Lapland wintern 2002-2003



Number of cases with < -25 C in Sodankylä

Winter 2001-2002

	Observed	T511	Kalman-2
00 z	20	2	13
12 z	12	1	10

Winter 2002-2003

	Observed	T511	Kalman-2
00 z	23	5	15
12 z	17	4	16

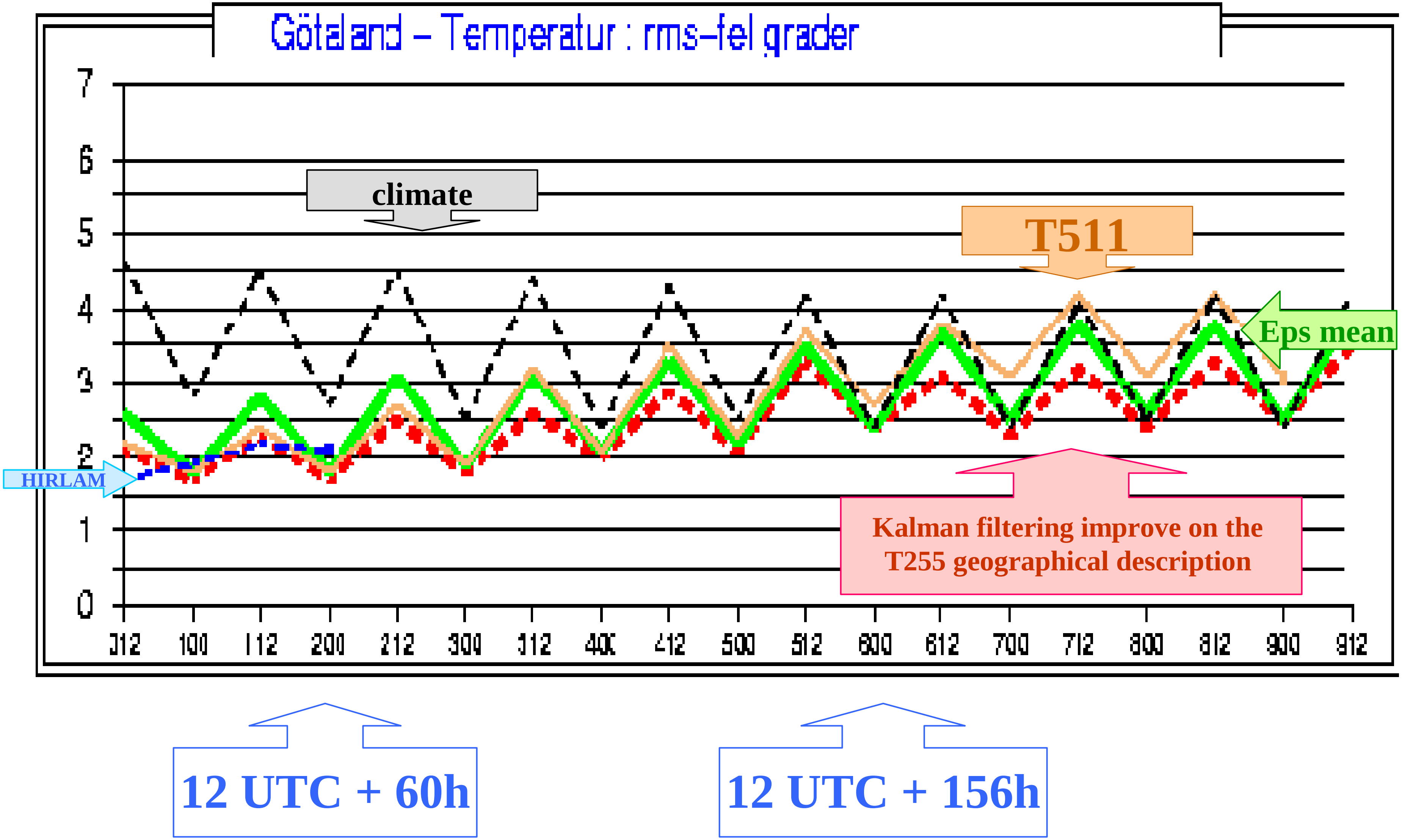
Number of cases < -25 C in Sodankylä

Winter 2001-2002

	Observed	T255	Kalman-2	EPS	Kalman-2
00 z	20	1	10	1.3	10.2
12 z	12	1	7	1.1	6.9

Winter 2002-2003

	Observed	T255	Kalman-2	EPS	Kalman-2
00 z	23	4	14	4.0	13.5
12 z	17	3.5	12.5	3.4	12.6


Götaland – Temperatur : rms–fel grader
7
6
5
4
3
2
1
0
climate
T511
Eps mean
HIRLAM
Kalman filtering improve on the T255 geographical description
012
100
112
200
212
300
312
400
412
500
512
600
612
700
712
800
812
900
912
12 UTC + 60h
12 UTC + 156h

