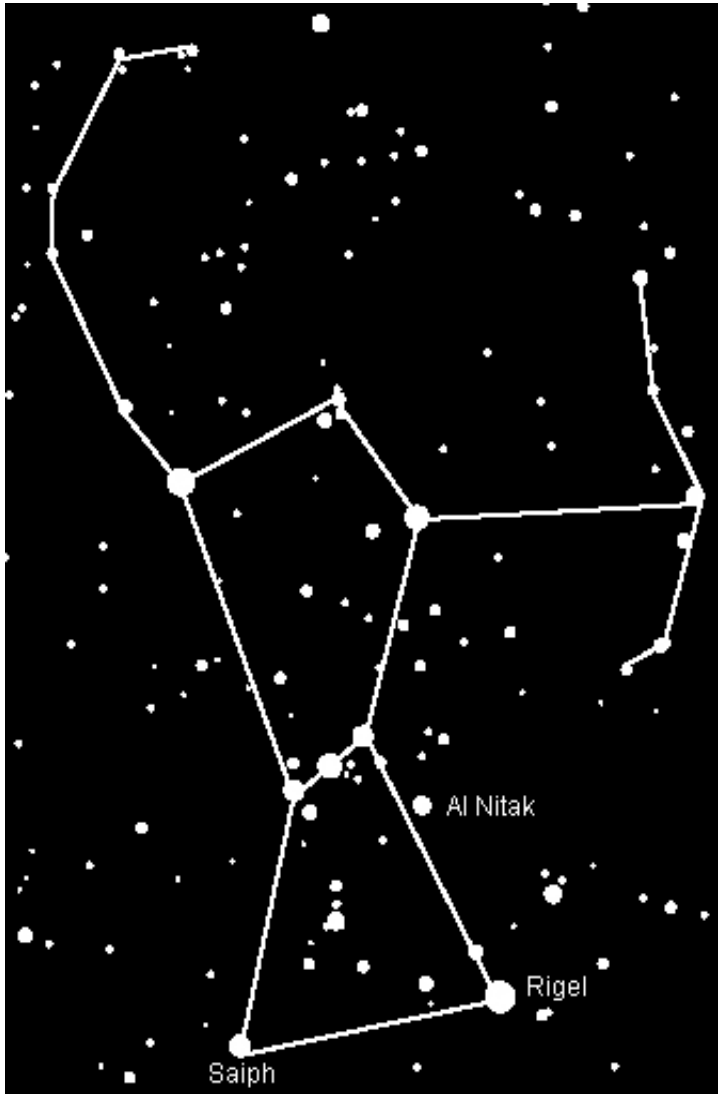
V The psychology of probabilities

3. The Regression to the mean

The problems with randomness



1. Conclusions from too small samples (notorious error!)
2. The Slutsky-Yule Effect, whereby time averaging yields spurious periodicities also from purely random data
3. **The <u>lethal</u> "Regression to the Mean Effect", which yields spurious tendencies and patterns**

V.3.1 Does the ECMWF has a serious model drift problem?

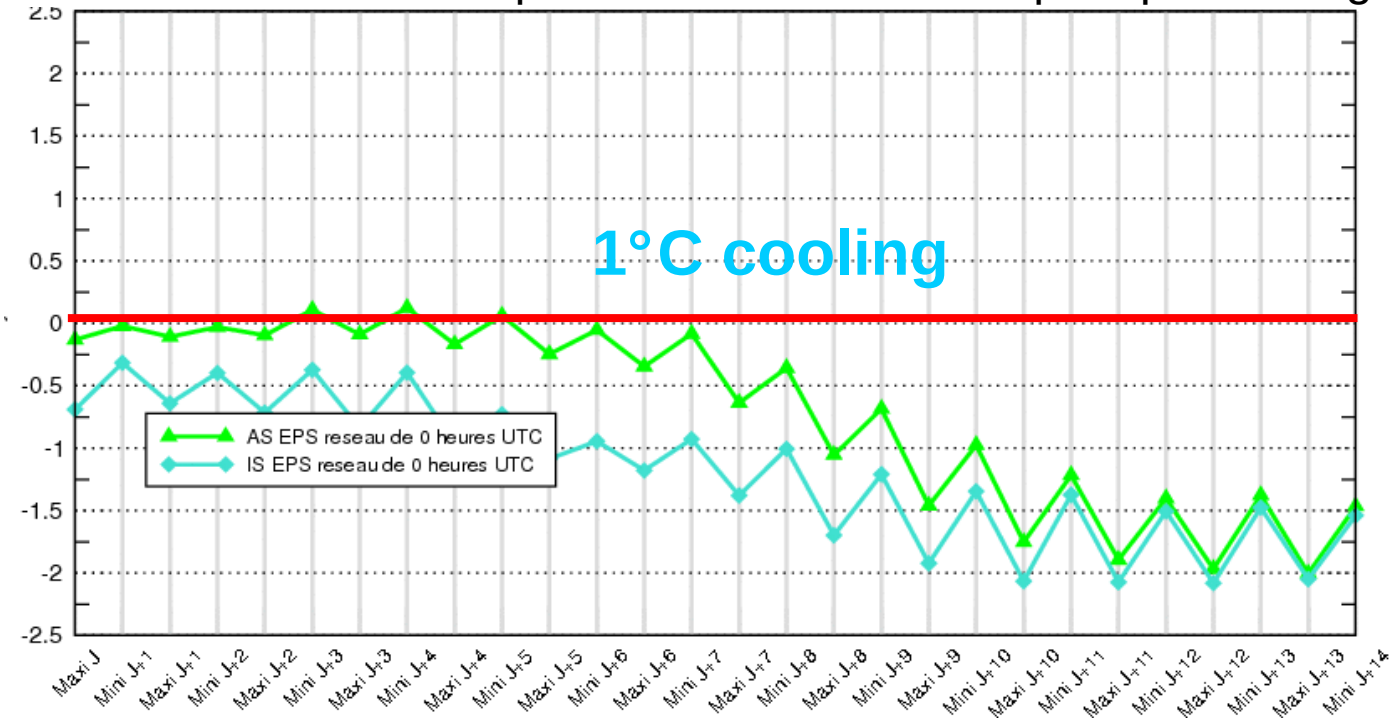
Regression to the mean deceptions in weather forecast verification

During an anomalous period, a non-biased NWP model will, due to *random* errors, display *systematic* mean errors increasing with forecast length.

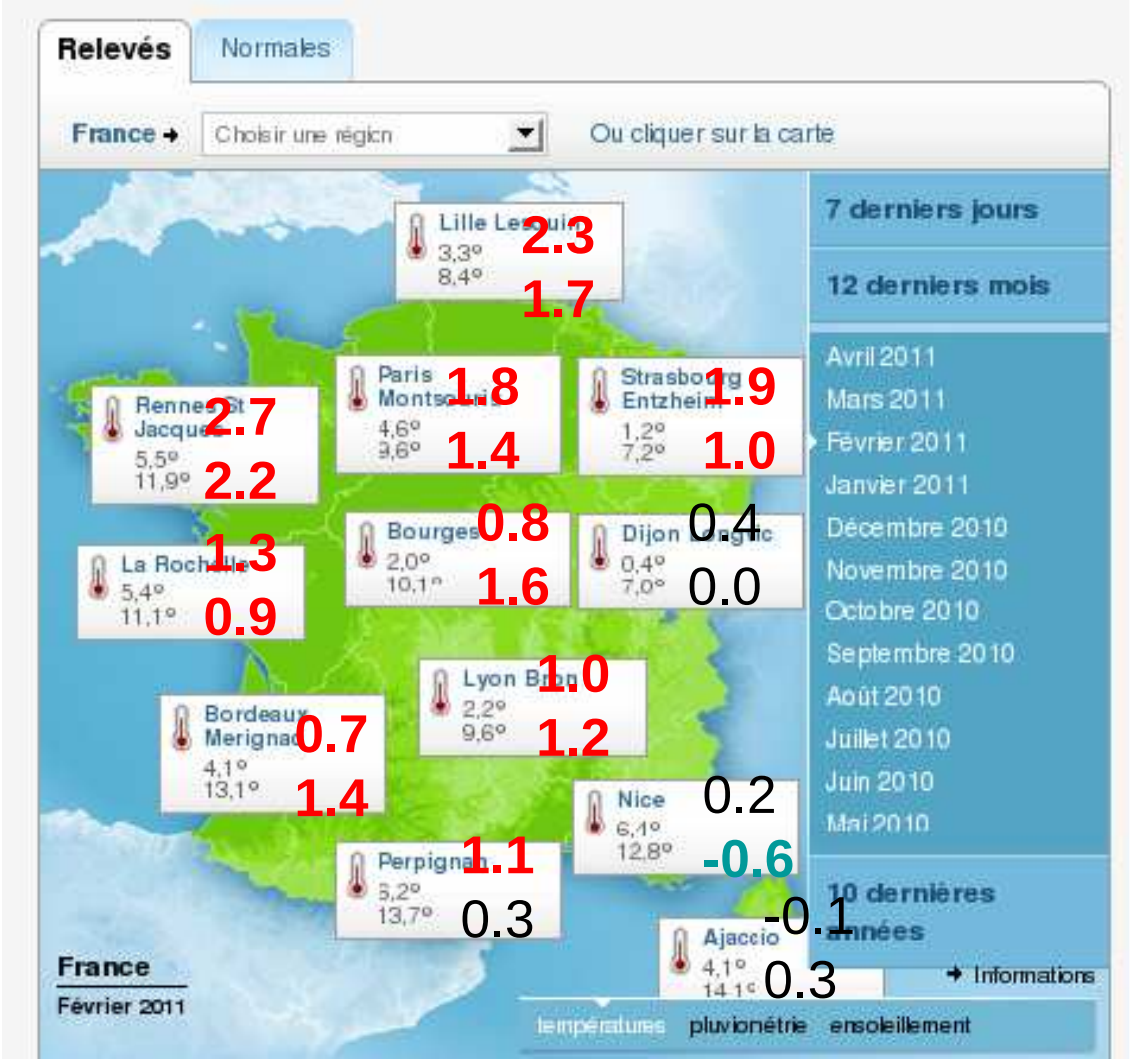
Mean error for the 2m-temperature based on EPS

