II. Frequentist probabilities

II.4 Statistical interpretation or calibration

II.4.1 What is statistical interpretation doing?

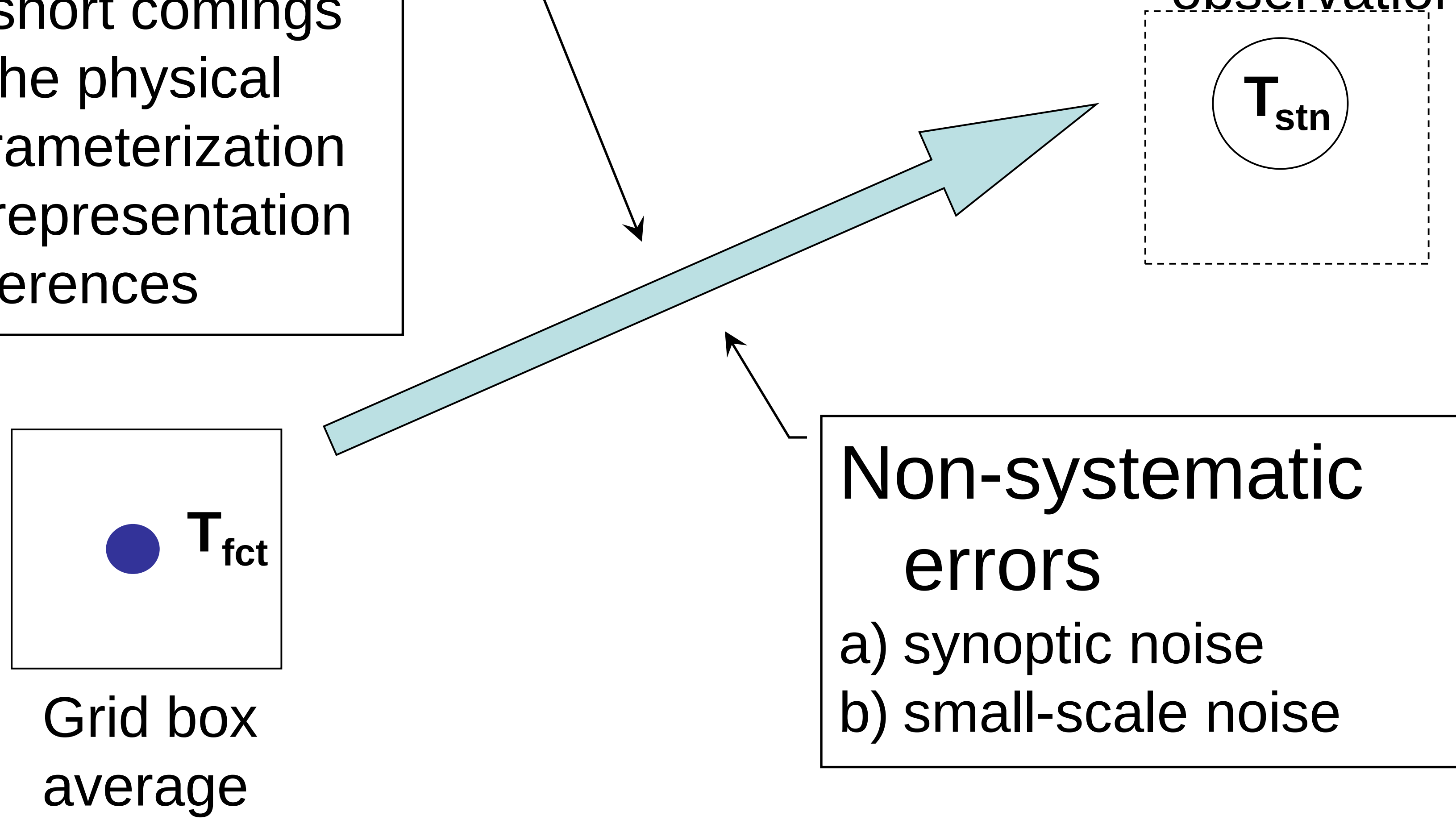
In light of a (non-perfect) forecast performance corrections are applied in order to improve the performance.

This applies to both systematic <u>and</u> non-systematic forecast errors

What we do

Systematic errors

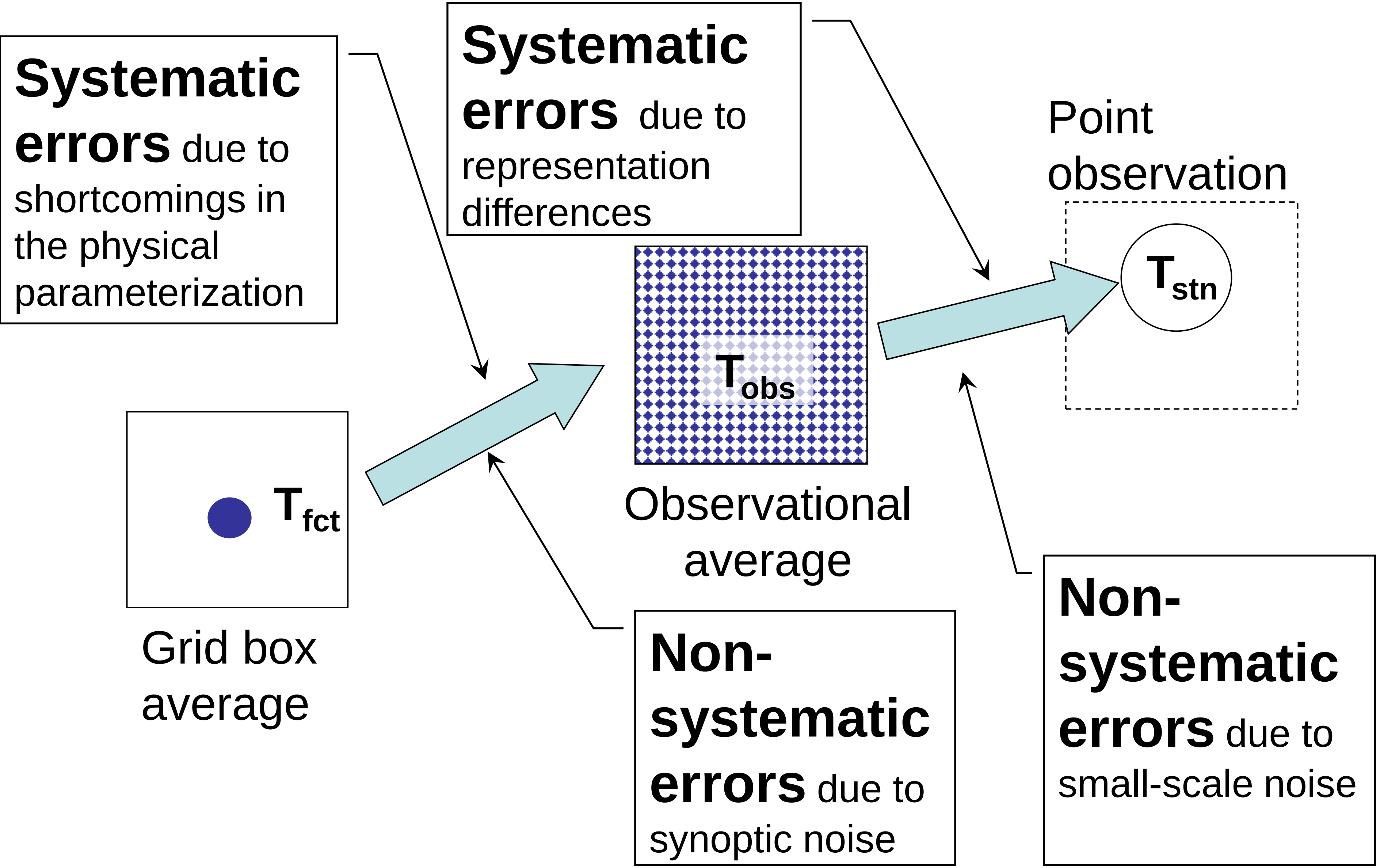
a) short comings in the physical parameterization
b) representation differences



Non-systematic errors

a) synoptic noise
b) small-scale noise

What we should do:



II.4.2 Biases and systematic errors

Common misconception:

-"Systematic errors" – that is the same as biases!

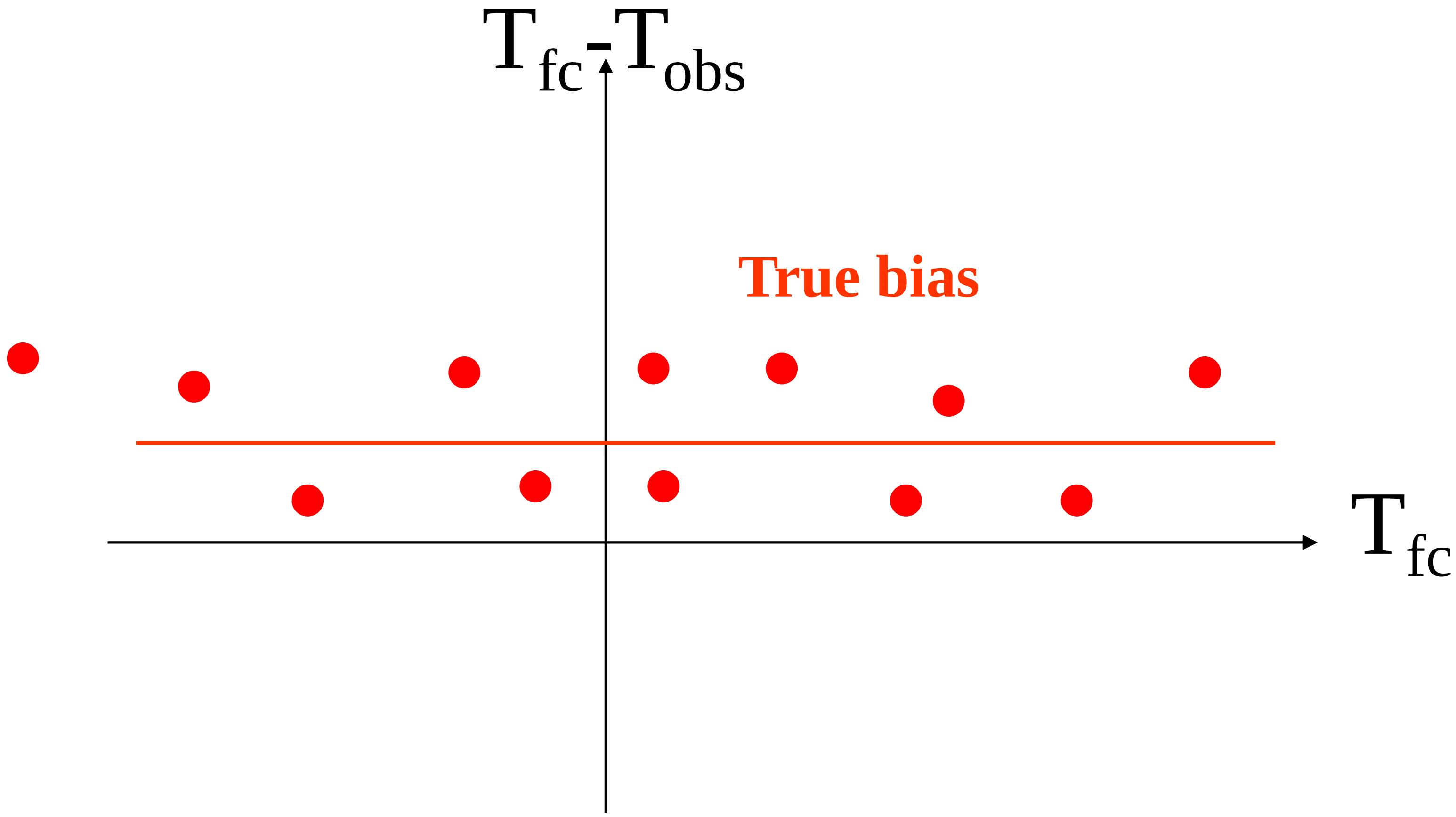
Definitions (according to AP!):

a) **Bias** = mean forecast error (independent of any parameter)

