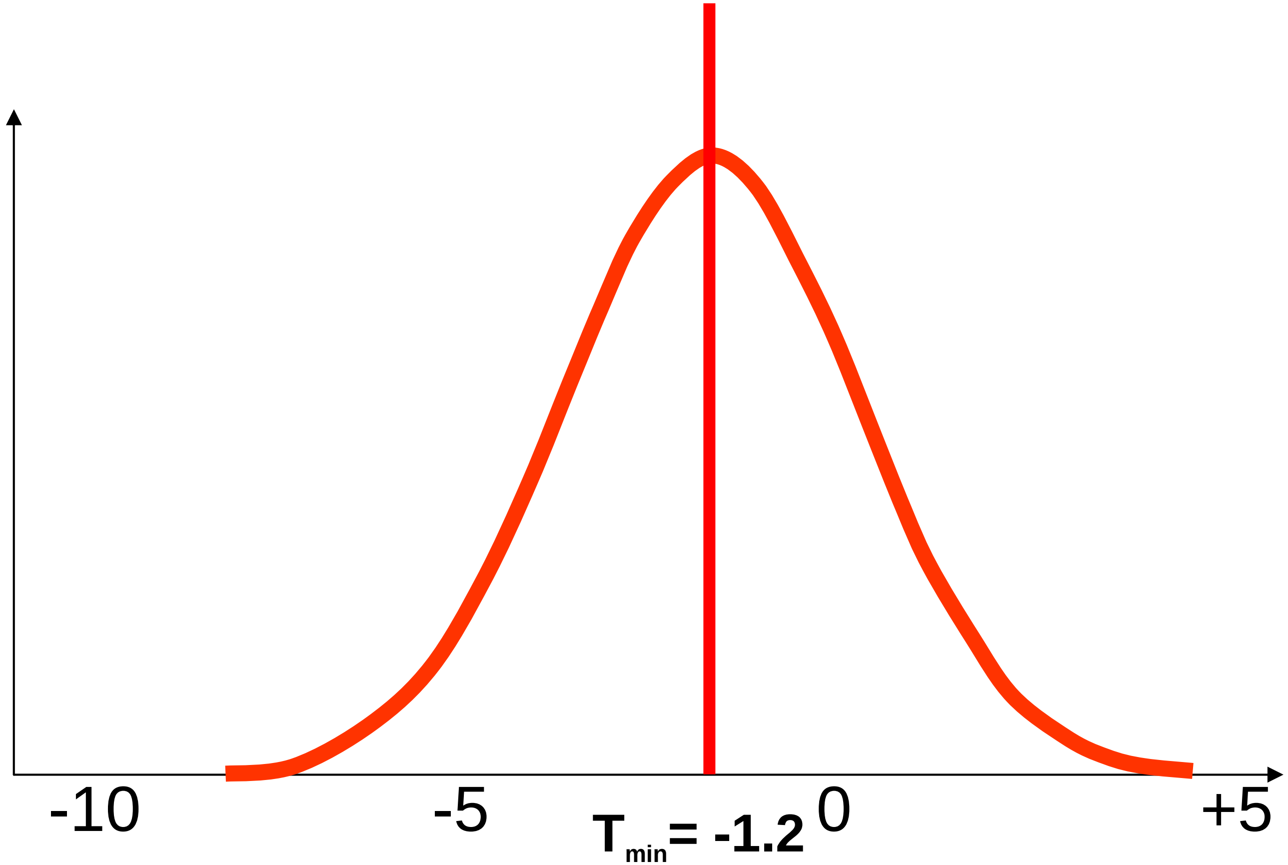
II. Frequentist probabilities

II.1 The problem with the "mean"

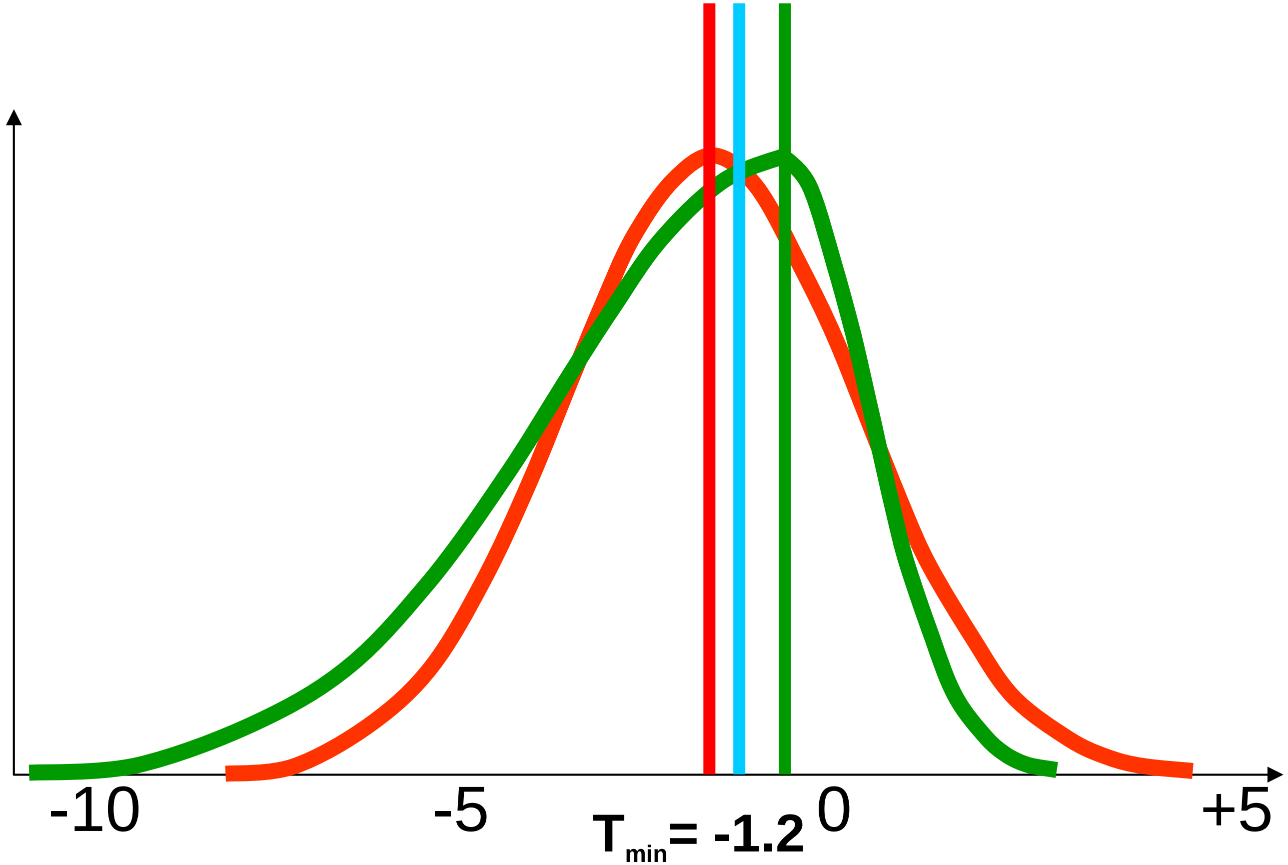
II.1.1 Probabilities is not the most controversial issue

In statistics probability is called "the 2nd moment" where "the 1st moment" is the mean or median

The 3rd moment is the skewness (asymmetry) of the distribution.