February 2011 – French stations

=direct model output = statistical post-processing



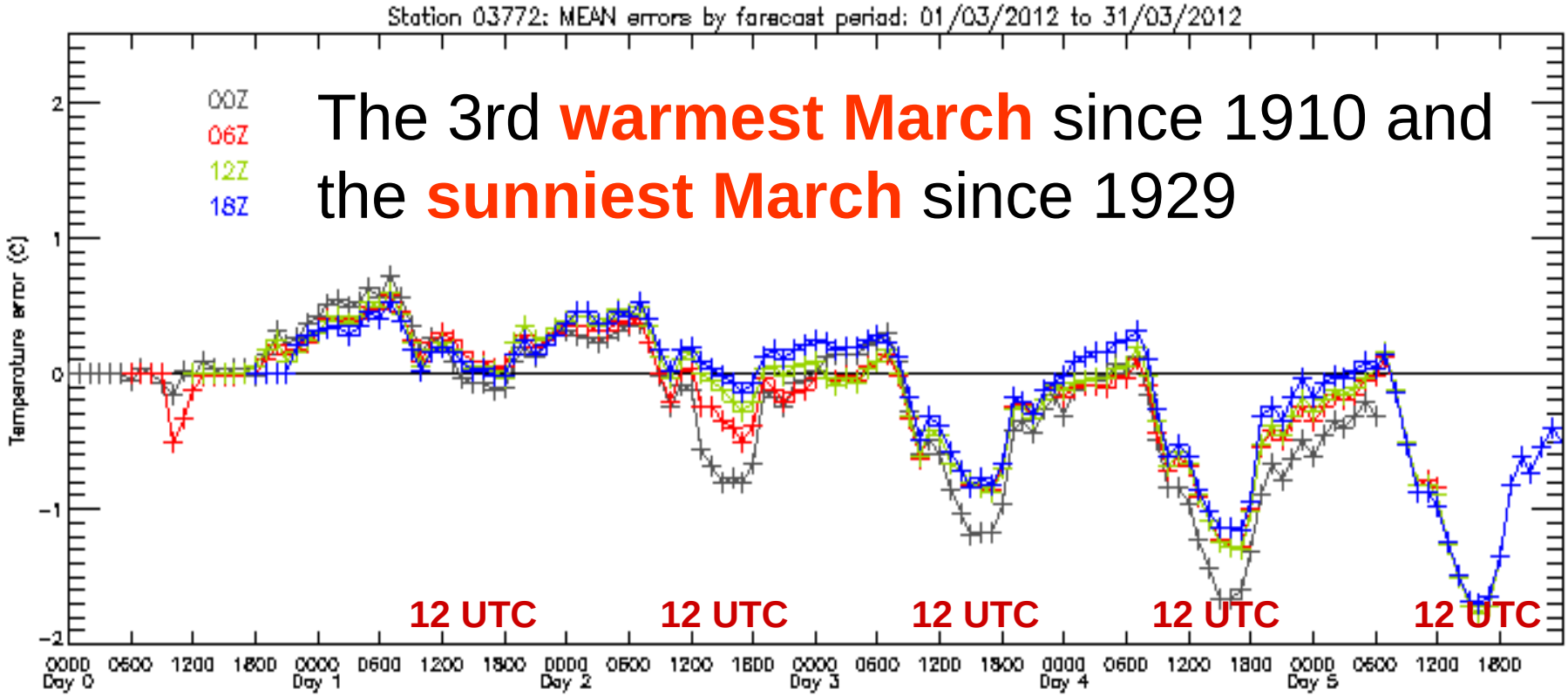
February 2011 was around **1½°** warmer than normal



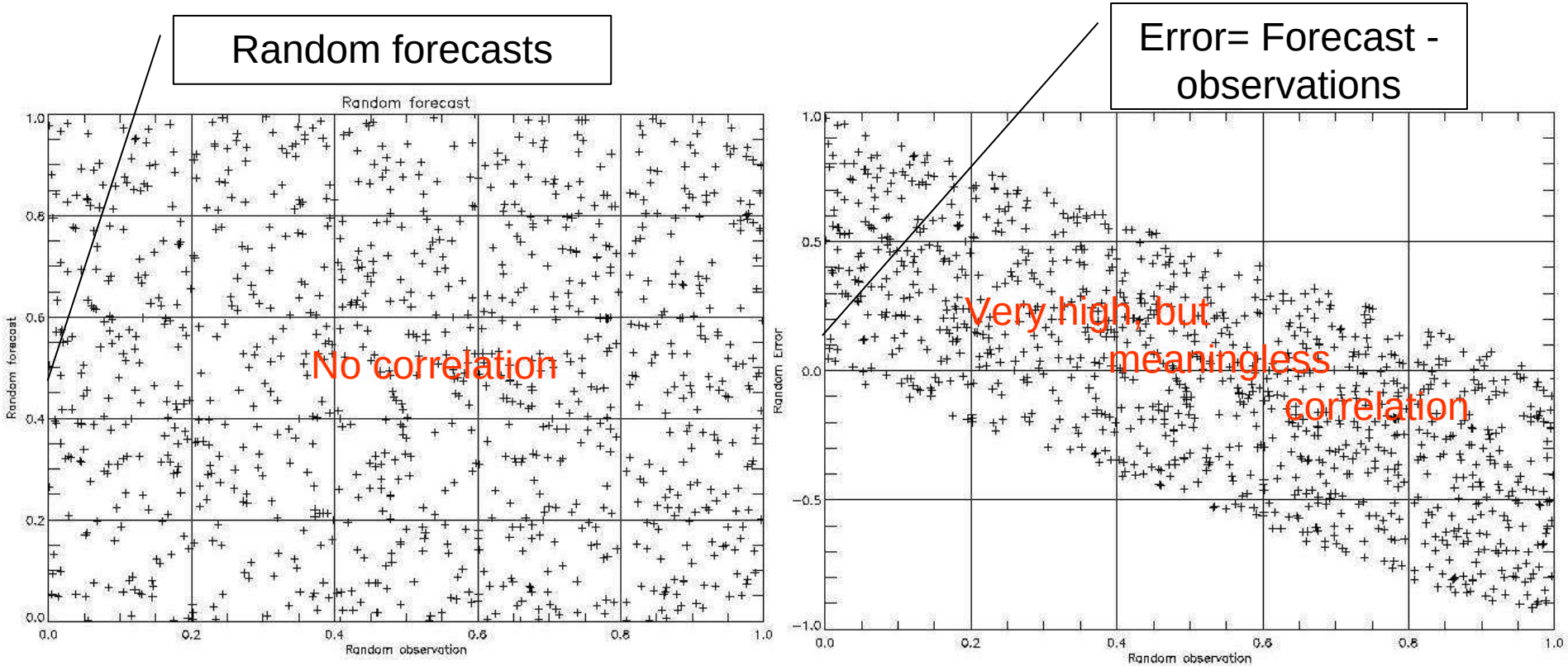
V.3.2. Why didn’t AP’s Kalman filter remove these mean errors???