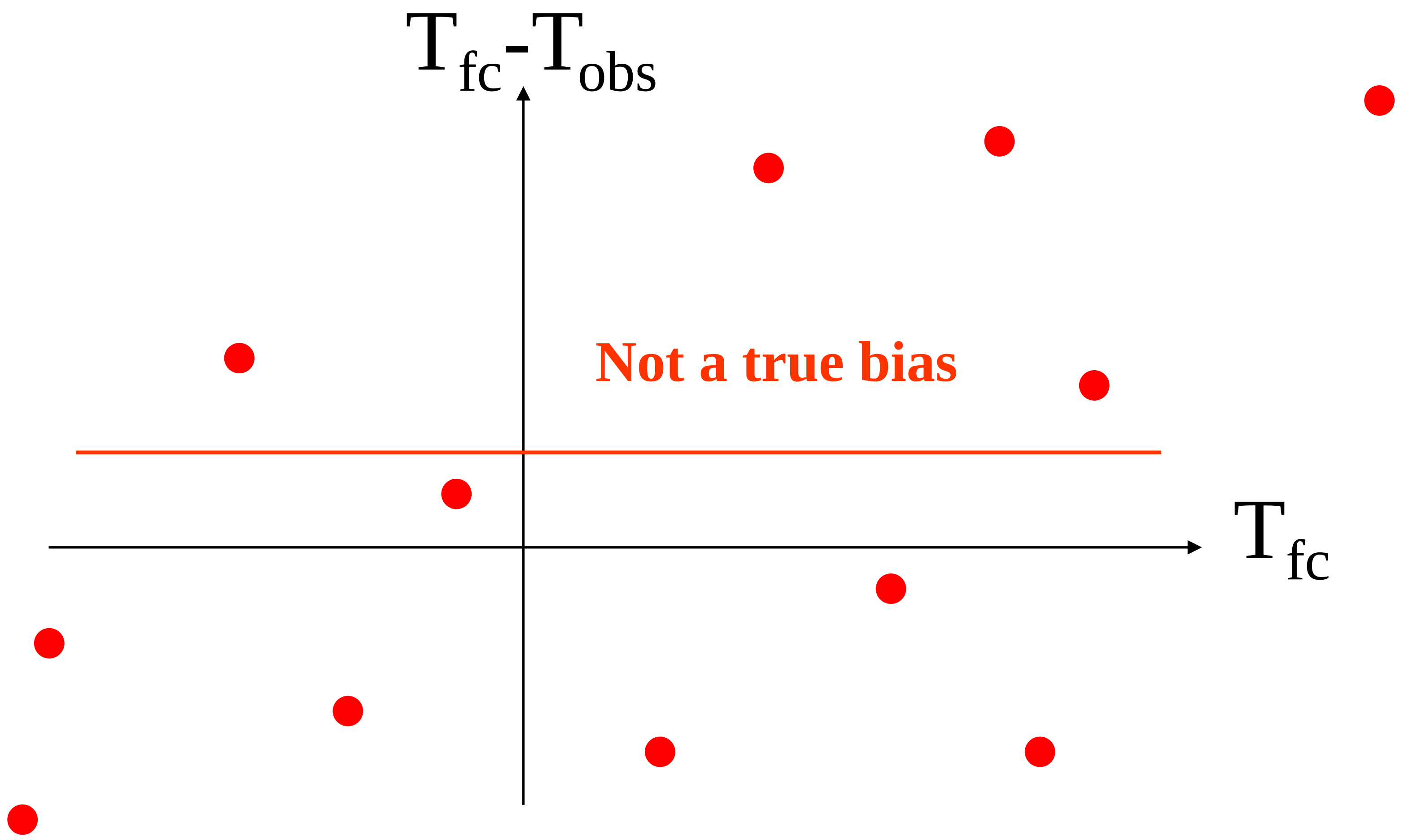
b) **Systematic error** = forecast error which can be mathematically modelled

c) **Non-systematic error** = forecast error which it has not (yet) been possible to model mathematically

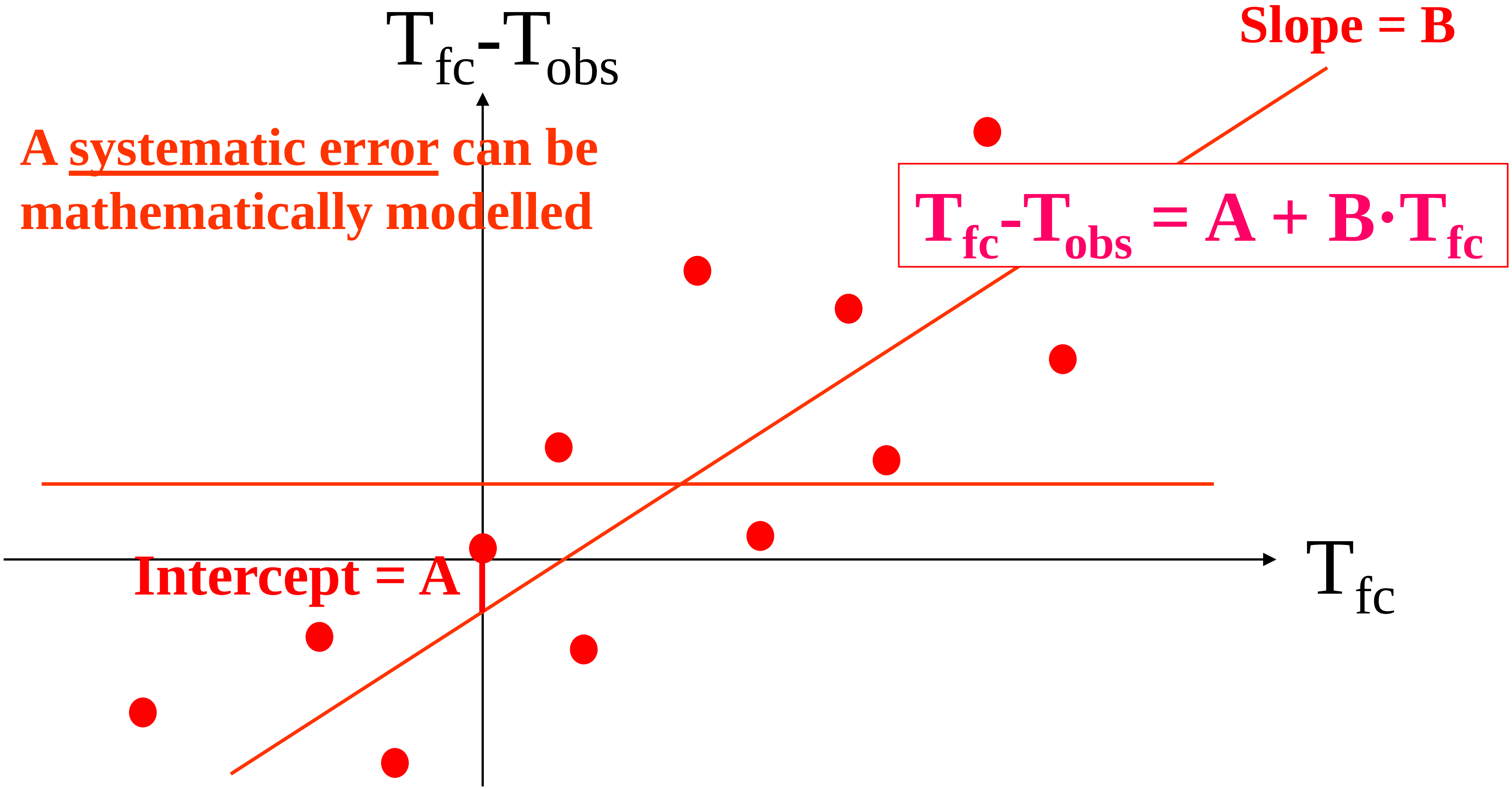
A “bias” or “true bias” when the mean error is quasi-constant independent of the forecast



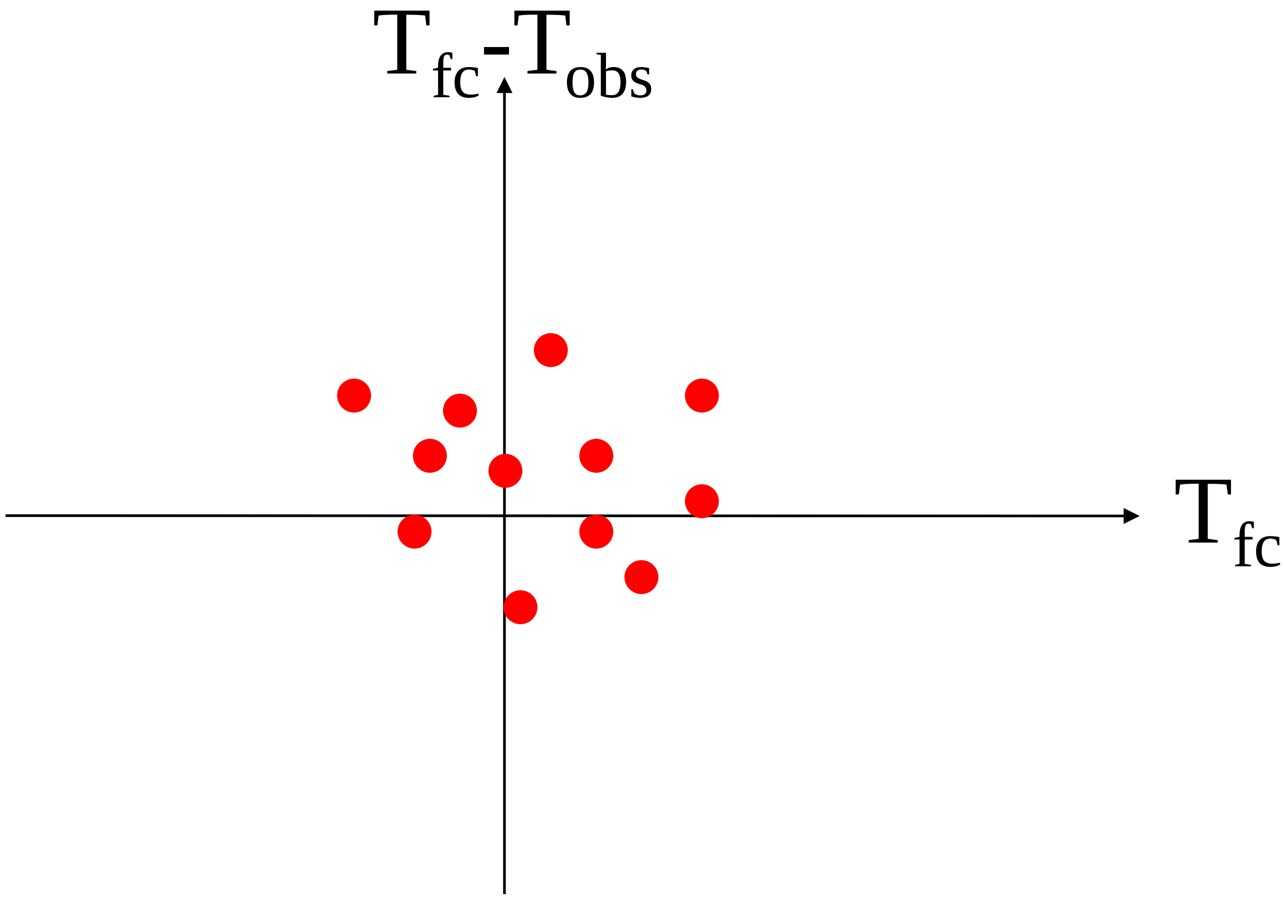
The same mean error as in the previous example but should not be called “bias” but just “mean error”