Mean, median or mode?



The probability information does not normally "hang in the air" – it is supplementing some sort of single value deterministic forecast:

-We expect winds around 9 m/s with a 20% possibility of gale force.

The "Best Data" Paradox

Probabilities are difficult to interpret and use, **but** they are fairly simple to produce

Categorical values, on the other hand, are easy to interpret **but**, paradoxically, difficult to produce

Should they be the ensemble **mean or median**,

Accurate, not "jumpy" and consistent with probabilities, but not always "physically realistic"

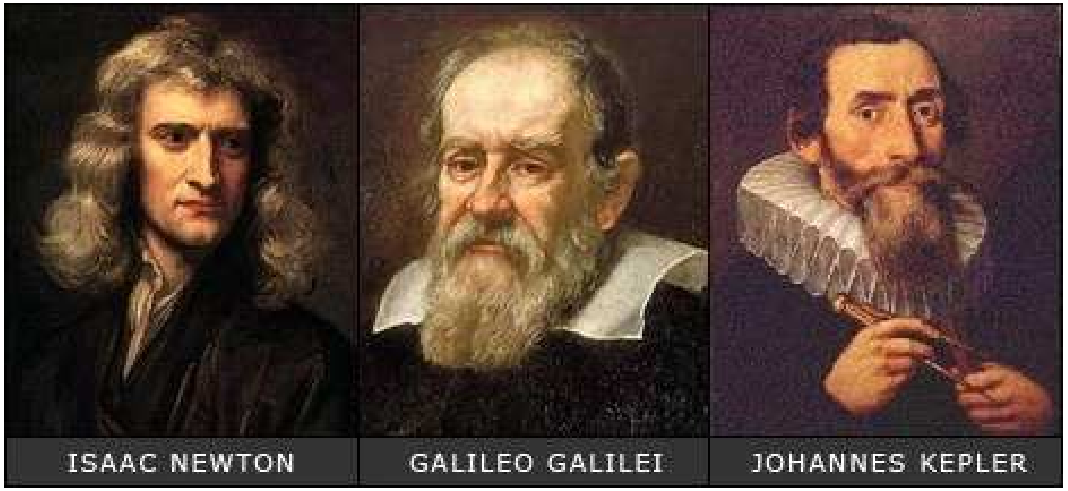
or just **DMO** from a favoured NWP model?

Physically realistic but less accurate, very "jumpy" and not consistent with the probabilities

The problem goes 250 years back in time . . .

II.1.2 Choosing the “best” observation in the 1700’s

Before the 1800's there was a poor understanding of randomness in measurement errors



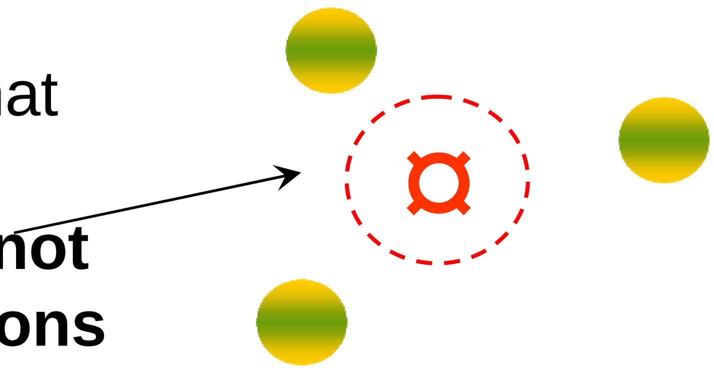
1. Scientists had the routine to select their "best" measurement
2. They didn't understand that measurement errors add up and randomly cancel out
3. **They disliked averages of observations since these did not normally agree with measured values**

18th century view on observation errors

1. Astronomers in the 1600:s and 1700:s tried to find out which of their diverging observations was the “right” one

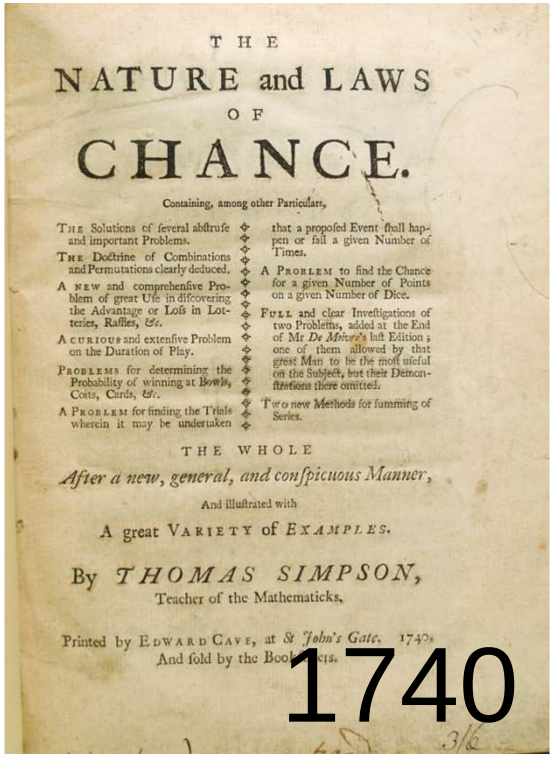
2. In the late 1700’ it was realized that that the observations should be combined **even if the result did not agree with any of the observations**

Where is Jupiter?



3. The first mathematical discussion on statistical inference

Thomas
Simpson
1710-61
Mathematician



THE
NATURE and LAWS
OF
CHANCE.

Containing, among other Particulars,

THE Solutions of feveral abftrufe and important Problems.

THE Doctrine of Combinations and Permutations clearly deduced.

A NEW and comprehenfive Problem of great Ufe in difcovering the Advantage or Lofs in Lotteries, Raffles, &c.

A CURIOUS and extenfive Problem on the Duration of Play.

PROBLEMS for determining the Probability of winning at Bowls, Coits, Cards, &c.

A PROBLEM for finding the Trials wherein it may be undertaken that a propofed Event fhall happen or fail a given Number of Times.

A PROBLEM to find the Chance for a given Number of Points on a given Number of Dice.

FULL and clear Inveftigations of two Problems, added at the End of Mr *De Moivre's* laft Edition; one of them allowed by that great Man to be the moft ufeful on the Subject, but their Demonftrations there omitted.

TWO new Methods for fumming of Series.