A similar example from the other side of the Channel in March 2012

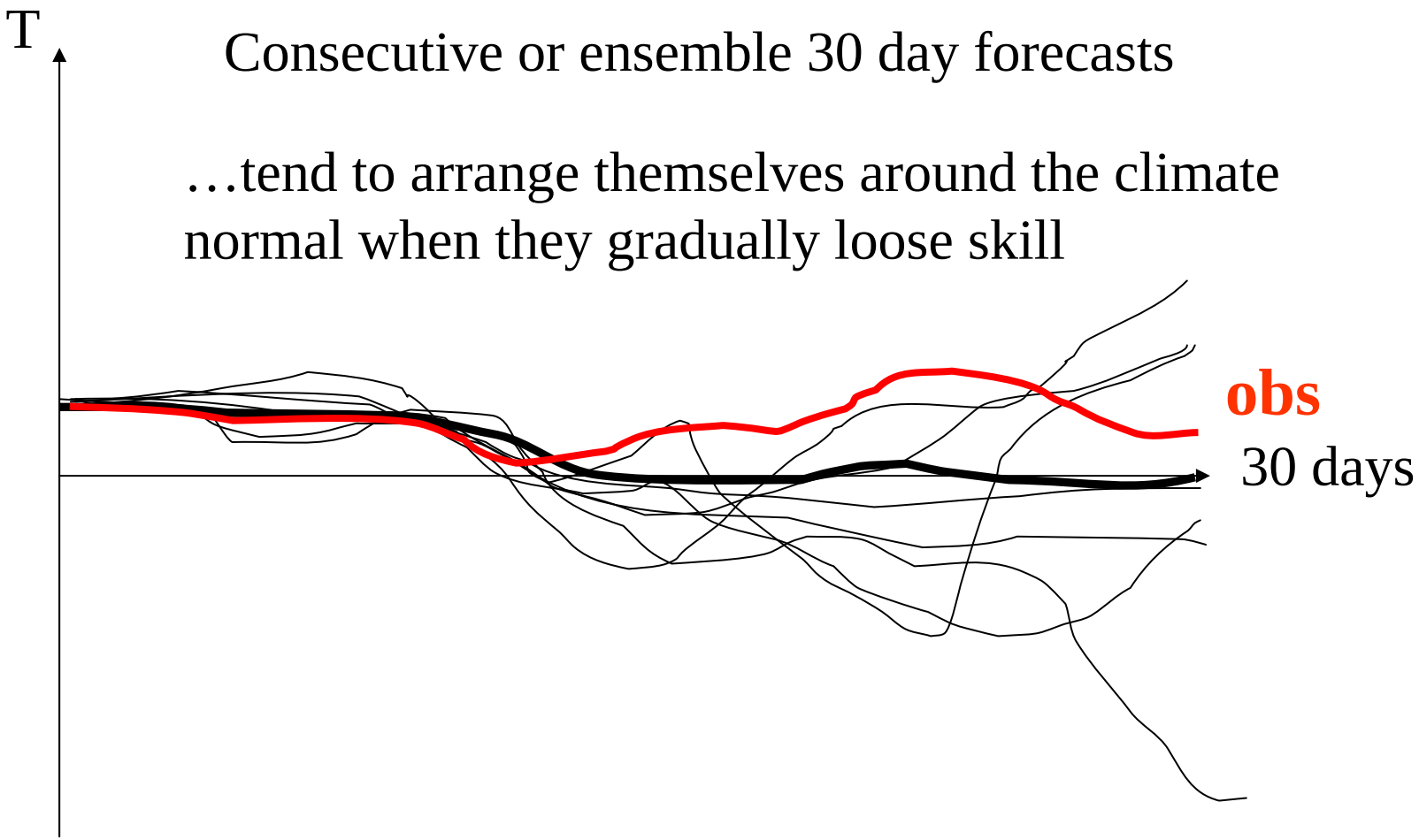
Mean error of UKMO 2 m temperature forecasts 03772 Heathrow



The Met Office, to convince themselves, plotted two graphs, one with random forecasts vs random observations, another with the errors vs the observations and created these spurious high correlations



V.3.3. Regression to the mean in medium-range forecasting

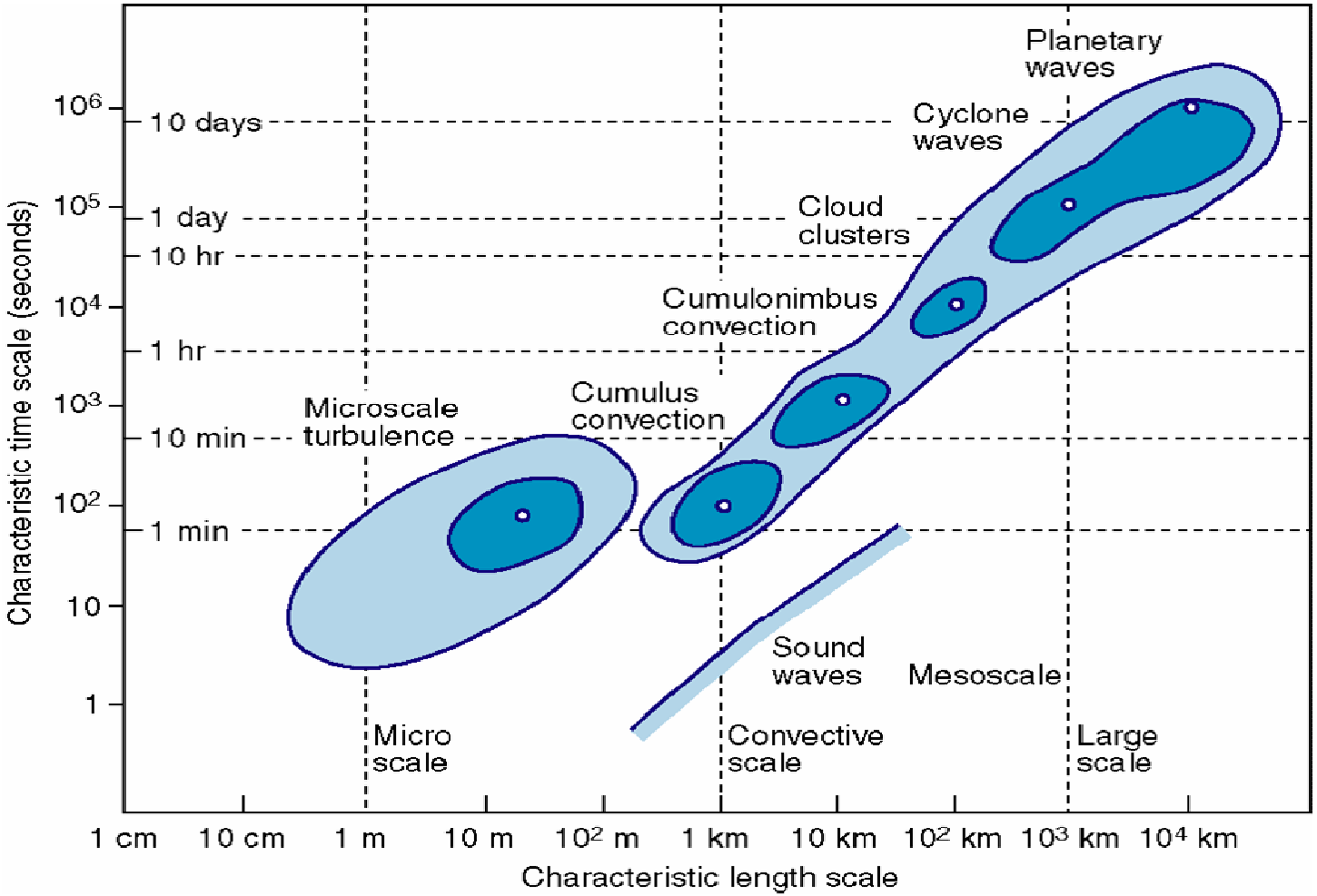


The result gives the impression of a mean error (“bias”)

V.3.4. Regression to the mean in real time forecasting

In the past (not nowadays, I think), when an aviation forecaster was totally taken by surprise by an extreme event (thunderstorm) he issued as his TREND **“GRADU CAVOK”** thus making use of the “regression to the mean effect”