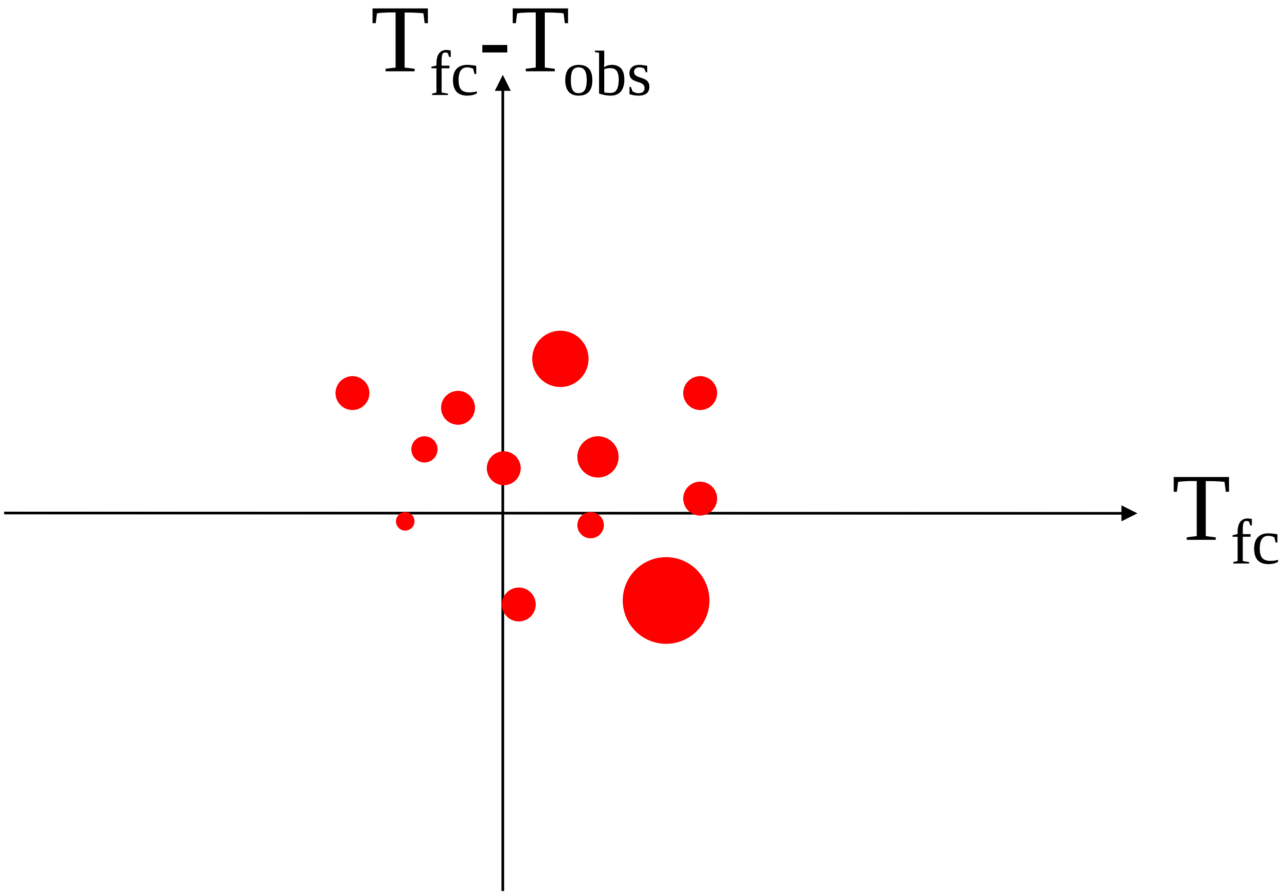
The same mean error as in the previous examples but not a "bias" but a "systematic error"



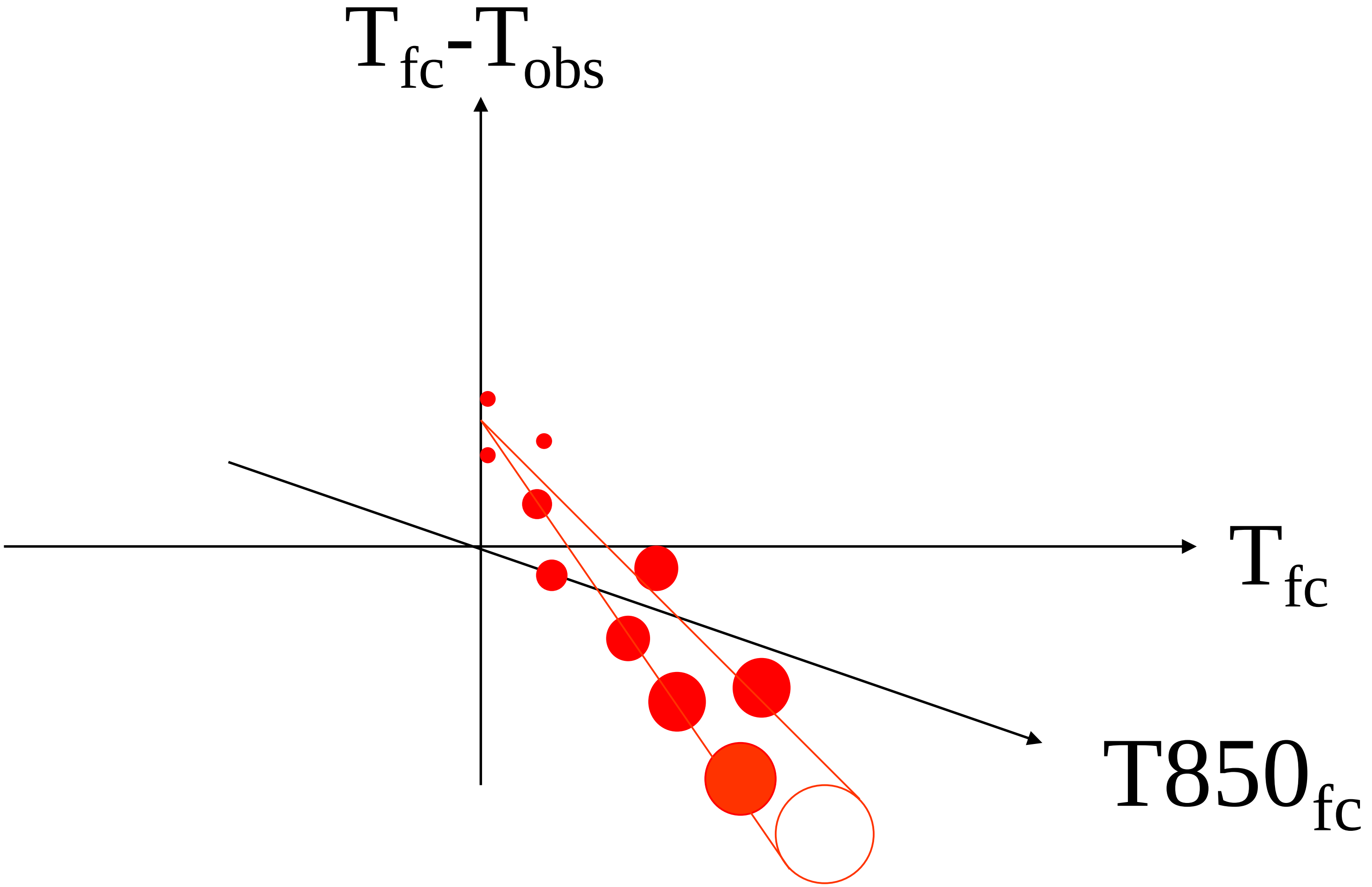
Apparent non-systematic errors..



But represented in three dimensions they might appear as the stars, some close, some far away . . .



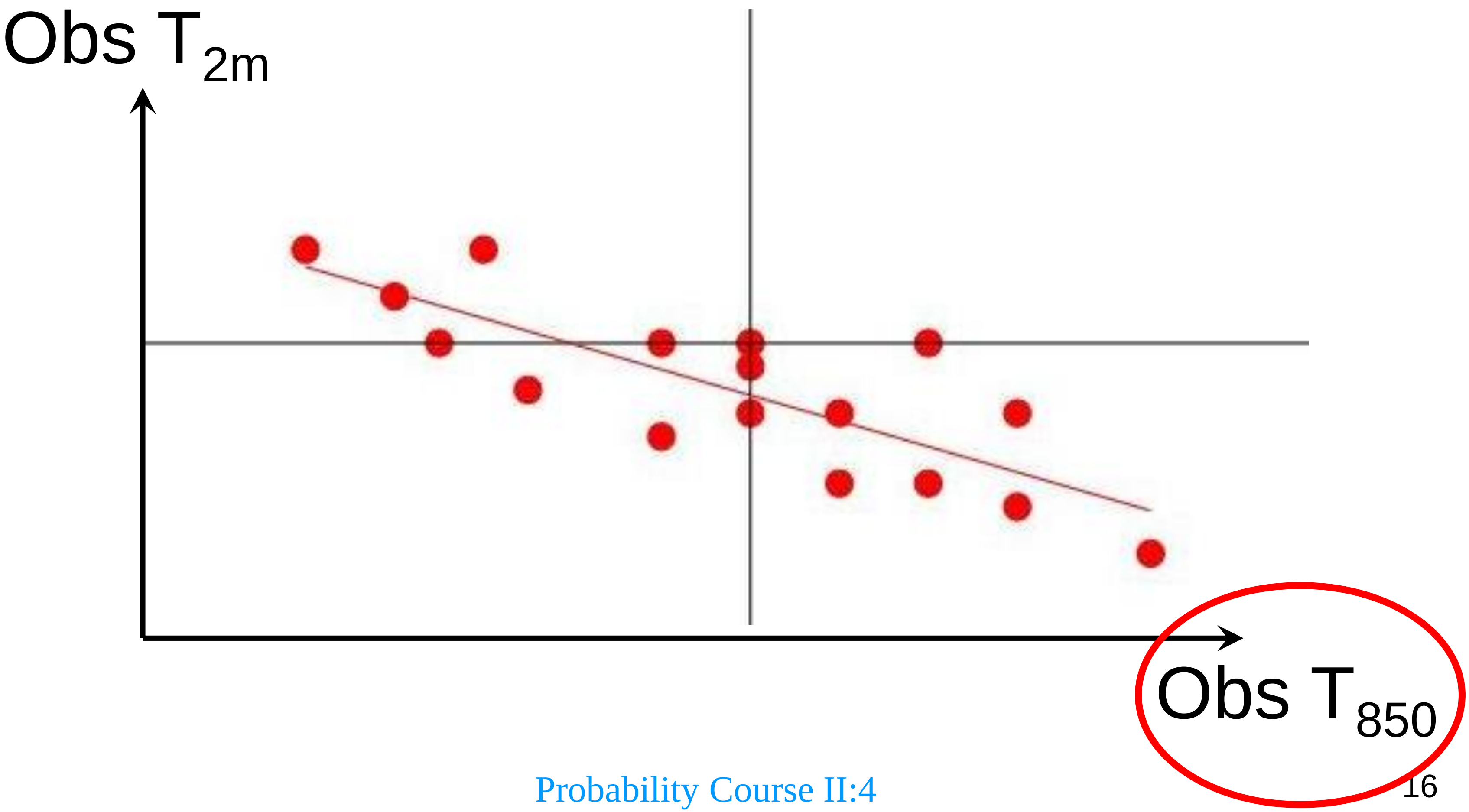
Projected into an additional dimension the errors appear to be systematic