THE WHOLE

After a new, general, and confpicuous Manner,

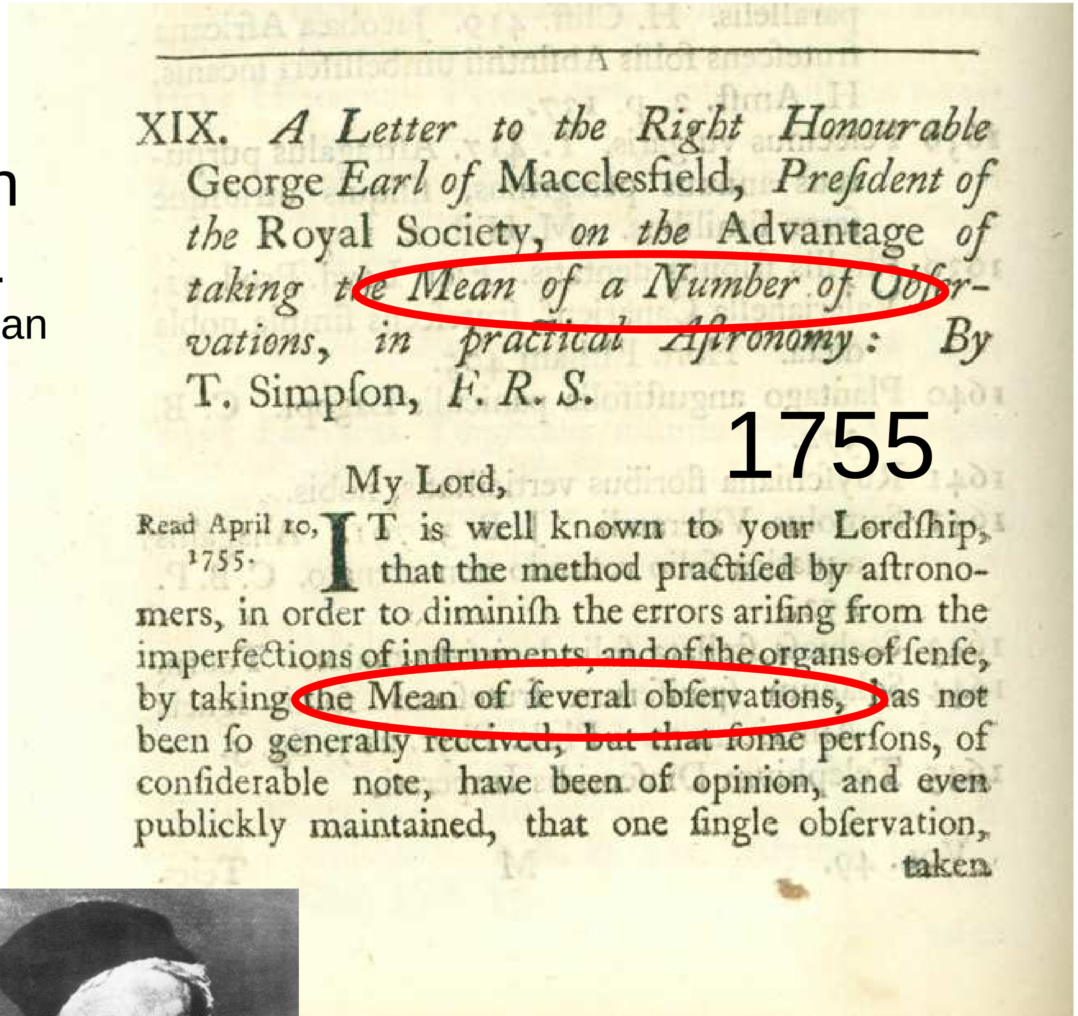
And illuftrated with

A great VARIETY of *EXAMPLES.*

By *THOMAS SIMPSON,*
Teacher of the Mathematicks.

Printed by EDWARD CAVE, at *St John's Gate.* 1740.
And fold by the Booksellers.

1740



XIX. *A Letter to the Right Honourable* George *Earl of* Macclesfield, *Prefident of the* Royal Society, *on the* Advantage *of taking the Mean of a Number of Obfervations, in practical Aftronomy: By* T. Simpfon, *F. R. S.*

1755

My Lord,

Read April 10, 1755.

IT is well known to your Lordfhip, that the method practifed by aftronomers, in order to diminifh the errors arifing from the imperfections of inftruments, and of the organs of fenfe, by taking the Mean of feveral obfervations, has not been fo generally received, but that fome perfons, of confiderable note, have been of opinion, and even publickly maintained, that one fingle obfervation, taken

Only accepted 50-60 years later thanks to the works by Lagrange and Gauss

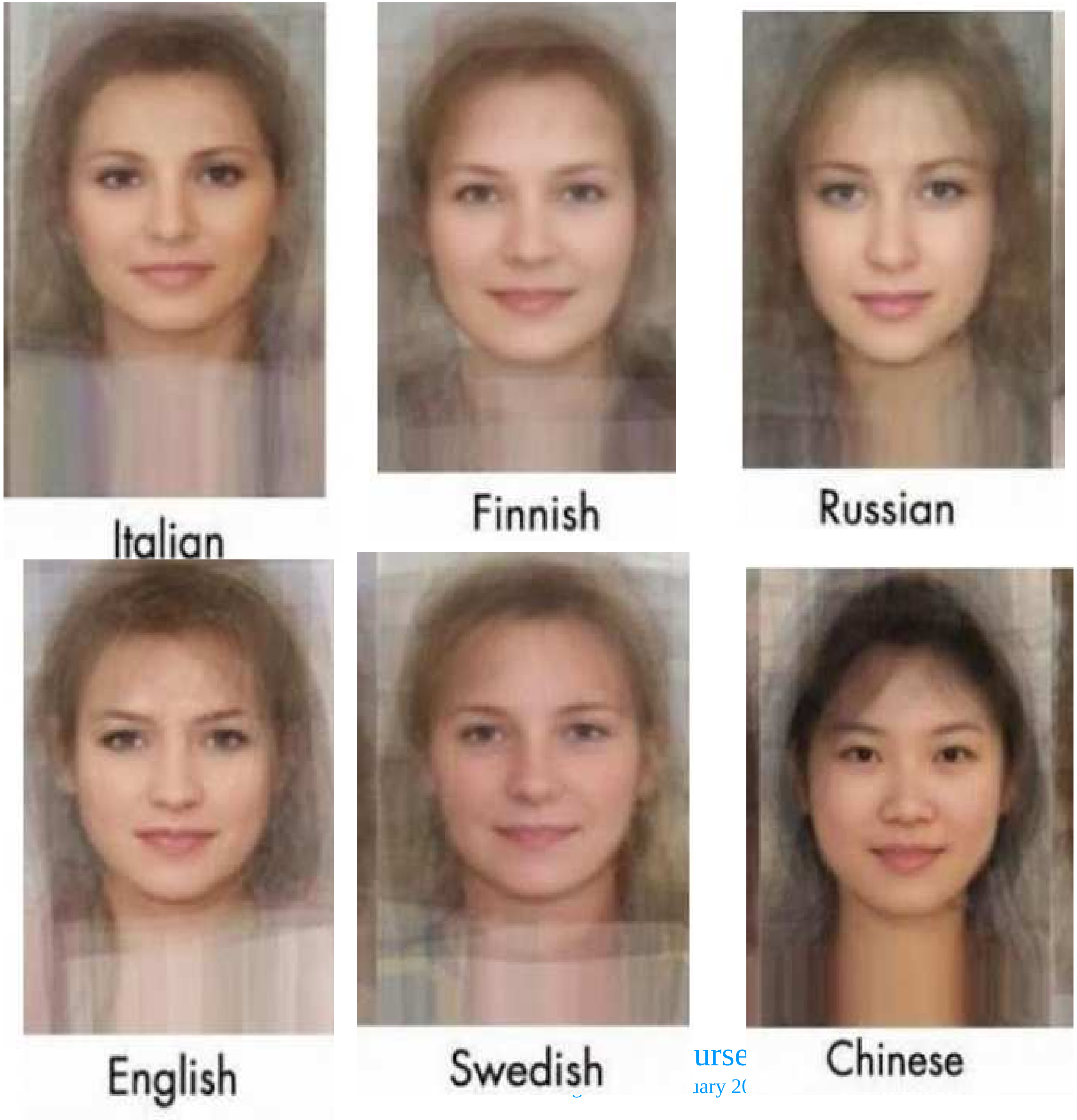
The Belgian meteorologist and statistician Adolphe Quételet (1796-1874) introduced in the mid 1800's the concept of "The Average Man" based on statistical averages from the population in Brussels.