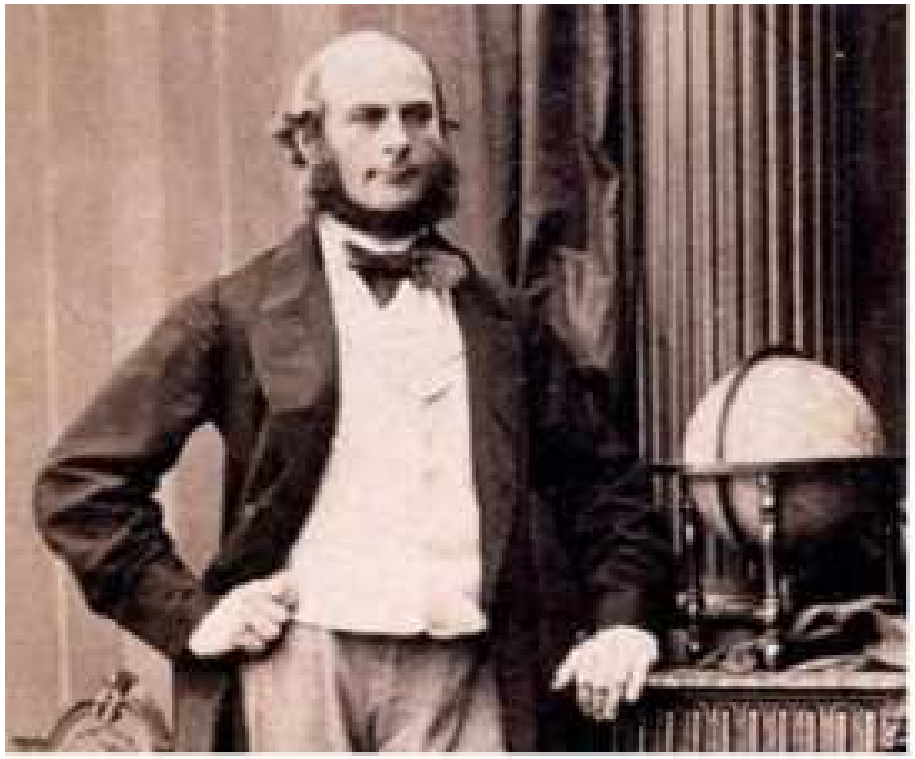
In an extreme weather situation, having no other information, a forecast of a “return to normal” can be issued based on knowledge of the typical life time of the feature



V.3.5 The father of the regression to the mean effect

Regression to the mean effect

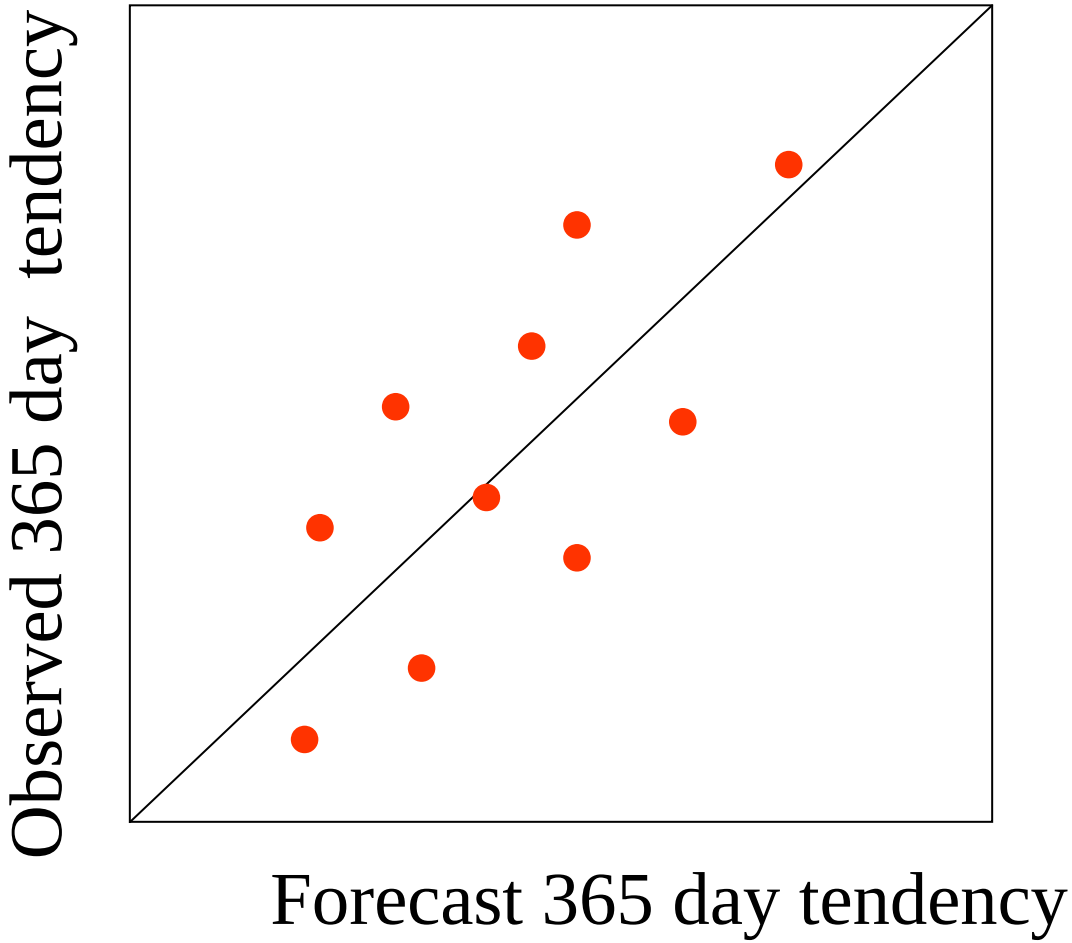
Below average length fathers have taller sons, above average length fathers have shorter sons



Francis Galton 1822-1911

V.3.6 Is it possible to forecast temperature changes a year ahead?

Is it possible to forecast temperature changes over 365 days?



30 years of mean max and min temperatures for Heathrow.

Take the ones for May each year

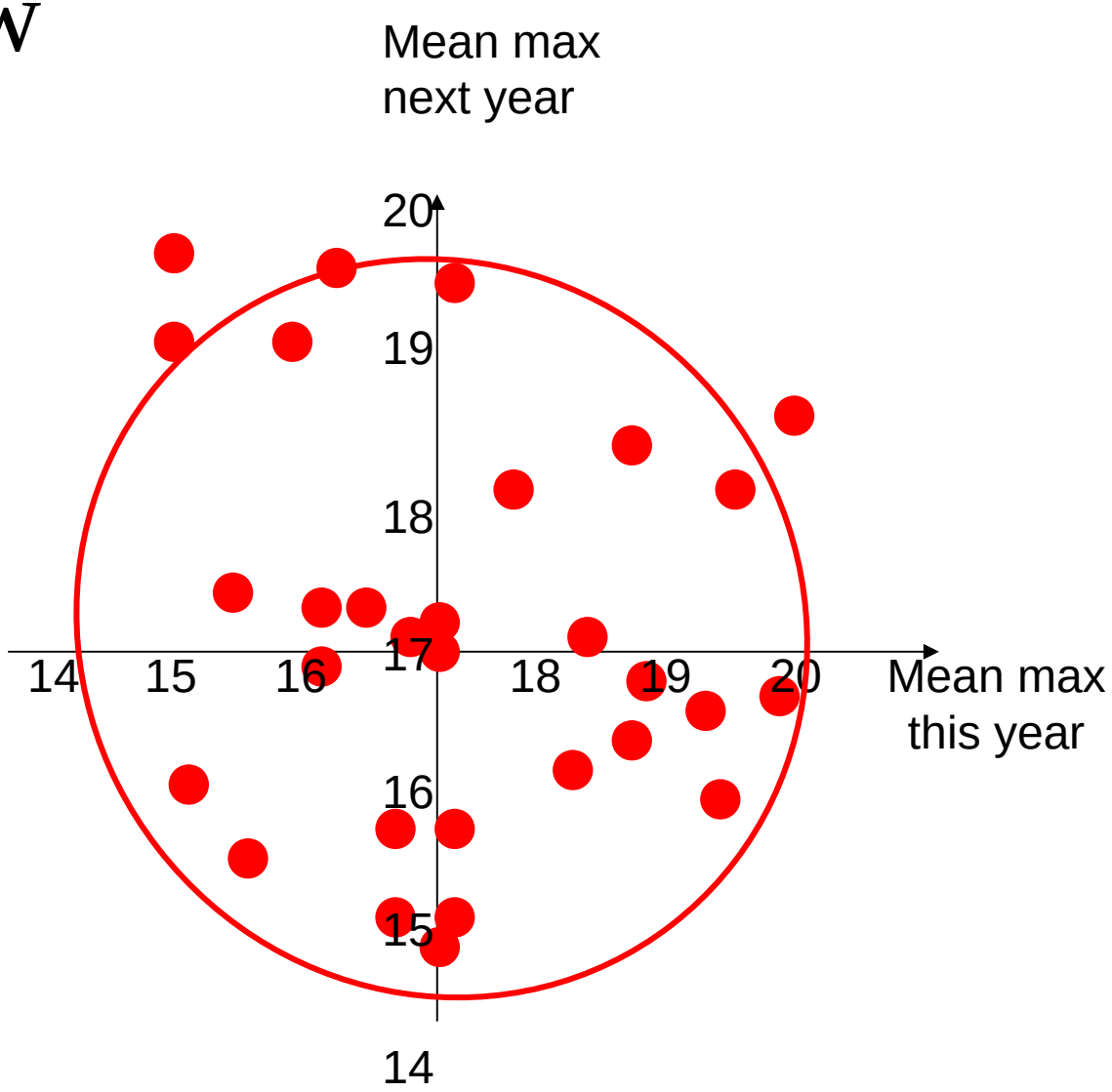
Extract of Heathrow
yyyy mm **tmax** tmin . . .