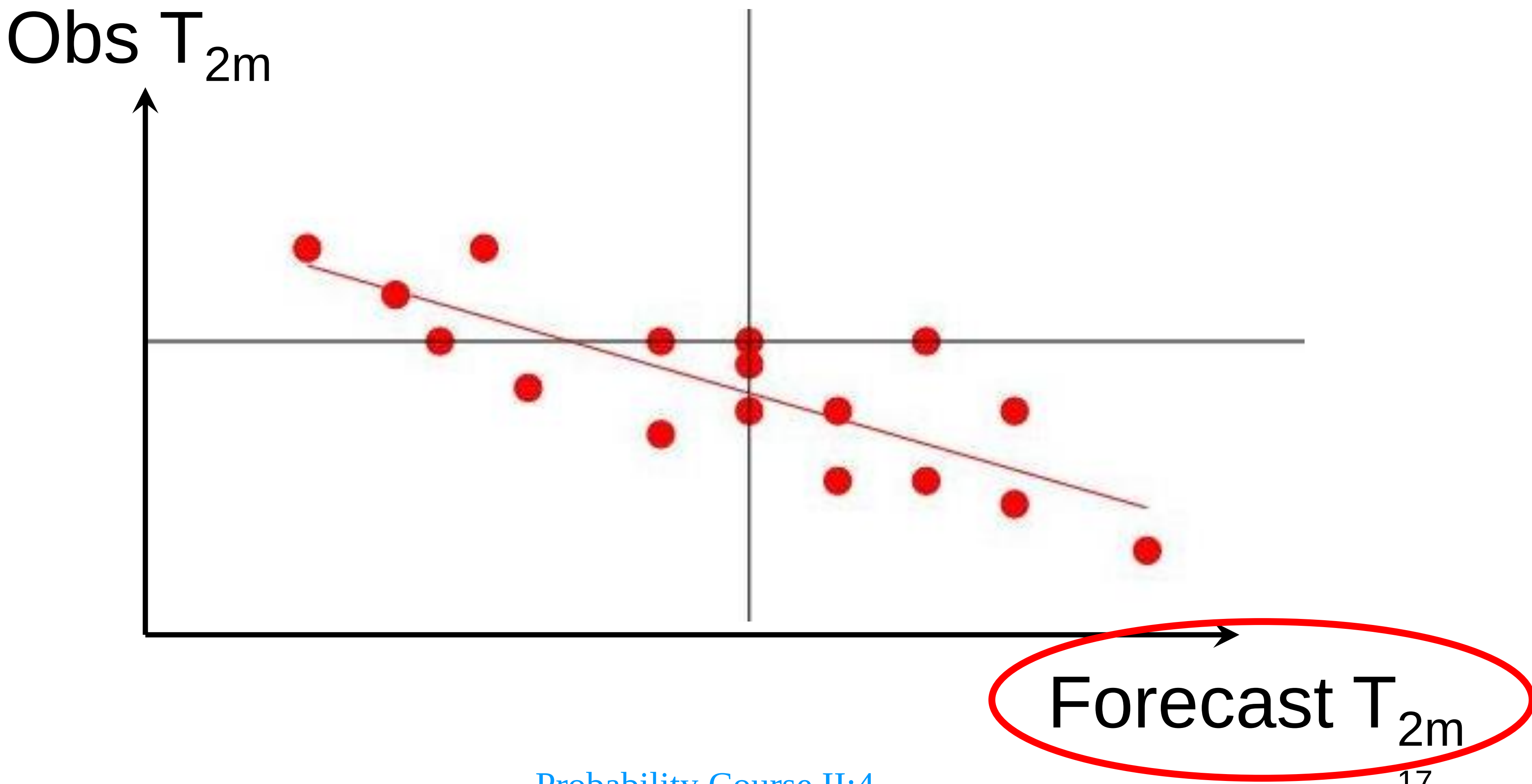
II.4.3 Different methods of statistical interpretation

The most commonly used statistical interpretation schemes assume that forecast errors can be linearly modelled

Perfect Prog Method (PPM) correlates analysed or observed values of easily predicted parameters with observations of a (often) less easily predicted weather parameter of interest



Model Output Statistics (MOS) and Adaptive Methods correlate values of different forecast parameters with observations of the weather parameter of interest



1. PPM can be based on regression analysis from historical data bases, preferably 30-50 years back,

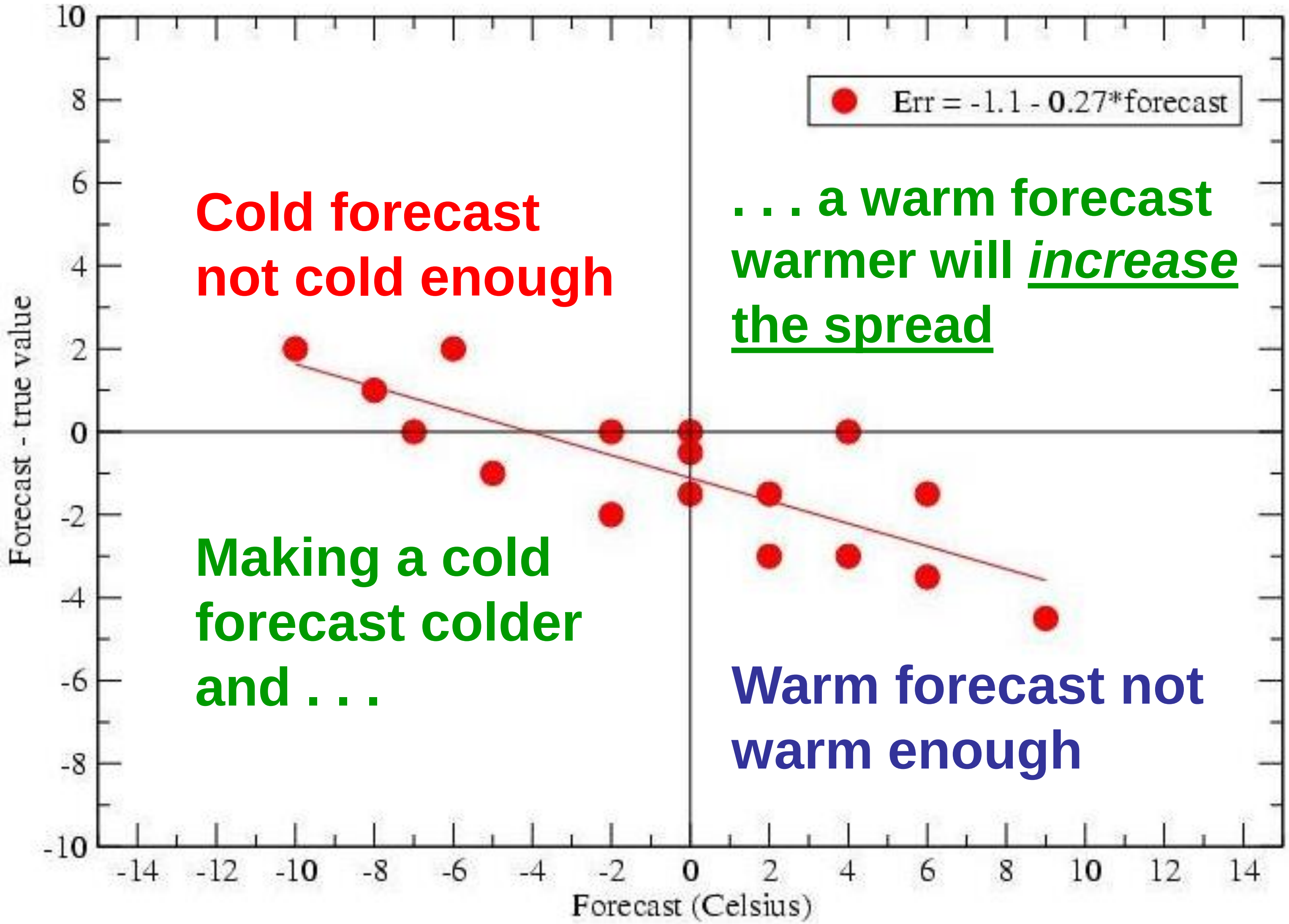
2. MOS can in principle be based on the same, but in practise only 3-5 years back, since the NWP model change their statistical characteristics.