He was criticised because there was nobody in Brussels who fitted this description

The Average Man?

Not very skilful average. But . . .

The Average Girl?



The “Average” Team Member

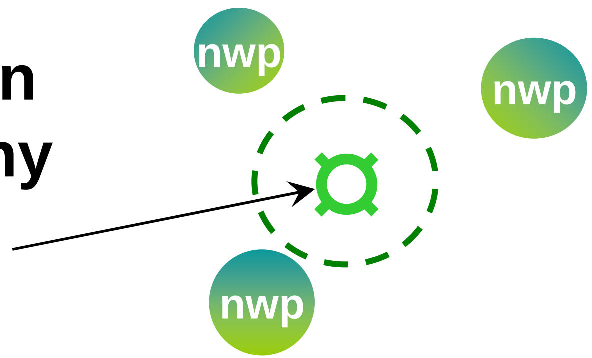
II.1.3 The Average Forecast

20th century discussion of forecast errors

1. Meteorologists in the 1900:s and early 2000:s still try to find out which of the diverging NWP is the “right” one

What is the weather?

2. It is not always realized that the observations should be combined **even if the result does not agree with any of the individual NWP**



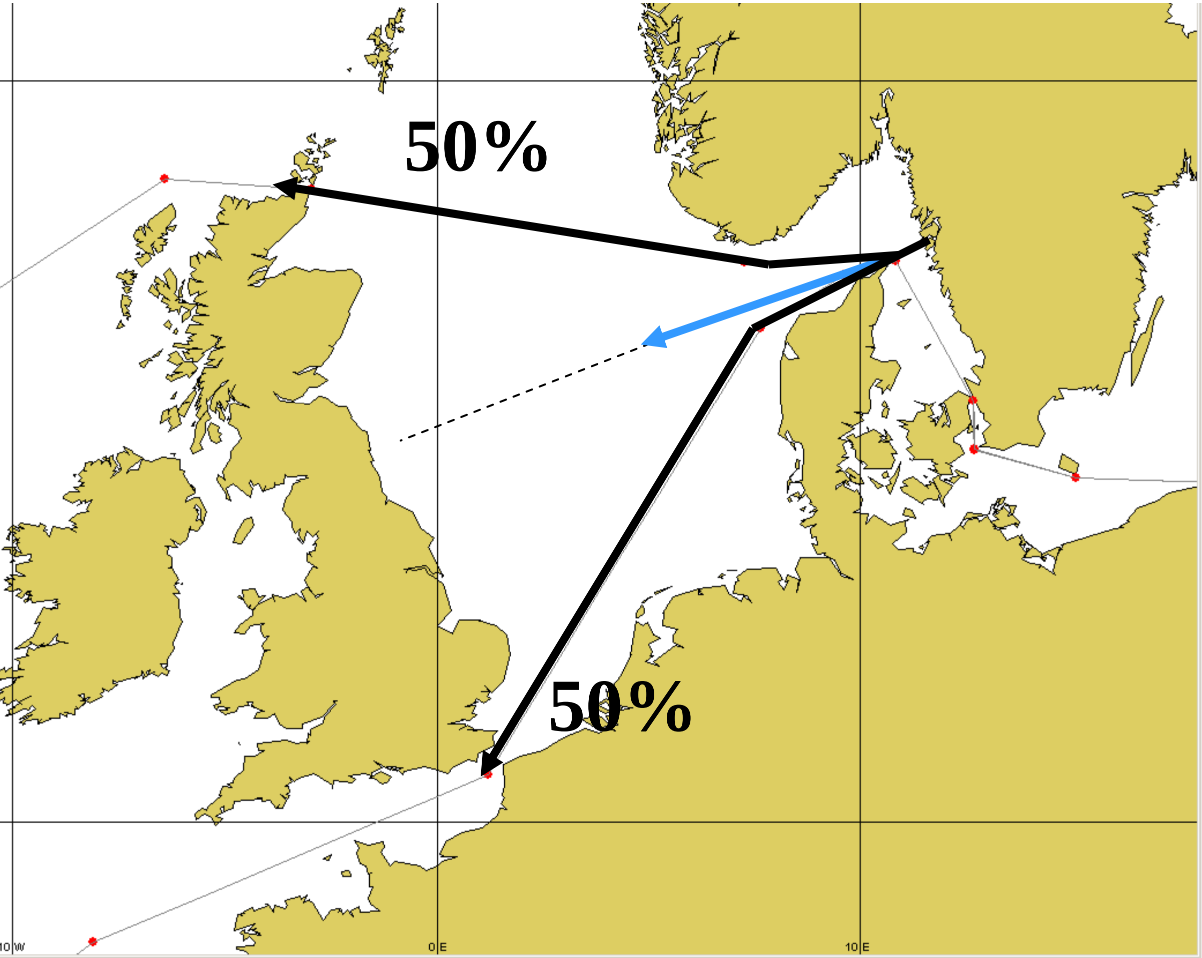
3. **A discussion on statistical inference is still needed . . .**

A common objection to the use of mean forecasts:

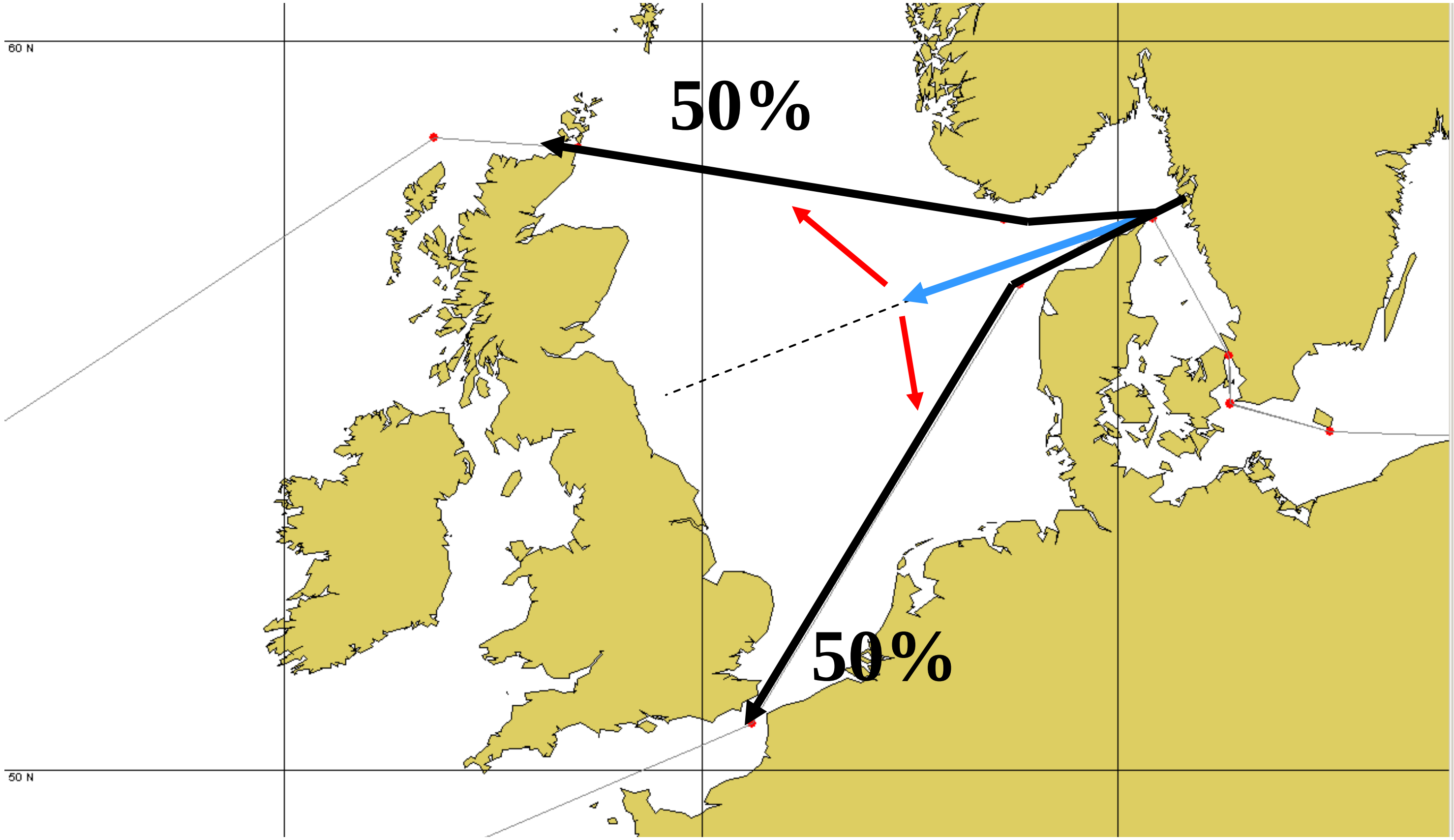
-It may lead to absurdities in bi-modal situations

A ship is leaving Gothenburg heading for the North Atlantic. Half of the indications point to taking the northerly route, half the Channel route

Using the “ensemble mean” would of course steer the ship towards Newcastle harbour!



But this is exactly what the ship routers would advice, as a “stand-by”



waiting for later, and hopefully, more reliable information

To repeat: The "Best Data" Paradox

Ensemble means are accurate, not "jumpy" and consistent with probabilities, but not necessarily "physically realistic"

Direct model output is physically realistic but, less accurate, very "jumpy" and not consistent with the probabilities

As we will see later in the course, interpreting the mean error is among the most difficult and treacherous things in science

Even more difficult than interpreting the standards **Root Mean Square Error** (RMSE) and the **Anomaly Correlation Coefficient** (ACC)

II.1.4 The Root Mean Square Error (RMSE)

A simple but powerful equation:

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

- $(a-b)^2$: Forecast error
- a^2: The NWP model behaviour
- b^2: The atmospheric behaviour
- $2ab$: Forecast agreement

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.

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

If we lived in an ideal world a lower RMSE would always be good and a higher RMSE always bad

But we don't, so . . .

What looks good might be bad, what looks bad might be good (Tim Palmer)

The Forecast Error equation

$$\overline{(f-o)^2}$$

f=forecast
o=observation
c=climate of the verifying day

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

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

- Forecast error
- The NWP model variability
- The atmospheric variability
- Forecast agreement or “skill”

II.1.5 Understanding the Anomaly Correlation Coefficient (ACC) and its relation to the RMSE

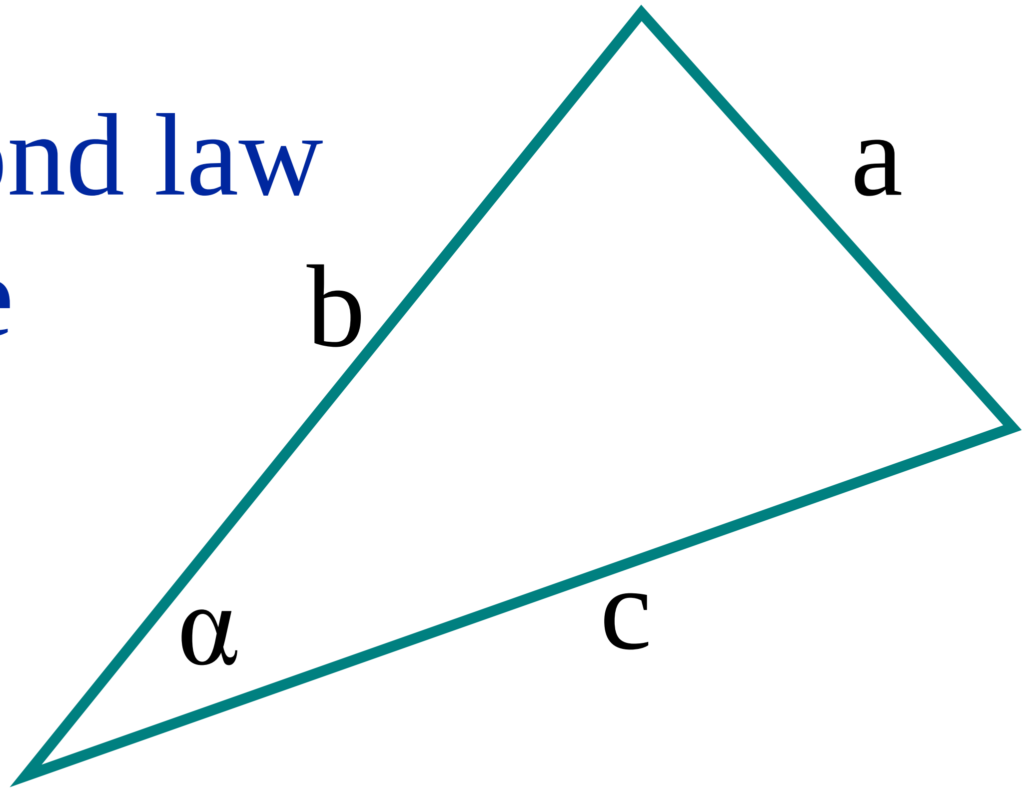
The RMSE:

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

The anomaly correlation coefficient:

$$ACC = \frac{\overline{(f-c)(a-c)}}{\overline{|f-c|}\;\overline{|a-c|}}$$

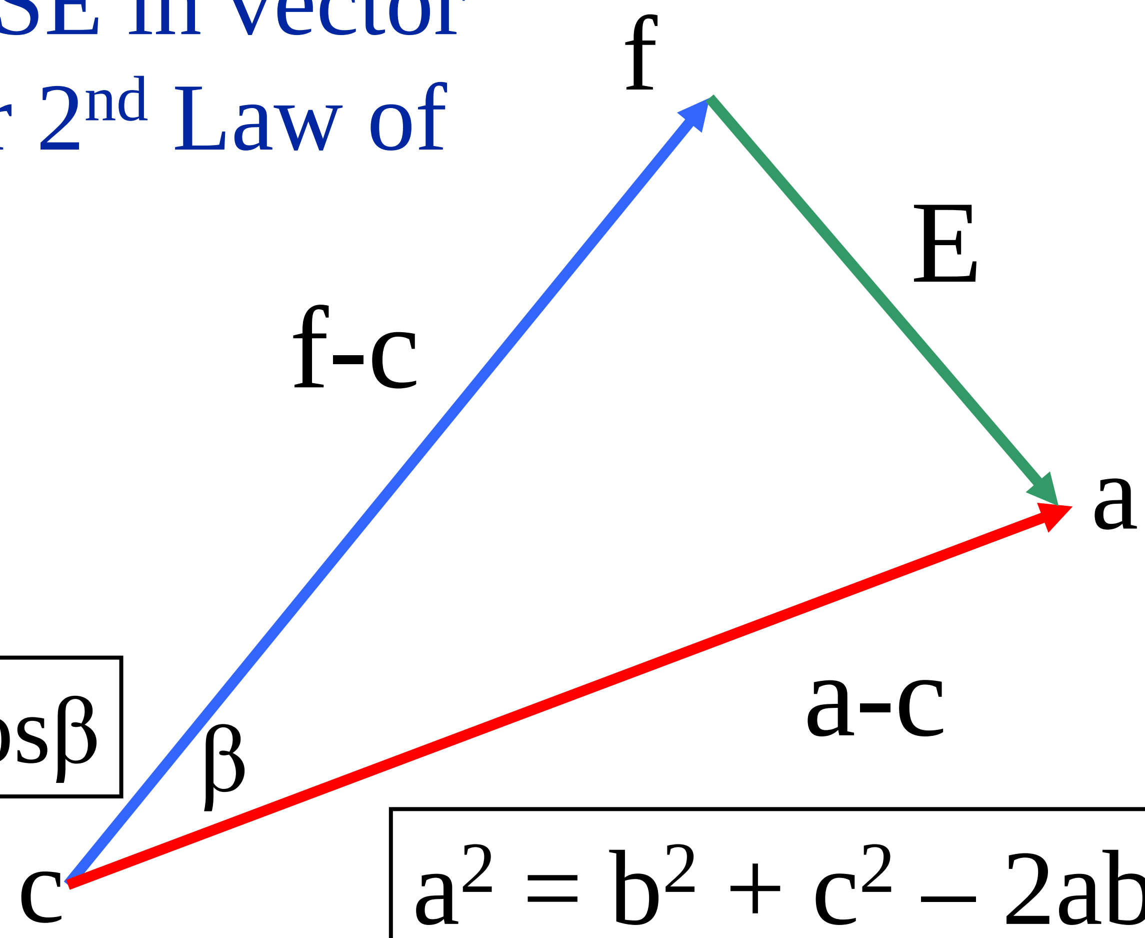
The second law of cosine



$$a^2 = b^2 + c^2 - 2ab \cdot \cos\alpha$$

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

The RMSE in vector Form (or 2nd Law of Cosine)