1957 1 8.7 2.7 5 39.5 53.0
1957 2 9.0 2.9 5 69.8 64.9
1957 3 13.9 5.7 2 25.4 96.7
1957 4 14.2 5.2 1 5.7 169.6
1957 5 **16.2** 6.5 0 21.3 195.0
1957 6 23.6 10.7 0 22.4 284.5
1957 7 22.5 13.8 0 87.0 152.3
1957 8 21.1 12.5 0 86.2 154.4
1957 9 17.6 10.1 0 51.7 88.5
1957 10 15.5 7.7 0 47.0 85.9
1957 11 9.4 4.3 2 59.5 67.5
1957 12 7.6 1.0 11 42.1 40.8

1958 1 6.8 0.9 10 64.3 40.1
1958 2 8.9 1.9 10 58.7 45.7
1958 3 8.1 1.1 10 26.0 105.2
1958 4 12.3 3.8 3 29.5 153.2
1958 5 **17.3** 7.8 0 59.5 189.2
1958 6 19.4 10.7 0 104.3 152.2
1958 7 21.7 12.9 0 51.9 190.5
1958 8 20.8 13.1 0 75.2 103.1
1958 9 20.0 12.1 0 83.8 134.8
1958 10 14.9 8.3 0 50.7 94.2
1958 11 9.7 4.4 1 50.7 40.8
1958 12 8.0 2.7 2 85.1 29.6

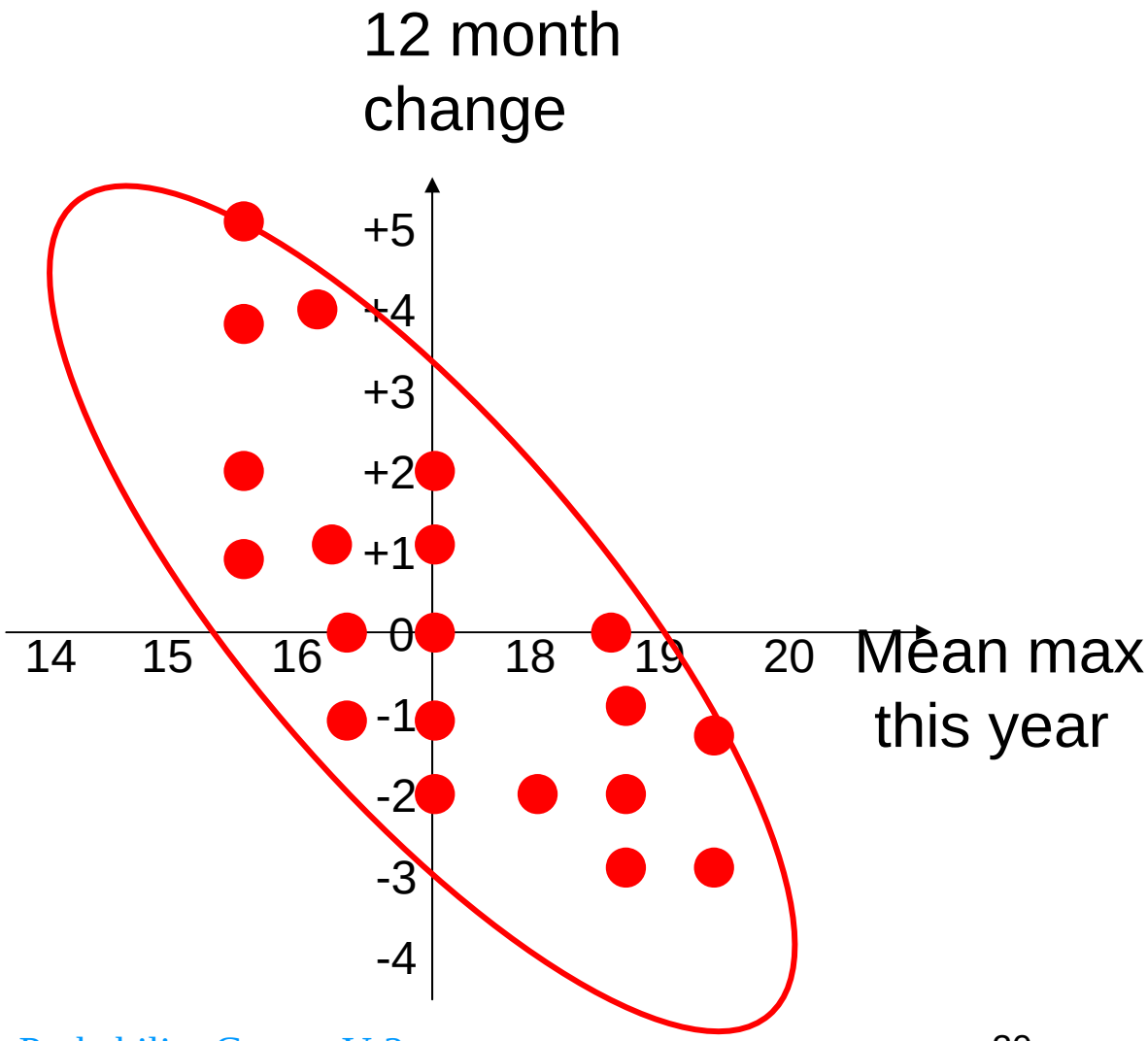
Mean monthly maximum in temperature in May at Heathrow 1948-1977

As expected, there is no clear correlation between the mean maximum temperature one year (the x-axis) and the next year (y-axis)