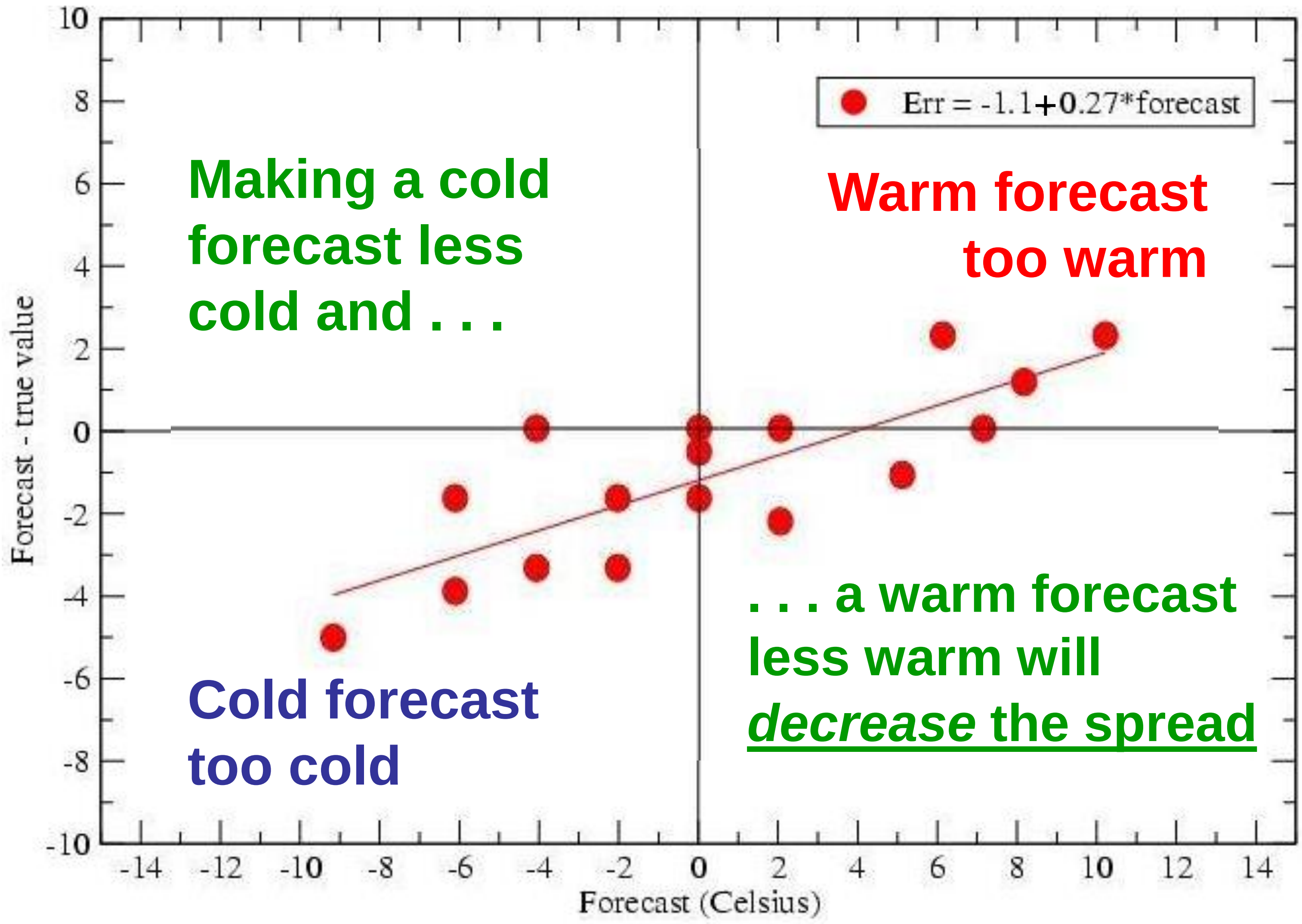
3. Adaptive methods do not use any historical data at all. Instead they make use of the most recent verifications as e.g. the simple running average system:

New correction = 0.9·yesterday's correction – 0.1·last error

II.4.4 The statistical interpretation schemes can also change the activity (spread)