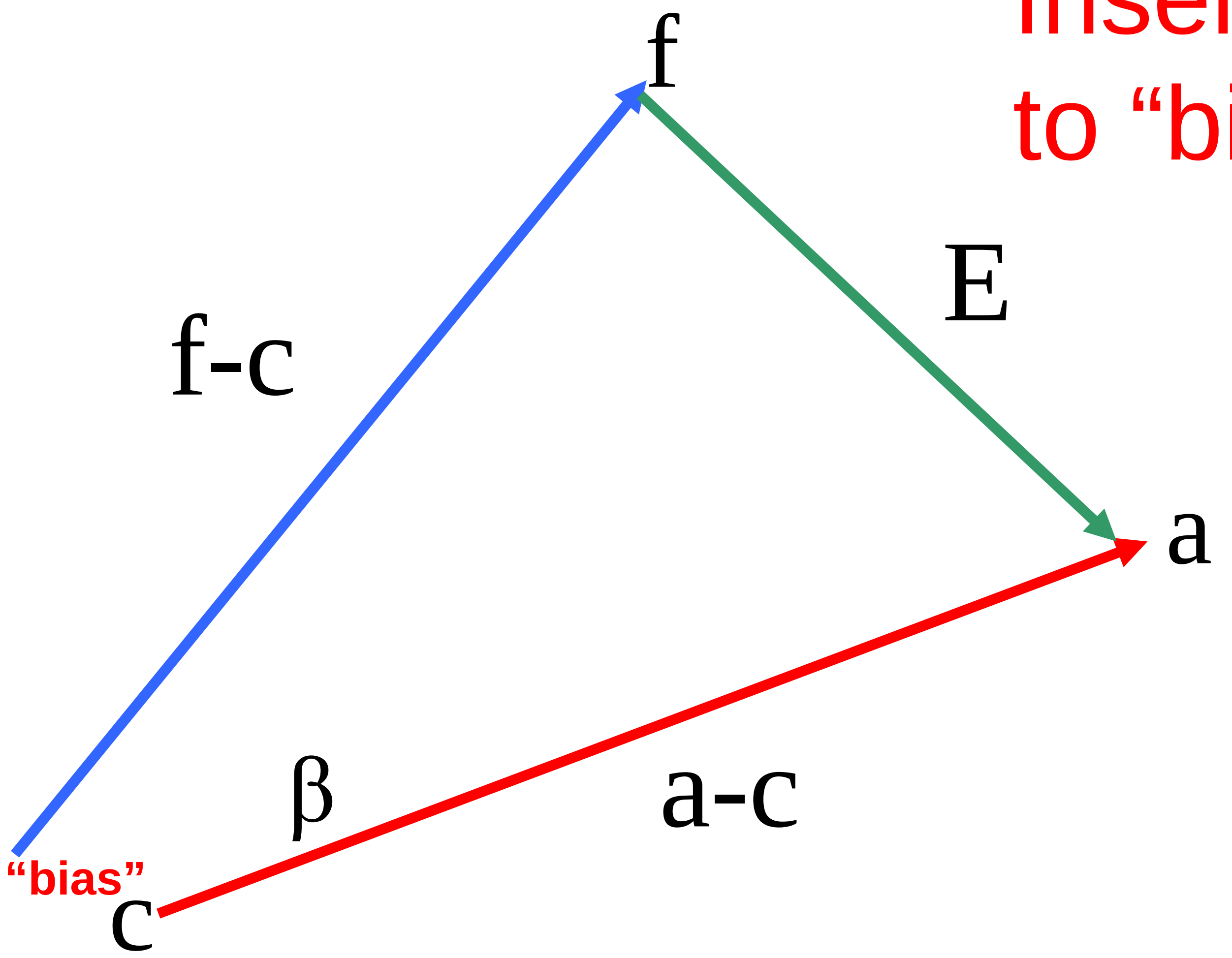


ACC=cosβ

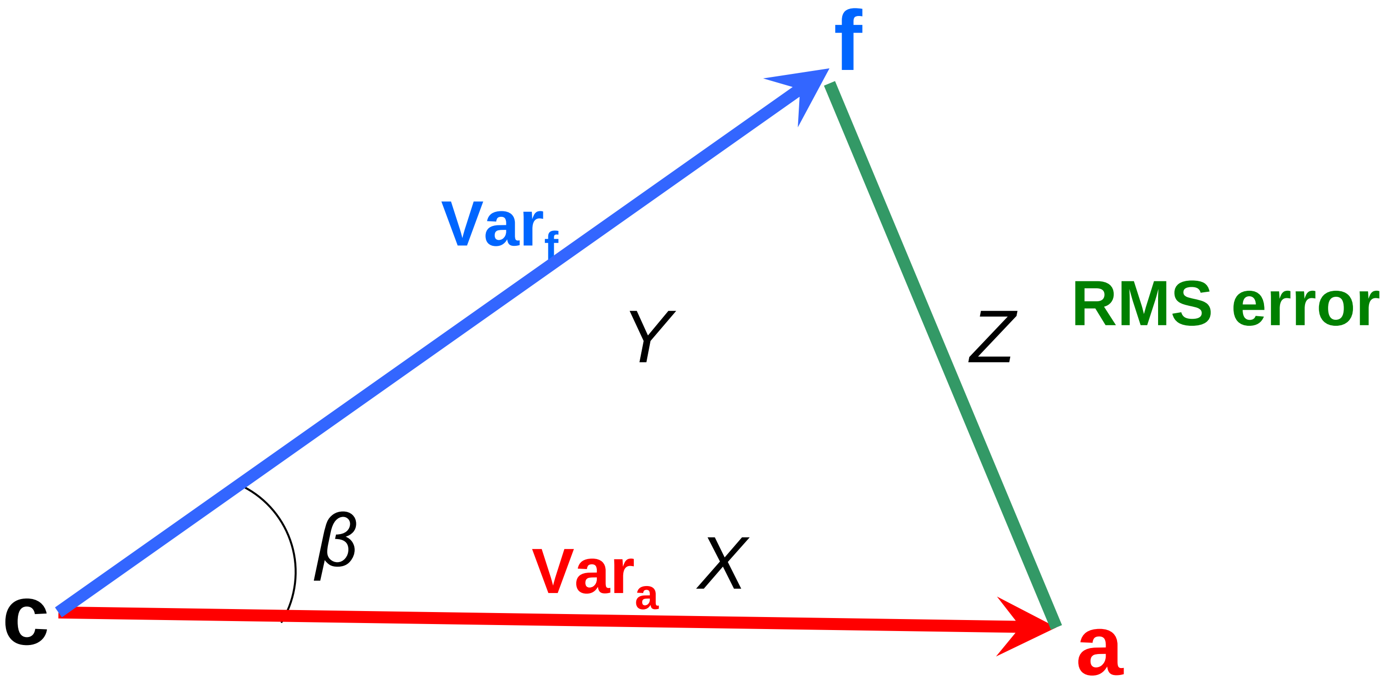
$a^2 = b^2 + c^2 - 2abc\cos\beta$

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

Insensitive to “biases”



Using the cosine theorem as a shortcut to understand the relation between RMS error, anomaly correlation coefficient (ACC) and model activity (variability)