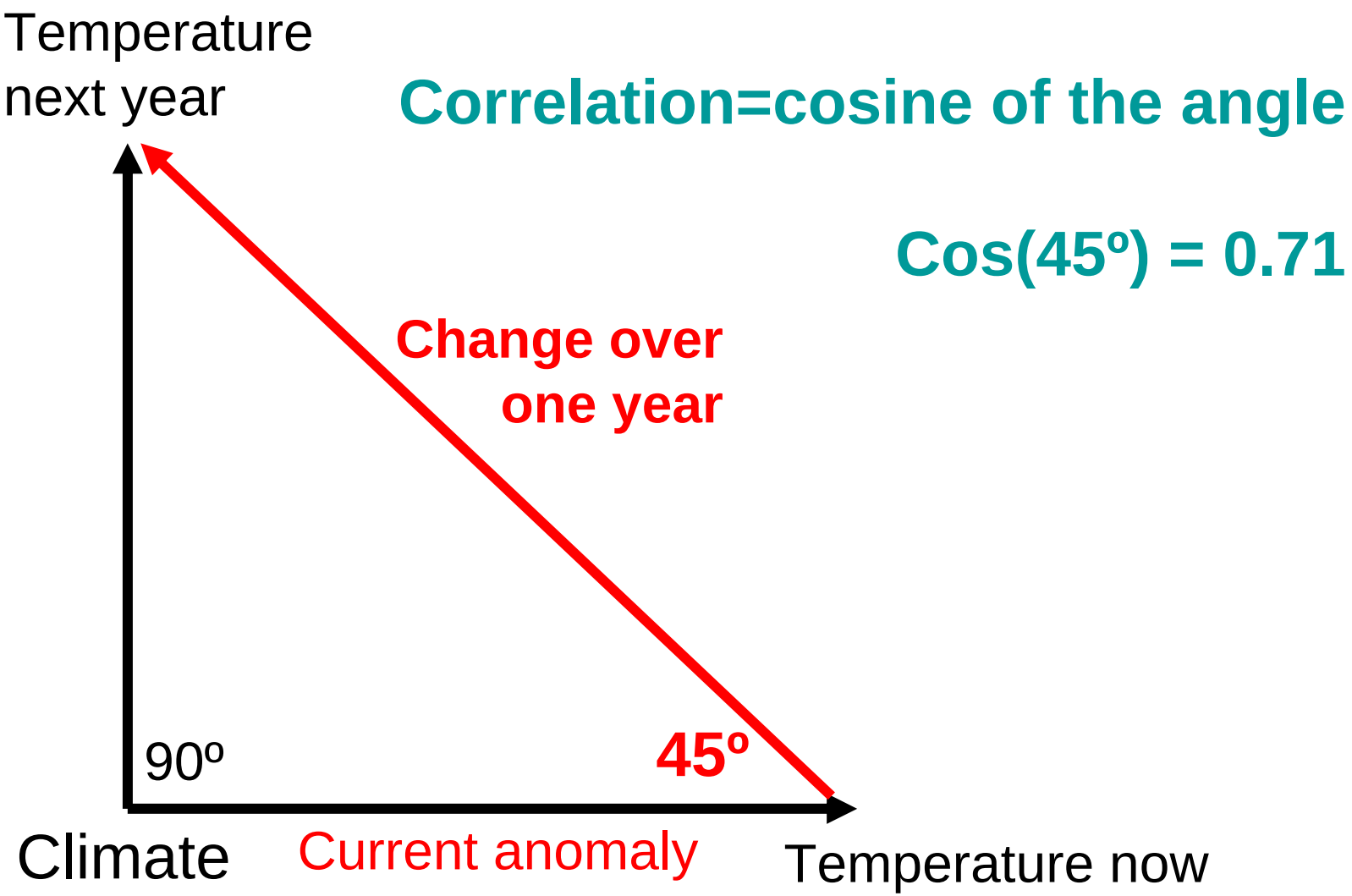


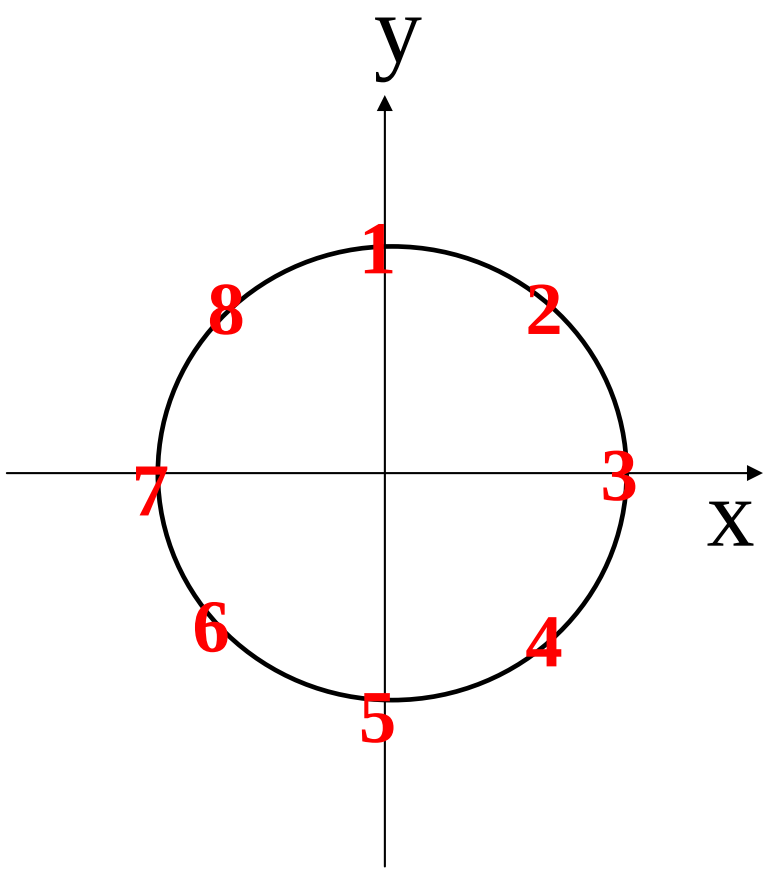
Mean monthly maximum temperature in May at Heathrow 1948-1977

There is now a clear correlation between the change over 12 months versus the mean max value the previous year. A forecast of the change over 365 days would verify with a 71% correlation



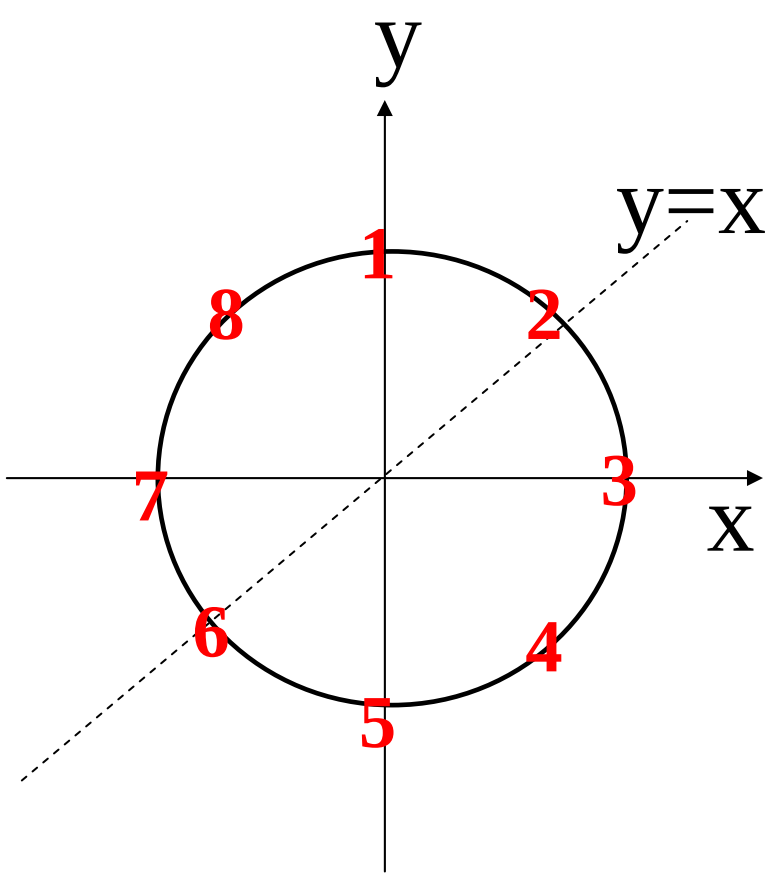
There are different ways to prove this mathematically:

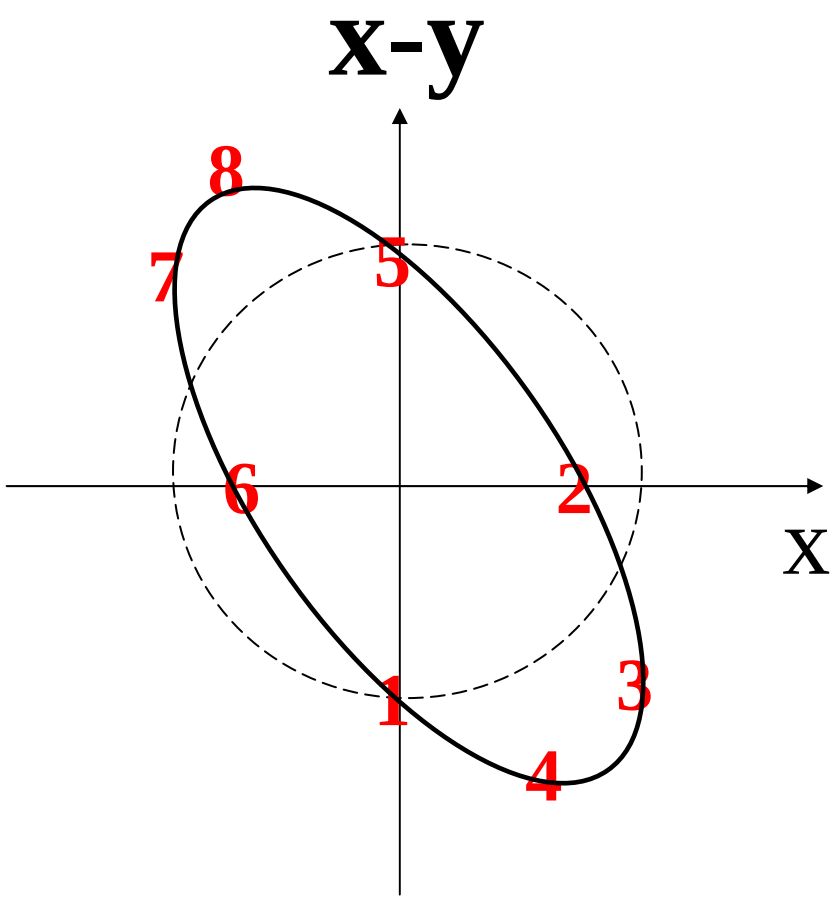




$x^2 + y^2 = 1$

Consider eight values along a circle. They are uncorrelated in a x-y representation, **but not in an x, y-x representation**





$x^2 + y^2 = 1$

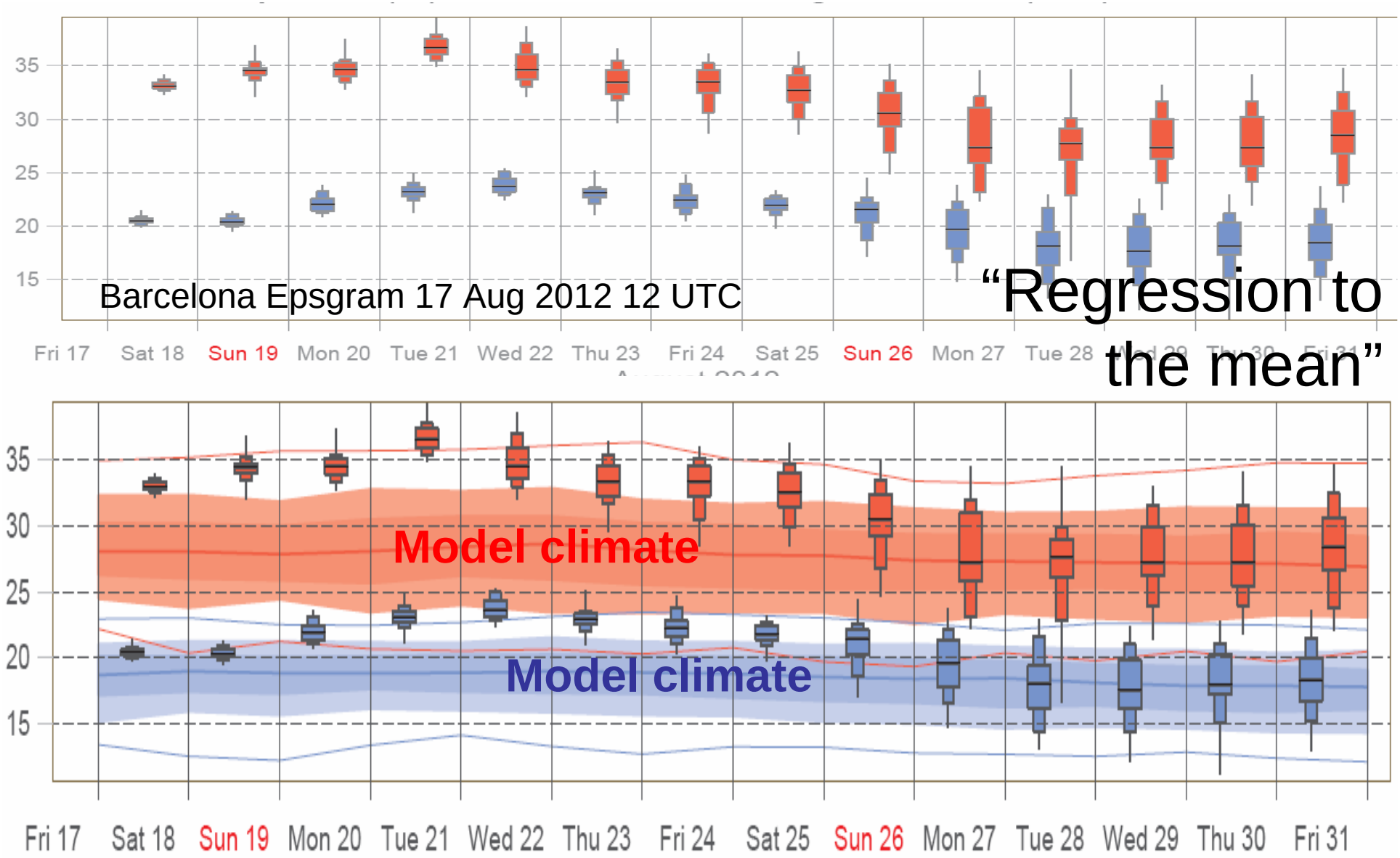
$(x-y)^2 = 1 \pm 2x\sqrt{1-x^2}$

V.3.7 Regression to the mean effect in medium range ensemble forecasts

During an anomalous period the ensemble probabilities will, due to *non-systematic* forecast errors, *systematically* drift towards the climatological average.

During a heat wave the weather seems to be cooling in the ensemble forecasts, during a prolonged cold spell the forecasts seem to indicate return to milder conditions.