A simple example of a linear regression model

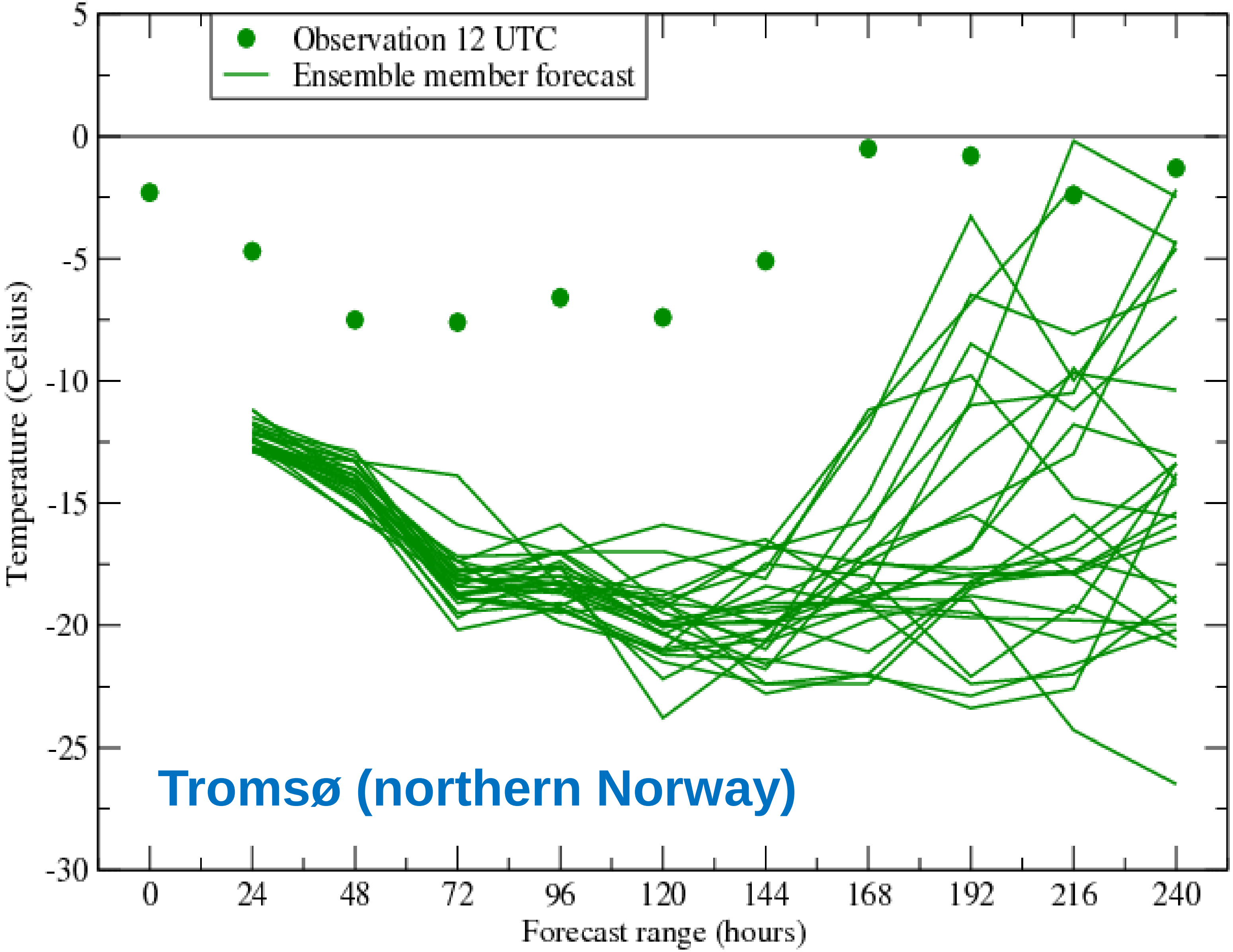


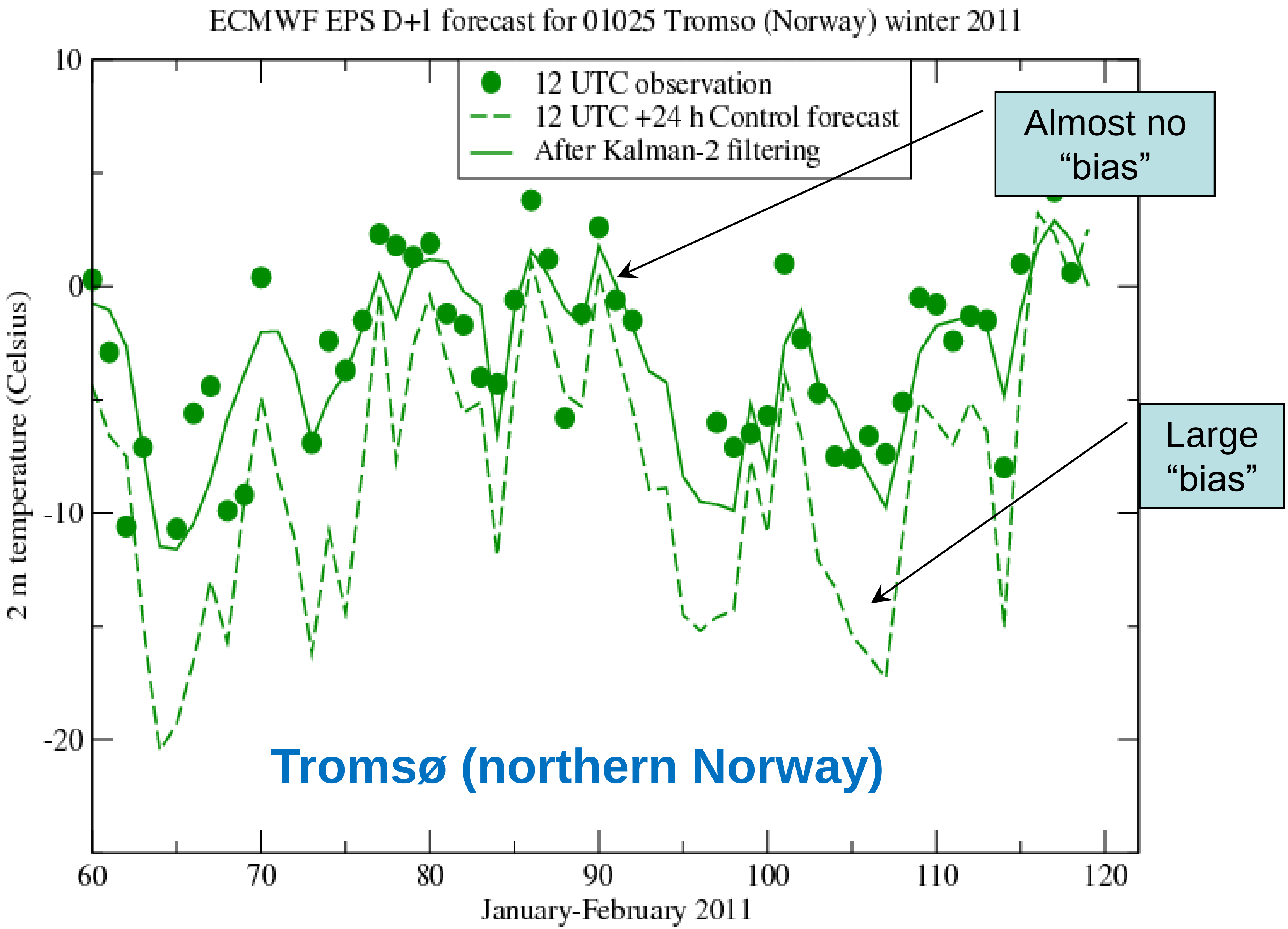


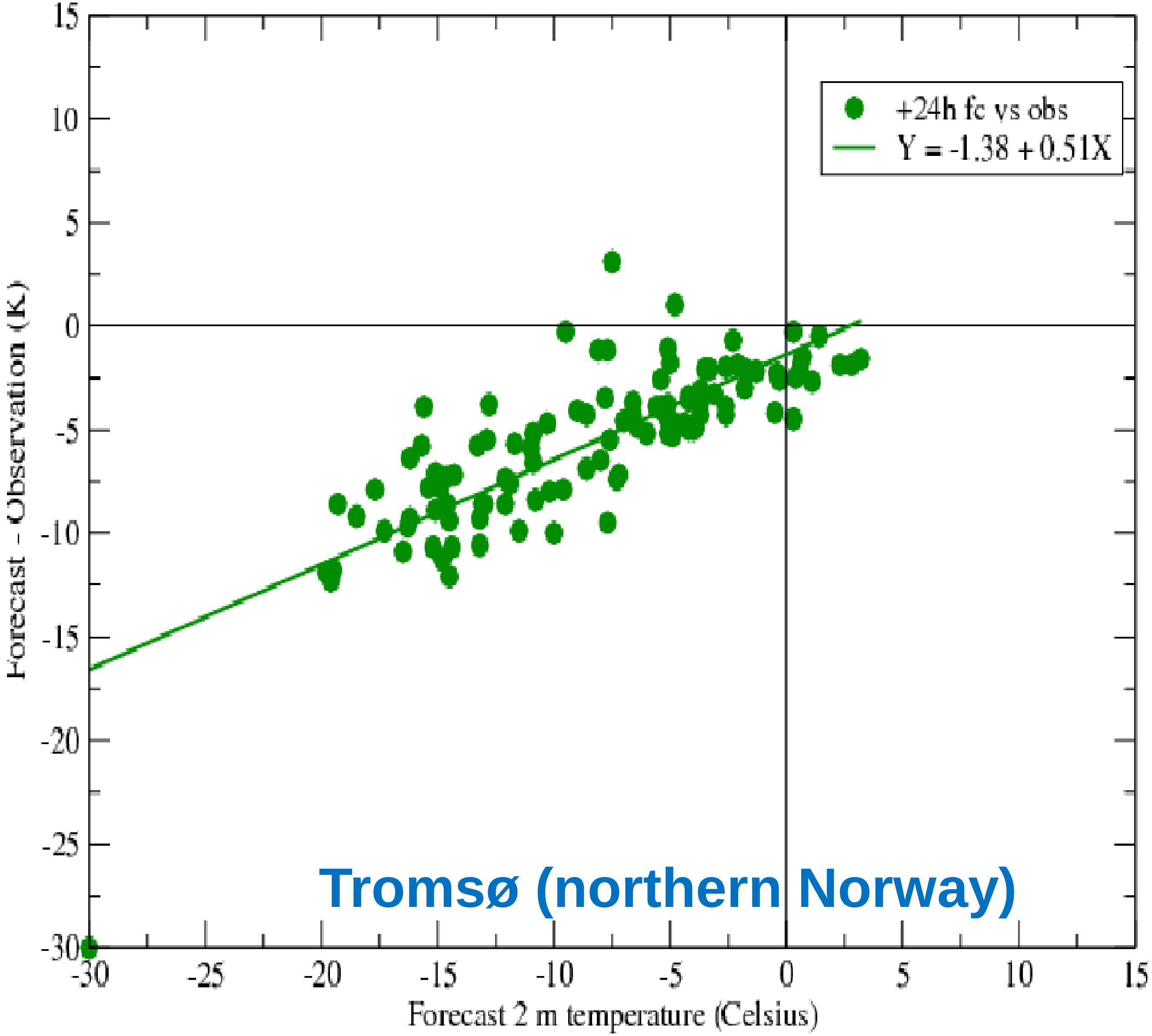
Err = -1.1+0.27*forecast
Making a cold forecast less cold and . . .
Warm forecast too warm
. . . a warm forecast less warm will *decrease* the spread
Cold forecast too cold
Forecast - true value
Forecast (Celsius)
10
8
6
4
2
0
-2
-4
-6
-8
-10
-14
-12
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
14

II.4.5 A practical example

A heavily "biased" temperature ensemble forecast

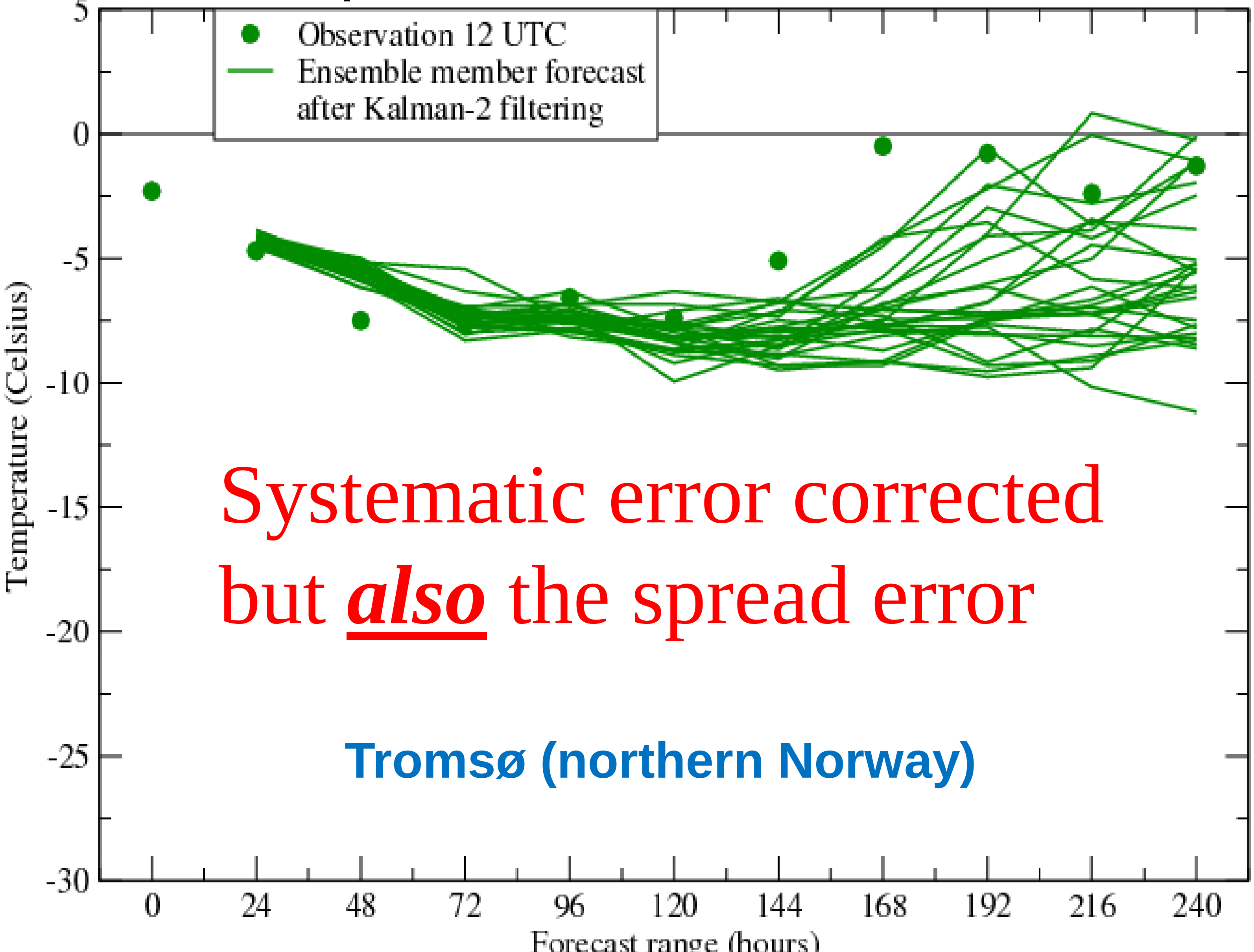



ECMWF EPS D+1 forecast for 01025 Tromso (Norway) winter 2011
12 UTC observation
12 UTC +24 h Control forecast
After Kalman-2 filtering
Almost no "bias"
Large "bias"
Tromsø (northern Norway)
2 m temperature (Celsius)
January-February 2011
10
0
-10
-20
60
70
80
90
100
110
120



No simple, straight bias. The mean error depends on the forecast

The EPS plume after statistical correction



Non-systematic errors cannot be removed – only NWP model and analysis improvements can achieve that

But they can be modified:

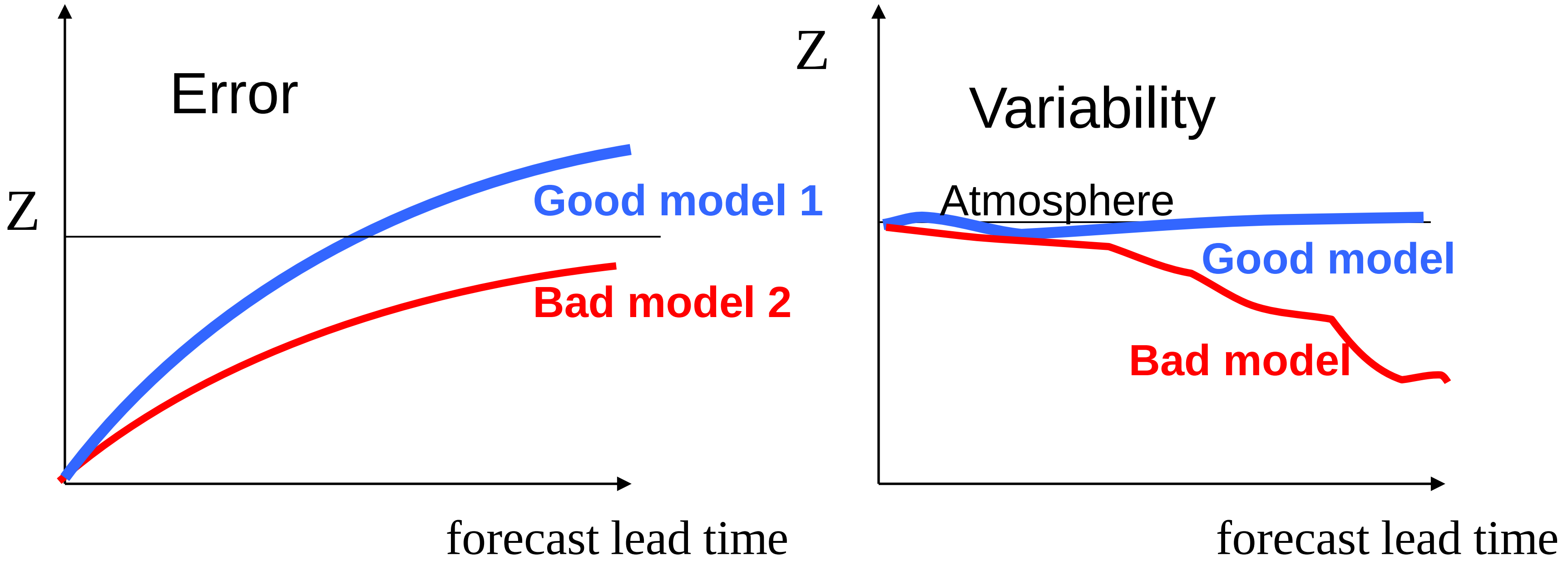
-Reducing the spread (variability) will dampen (decrease) the non syst. errors

-Increasing the spread (variability) will amplify (increase) the non syst. errors