The complete formula for RMSE

The full mathematical expression for the RMS error (E_j) of a **j**-day forecast issued on day **i** verified over **N** gridpoints over a period of **T** days

$$E_j = \sqrt{\frac{1}{T}\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}(f_{i,j} - a_{i+j})^2}$$

From the RMSE to the MSE

We make things easier for us by considering the ***square*** of the RMSE

$$E_j^2 = \frac{1}{T}\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}(f_{i,j} - a_{i+j})^2$$

Simplifying the notations

$$E_j^2 = \frac{1}{T}\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}(f_{i,j} - a_{i+j})^2$$

The notation is further simplified by replacing the Σs with an overbar symbolising all temporal and spatial averages. We also skip all the indices.

$$E^2 = \overline{(f - a)^2}$$

The power of simple mathematics

The equation looks trivial,

$$E^2 = \overline{(f-a)^2}$$

but reveals its deeper implication when considered in connection with the apparently equally "trivial"

$$(a+b)^2 = a^2 + b^2 + 2ab$$

Decomposing the MSE around climate

Decomposing the RMSE

$$E^2 = \overline{(f-a)^2}$$

Introduce **c** as the climate value of the verifying day

$$E^2 = \overline{(f-c+c-a)^2}$$

Reposition **c** to form **f-c** and **a-c**

$$E^2 = \overline{((f-c)-(a-c))^2}$$

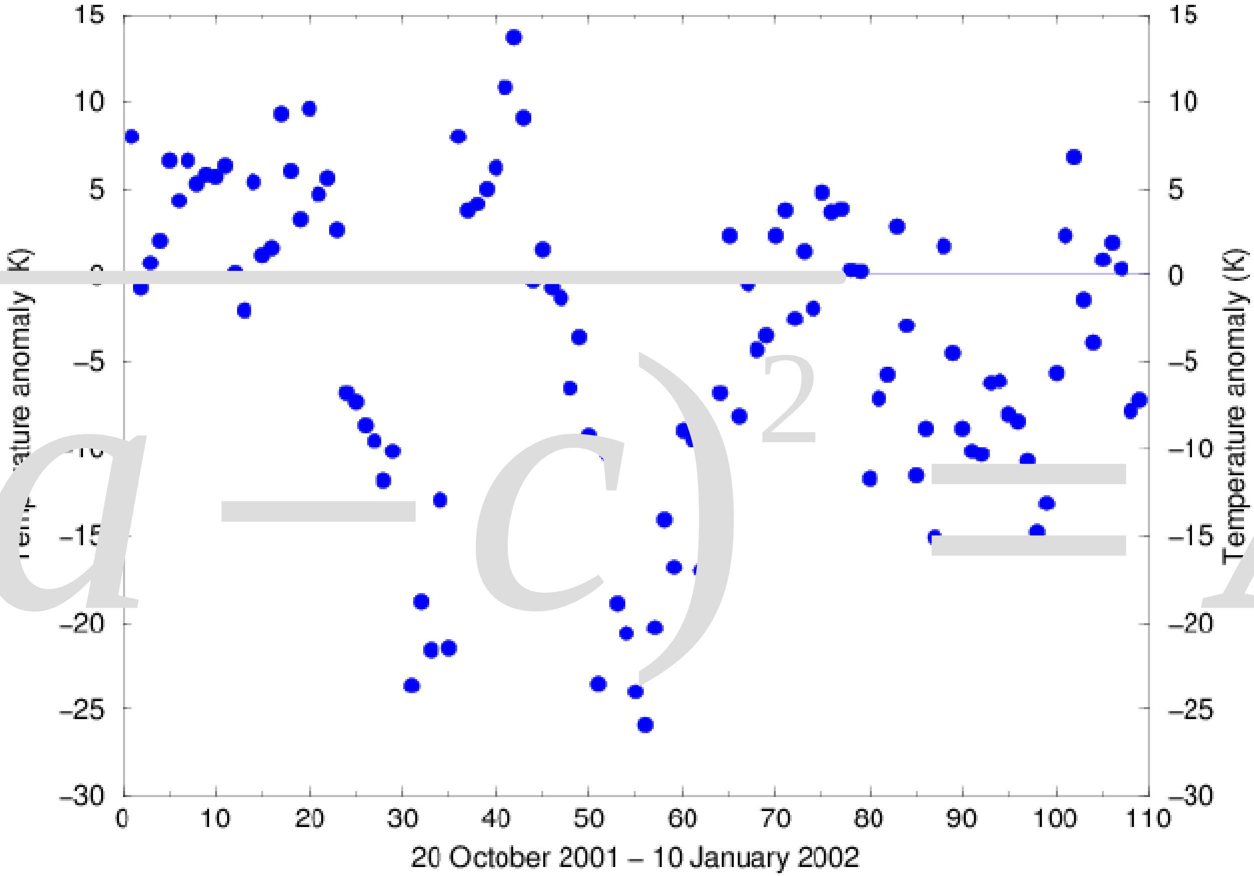
Apply $\mathbf{(a+b)^2=a^2+b^2+2ab}$

$$E^2 = \overline{(f-c)^2} + \overline{(a-c)^2} - 2\overline{(f-c)(a-c)}$$

Each of these three terms has its own story to tell

The interpretation of the first term

The *observed* variability around the climatological mean

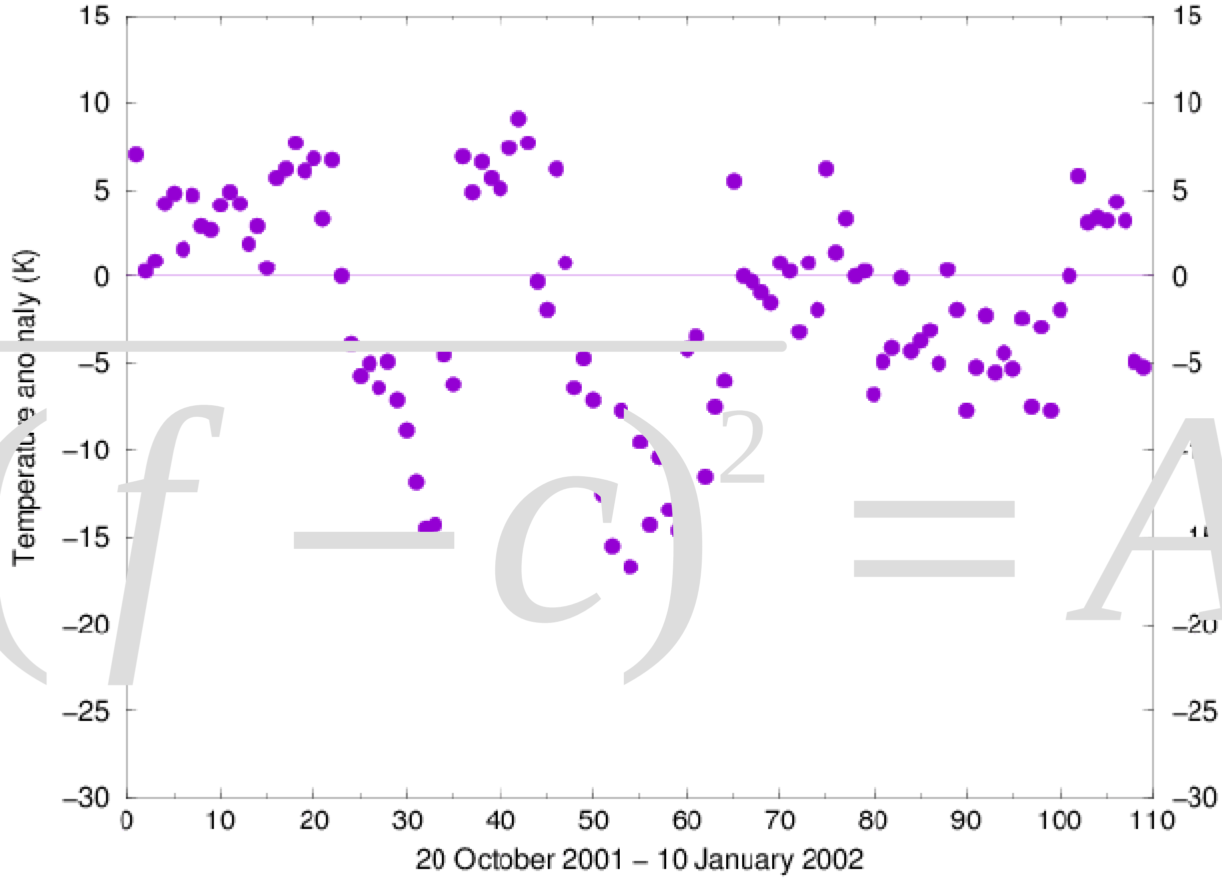


$$(a-c)^2 = A_a^2$$

The magnitude of this term can not be affected by human intervention

The interpretation of the second term

The *forecast* variability around the climatological mean

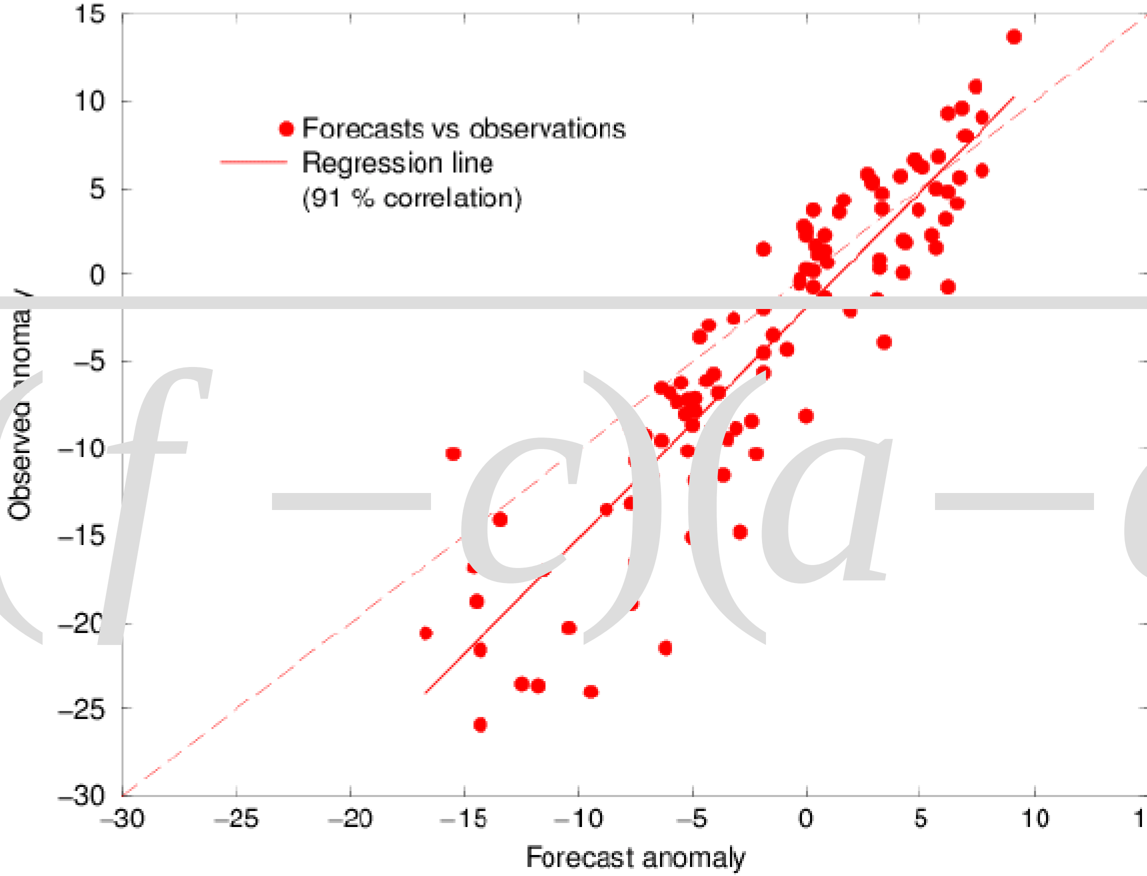


$$\overline{(f-c)^2} = A_f^2$$

The magnitude of this term can indeed be affected by human intervention

The interpretation of the “skill” term

The correspondence between f-c and a-c



$$\overline{2(f-c)(a-c)}$$

This is the ***only*** term in the RMSE decomposition which is related to the predictive skill of the model

What looks bad might be good...

The *improvement* of the model, as a simulation of the atmospheric system, may therefore appear as *deterioration* of the quality of the model!