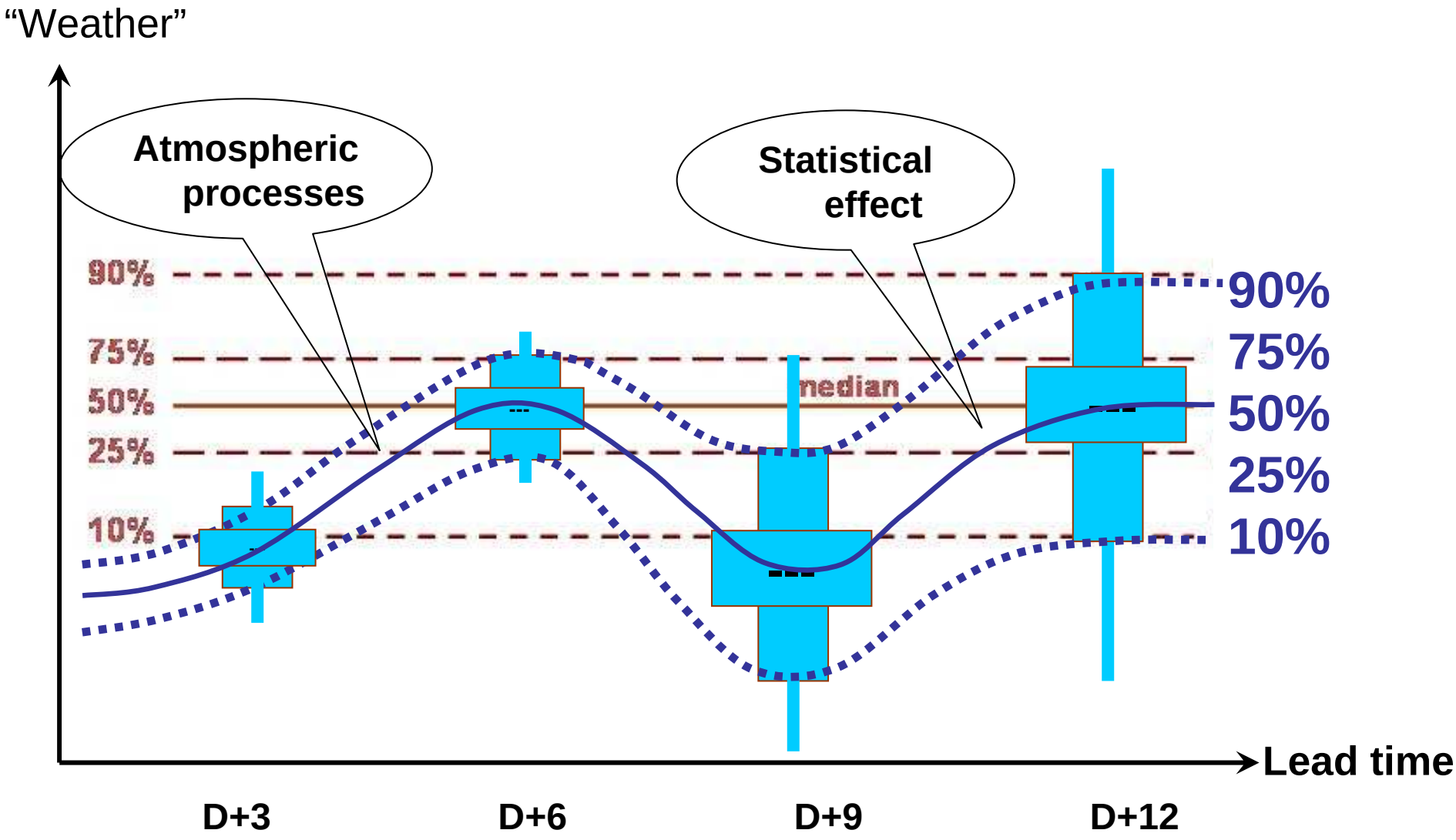
Example of regression to the mean



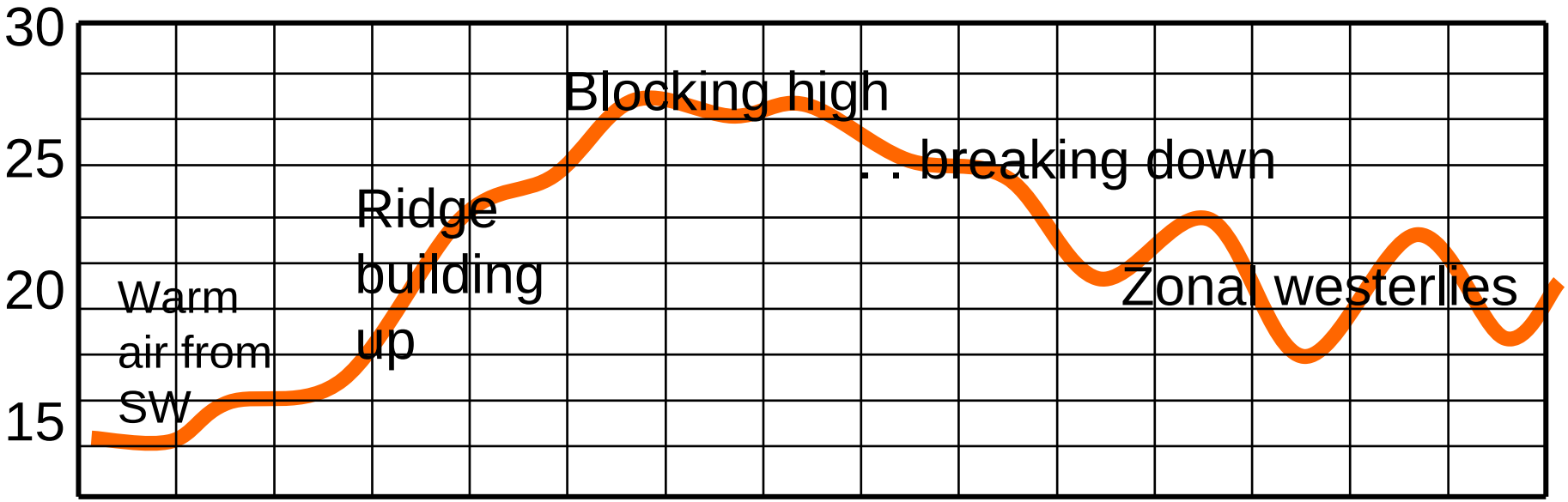
"Regression to the mean"

Has randomness a message?

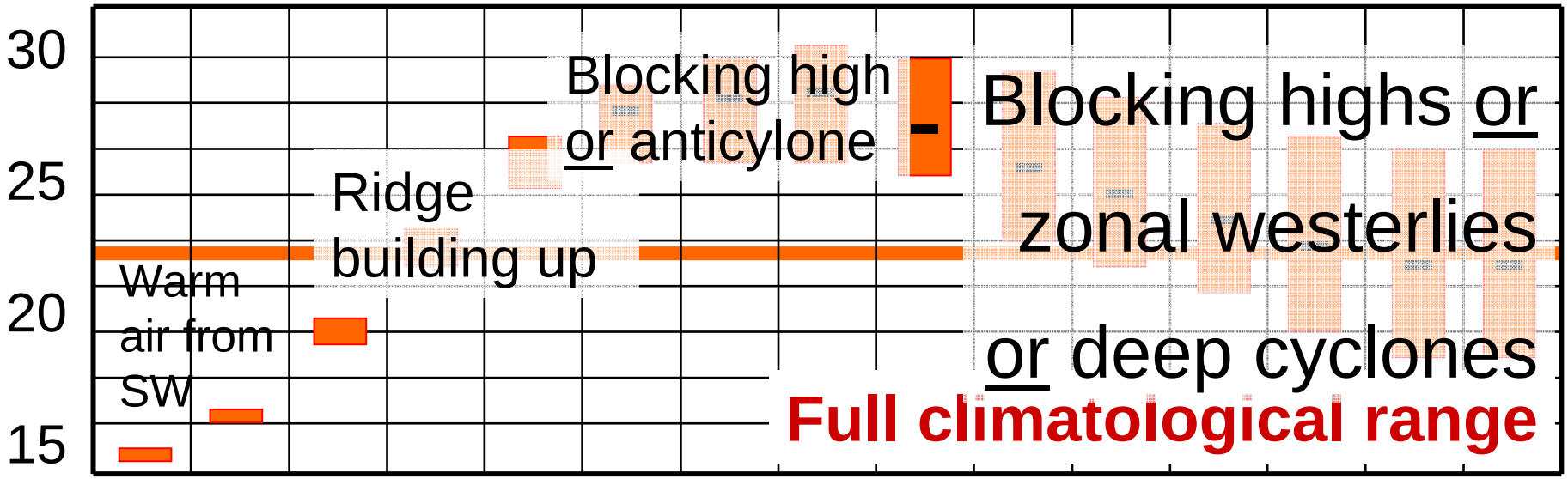
An example of the “regression to the mean effect”



A deterministic meteogram can be given a dynamically (physically) consistent interpretation in common synoptic terms



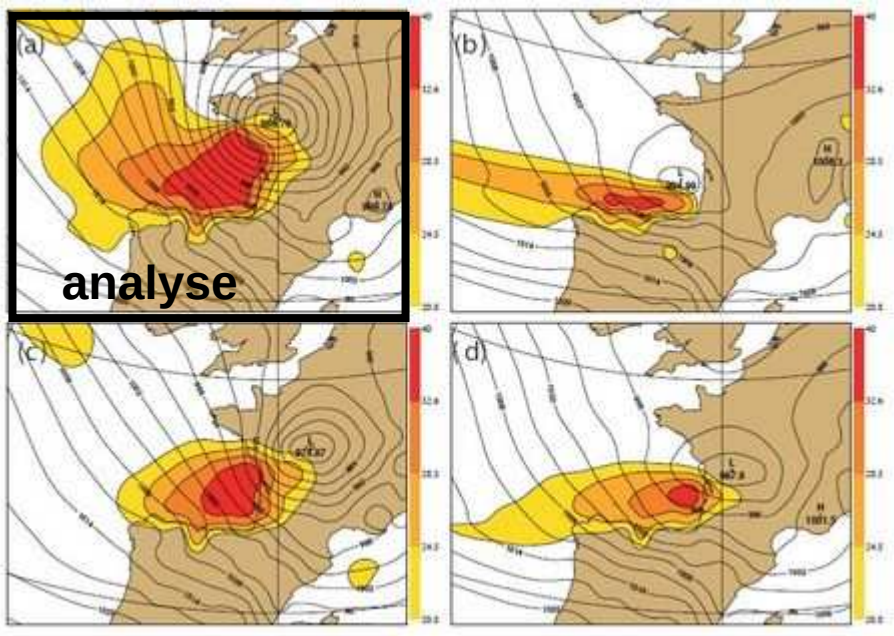
In a probabilistic epsgram this is only possible when the spread is small, normally in the early part of the forecasts.



V.3.8. Is there a “Regression **from** the mean” effect?

This might be inverted into “regression *from* the mean”:

“Following a missed extreme weather event, we have to over-forecast the next”



The operational forecasters success with the 2nd Xmas storm 1999, with conflicting NWP guidance, rested solely on the “Regression *from* the mean” method

But the cost-function they use apply to them, not to any customer

End