II.4.6 Variability and error

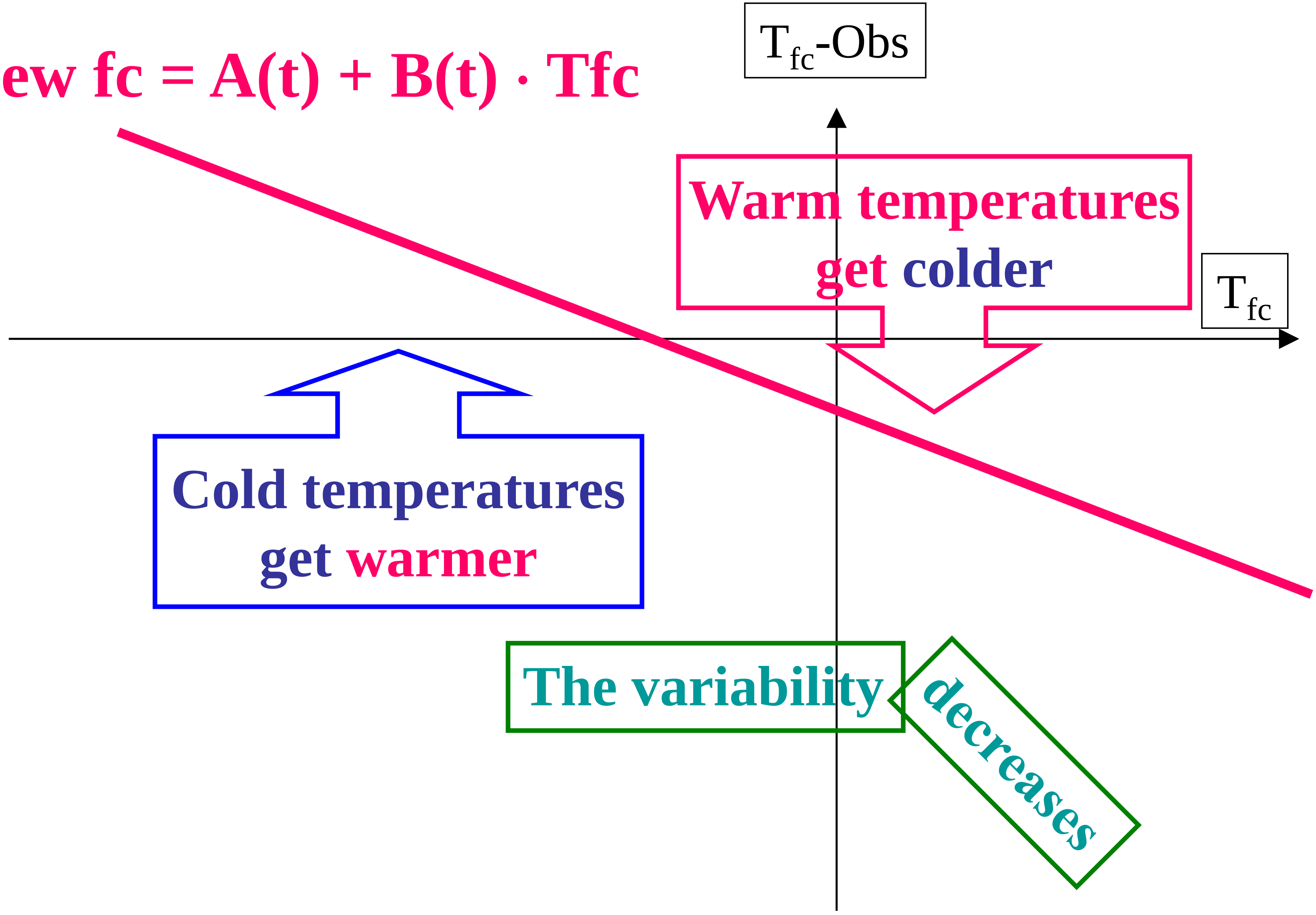
(a repeat)

A bad NWP model with under-variability might have lower RMSE than . . .