An increase in forecast variability increases $\mathbf{A_f}^2$

...compensates the decrease of the RMSE due to improved forecasts

$$E^2 = \overline{(f-c)^2} + \overline{(a-c)^2} - 2\overline{(f-c)(a-c)}$$

...to the level of the observed variability $\mathbf{A_a}^2$

The accuracy of a climate statement

When **f=c** the first and last terms disappear

$$E^2 = \overline{(f-c)^2} + \overline{(a-c)^2} - 2\overline{(f-c)(a-c)}$$

$$E^2 \rightarrow \overline{(a-c)^2}$$

$$E^2 \rightarrow A_a^2$$

We take the square root….

$$E \rightarrow A_a$$

Which is the error level for a purely climatological statement

The RMS error saturation level

When the forecasts start to lose skill and the RMSE start to approach high error levels the last term disappears

$$E^2 = \overline{(f-c)^2} + \overline{(a-c)^2} - 2\overline{(f-c)(a-c)}$$

$$E^2 \rightarrow A_f^2 + A_a^2$$

$$E^2 \rightarrow 2A_a^2$$

$$E \rightarrow A_a\sqrt{2}$$

Which is the uppermost error level for a realistic NWP model, 41% above the error level of a purely climatological statement

It is also called *The Error Saturation Level (ESL)*