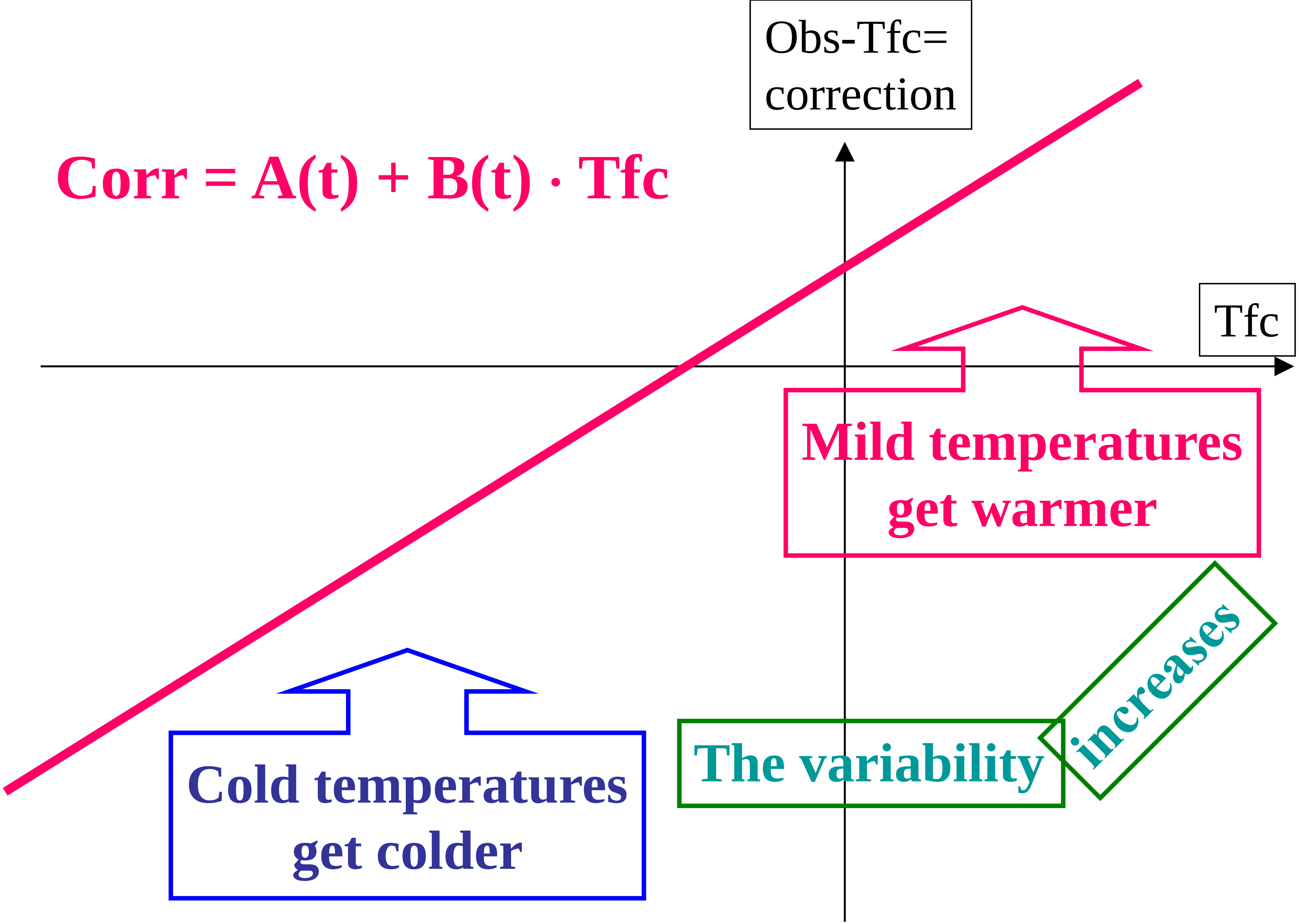


. . . a good NWP model with correct variability and therefore higher RMSE

New fc = A(t) + B(t) · Tfc

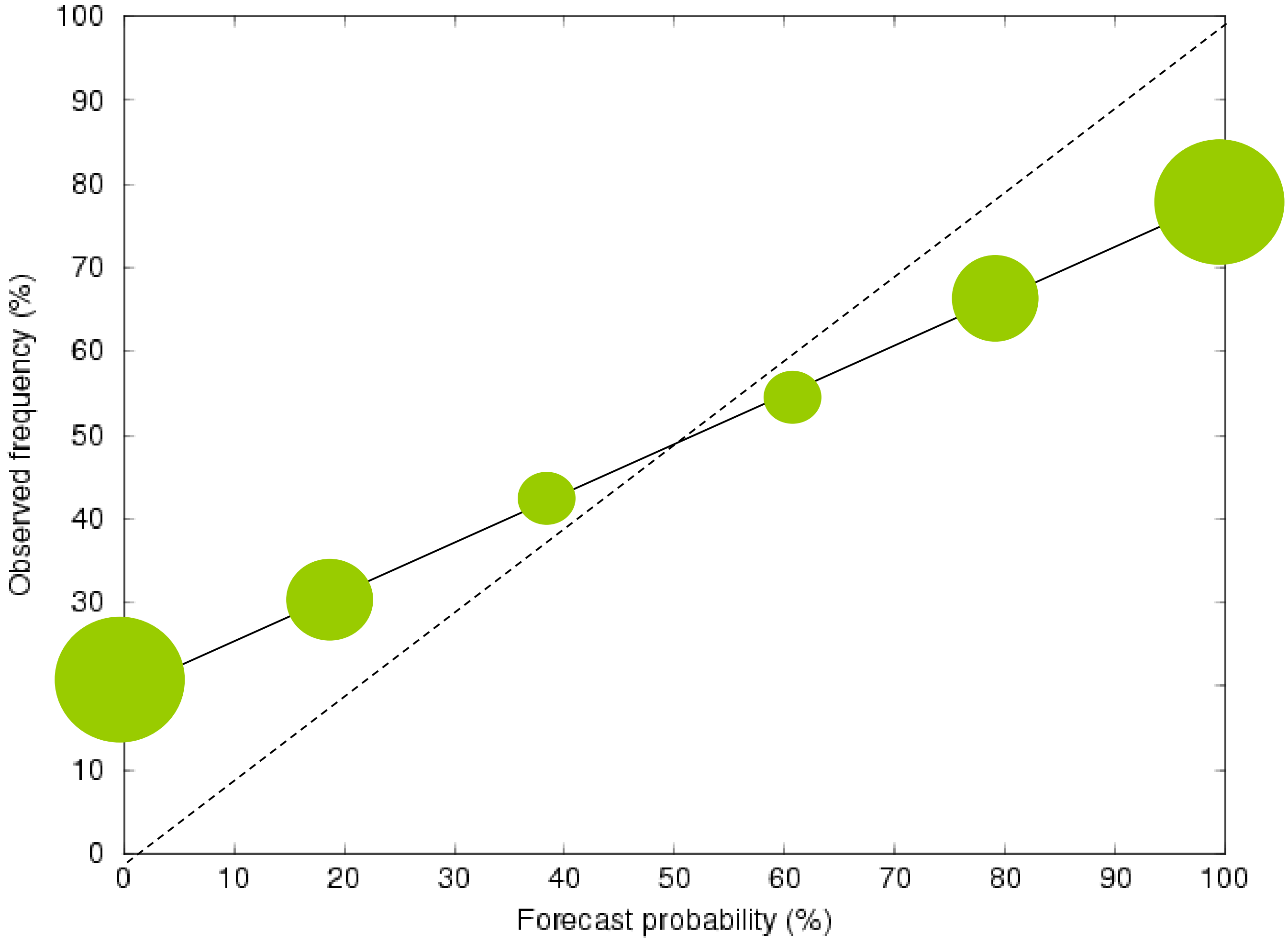


$$Corr = A(t) + B(t) \cdot Tfc$$



II.2.7 Calibration of typical probability forecast errors

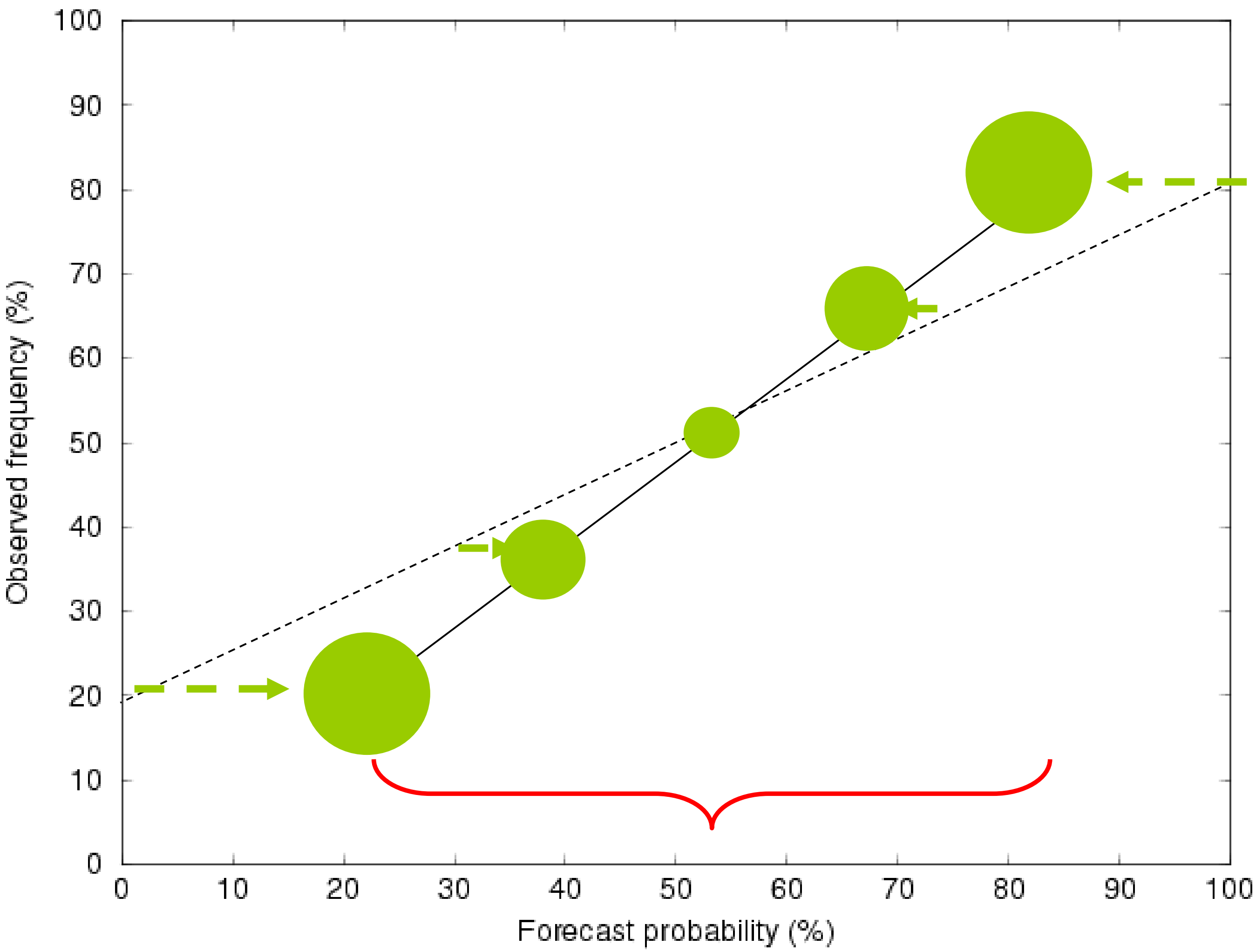
Over-confidence: a typical outcome of operational probability forecasting (forecasters, statistical schemes and EPS)



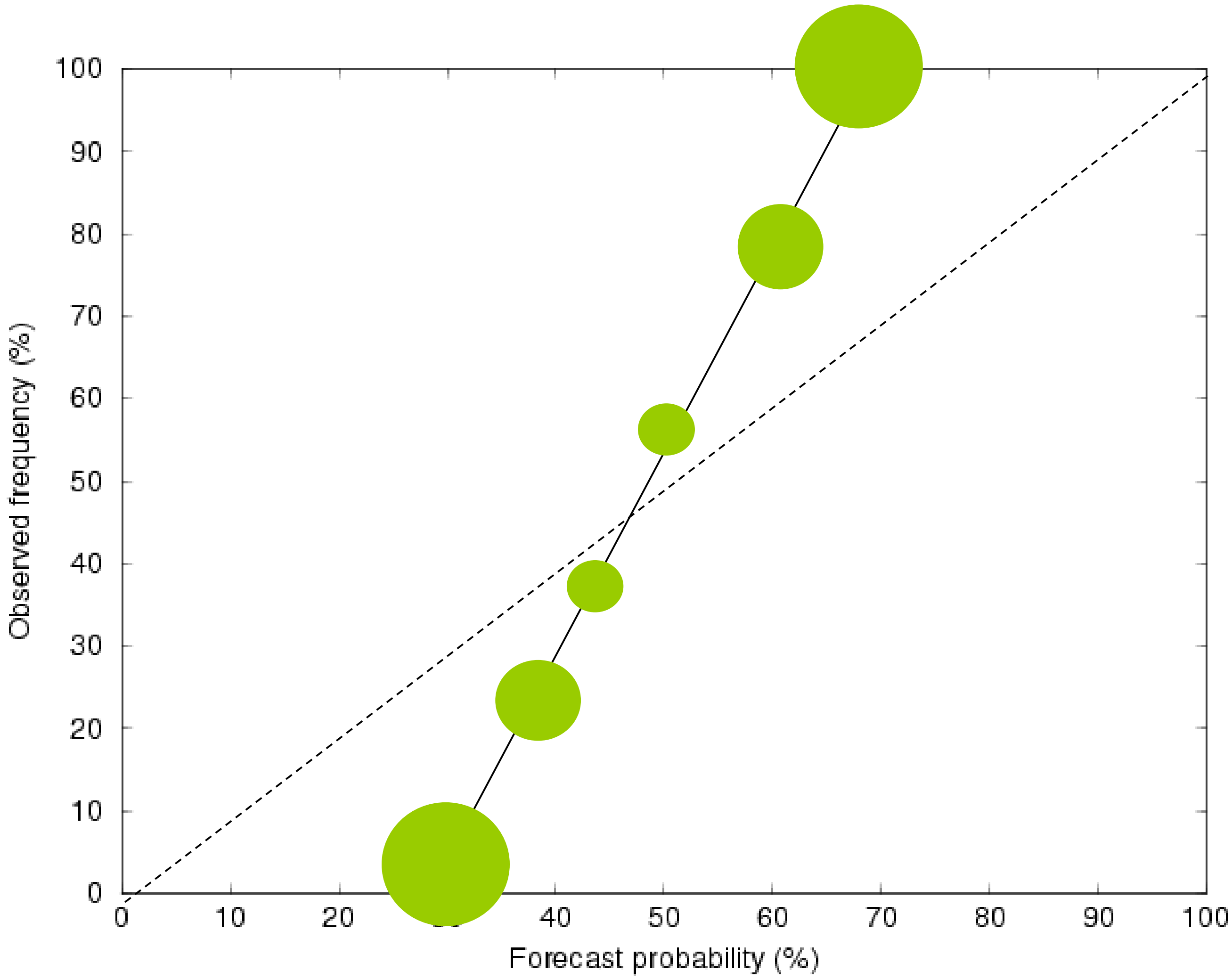
Mathematical calibration – or manual

When you intend to say 0% change it to 30%, when you intend to say 100% say 80%.

Note the reduced range

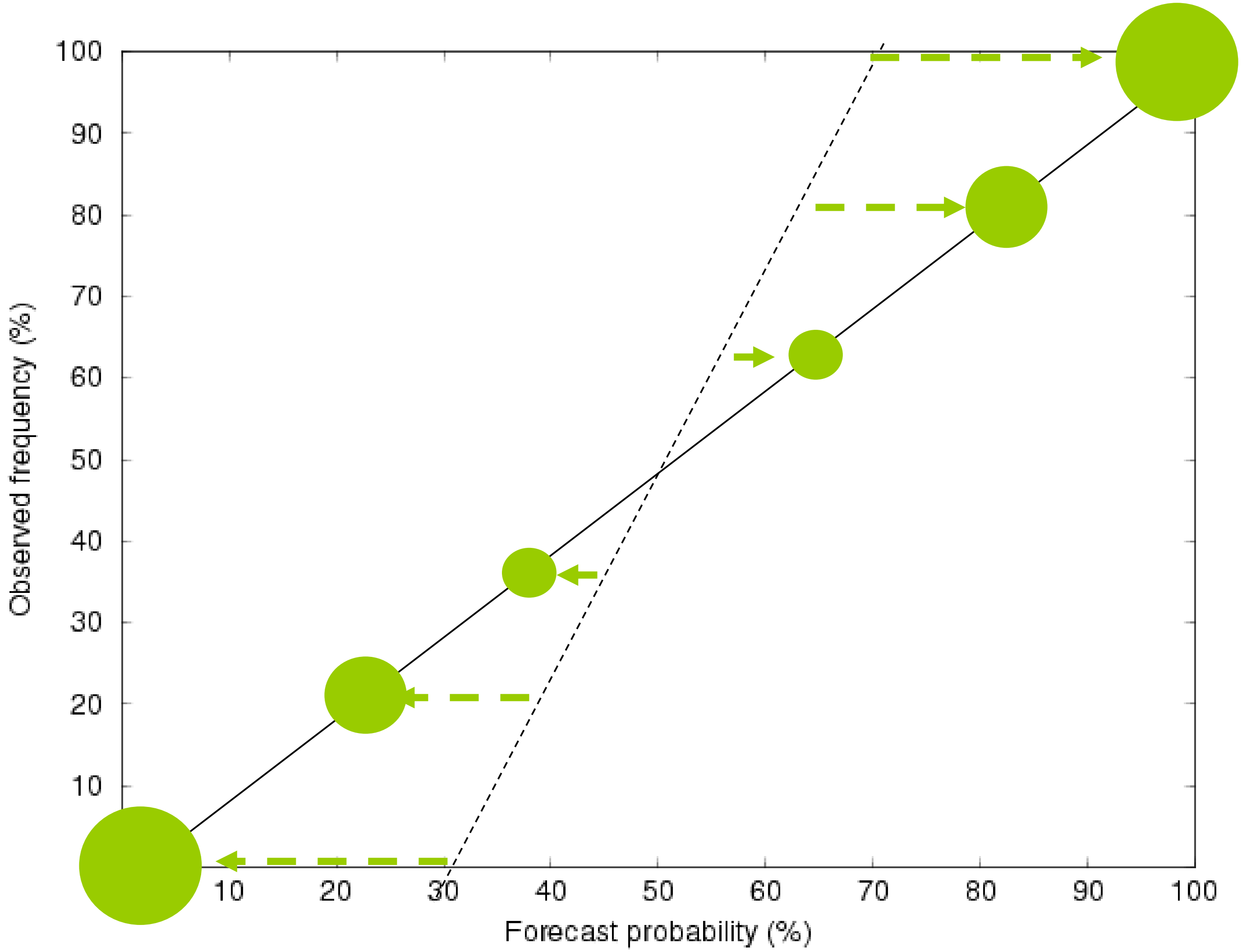


Under-confidence: **a rather rare profile (according to my experience). May occur for long range forecasts as a hesitation to deviate too much from the climatological average**



Mathematical calibration – or manual

When you intend to say 30% change it to 0%, when you intend to say 30% say 20%



END