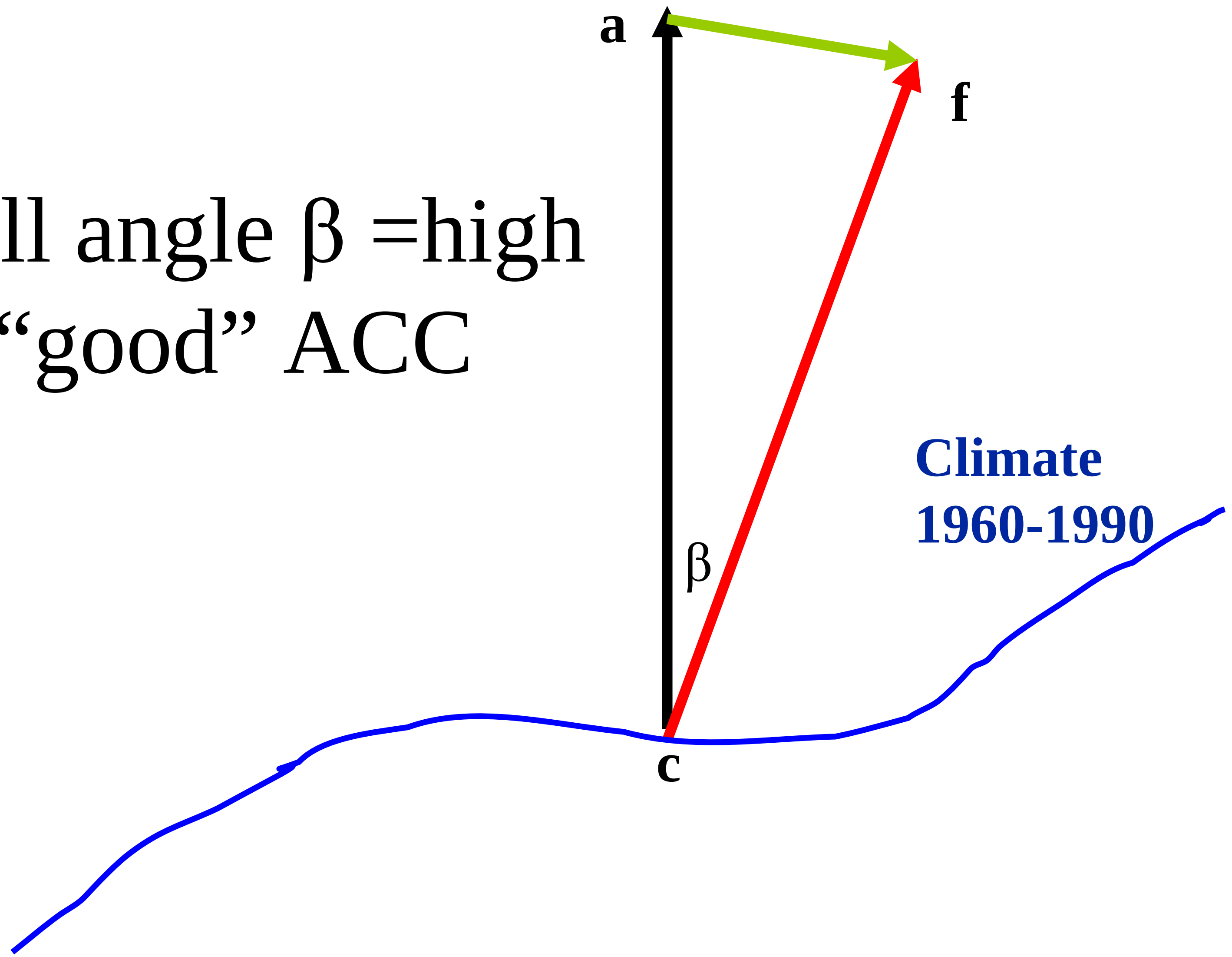


$cos\beta = ACC$

II.1.6 Interpreting the RMSE

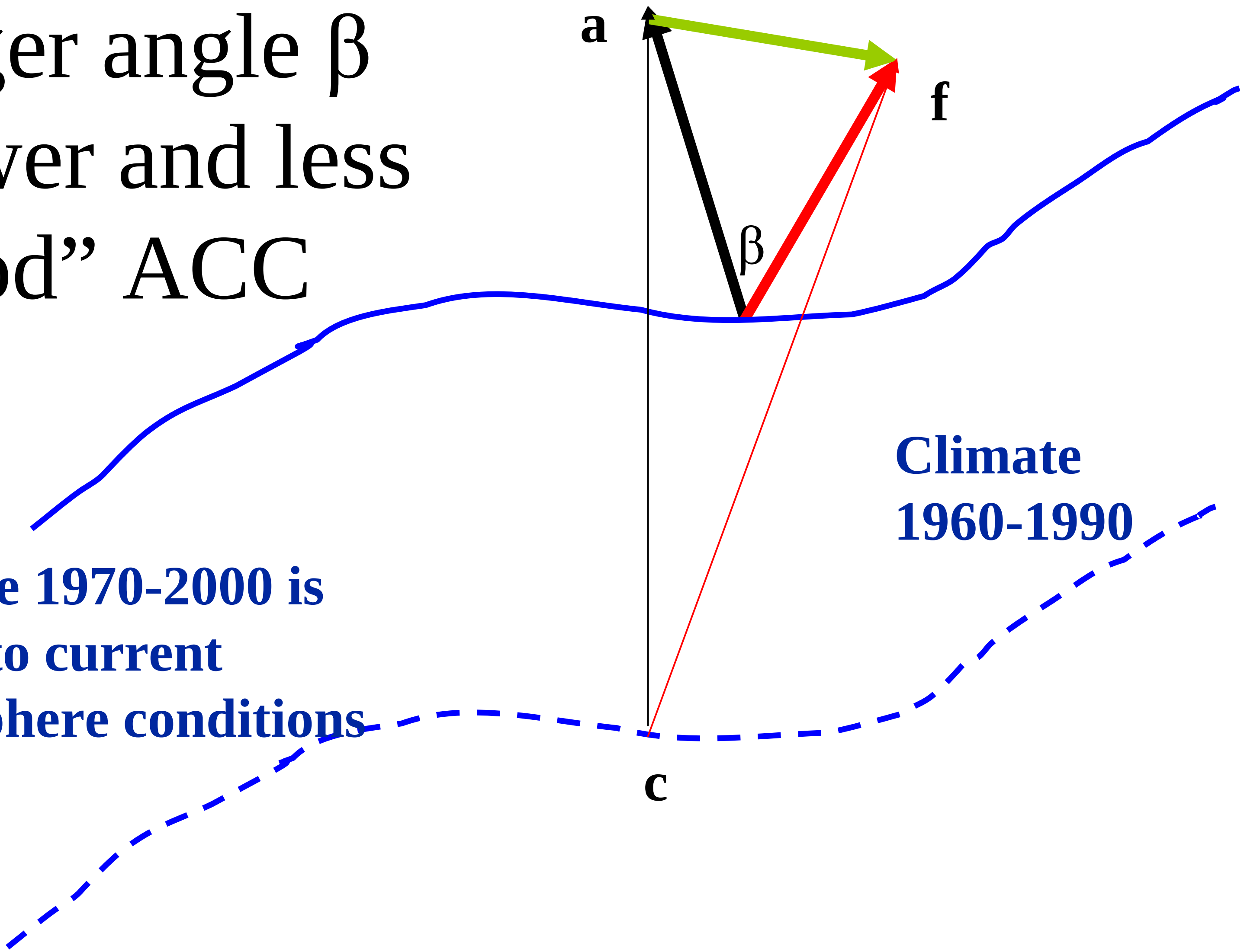
Anomaly correlations and climate reference

Small angle β =high and “good” ACC



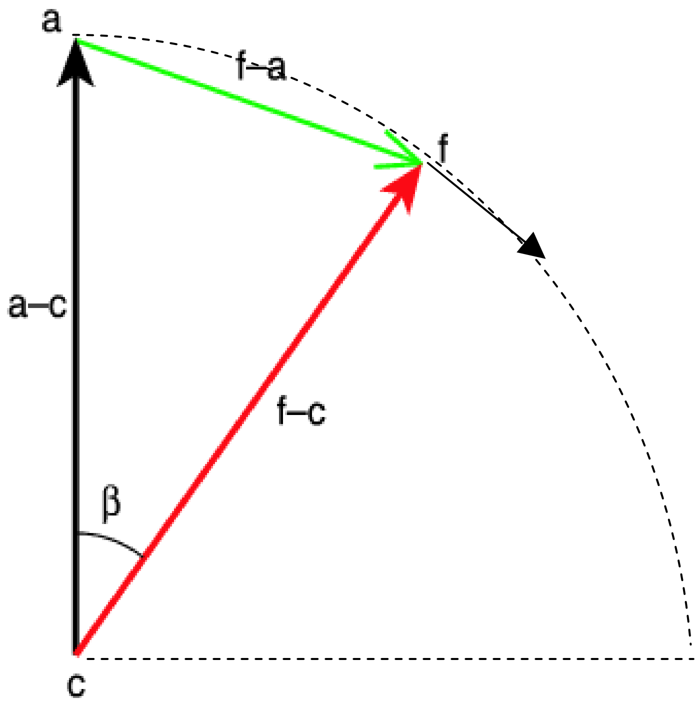
Anomaly correlations and climate reference

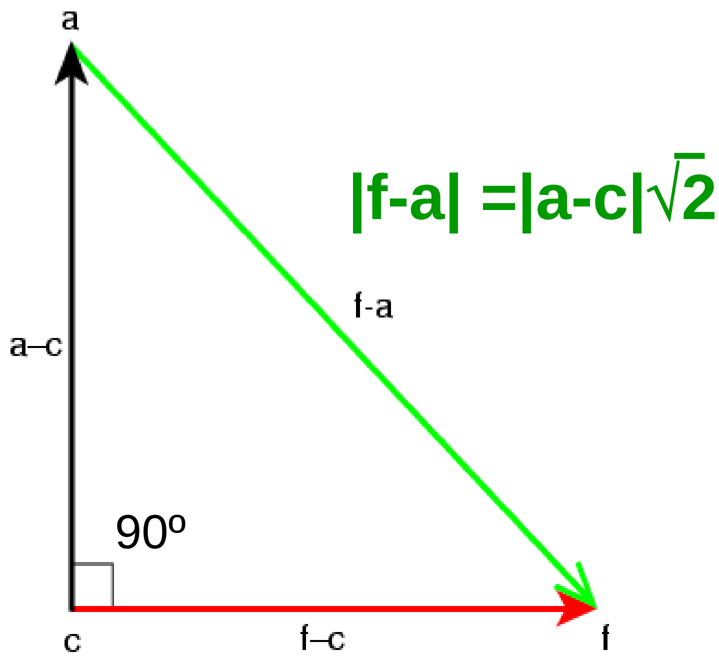
Larger angle β
=lower and less
"good" ACC



Climate 1970-2000 is closer to current atmosphere conditions

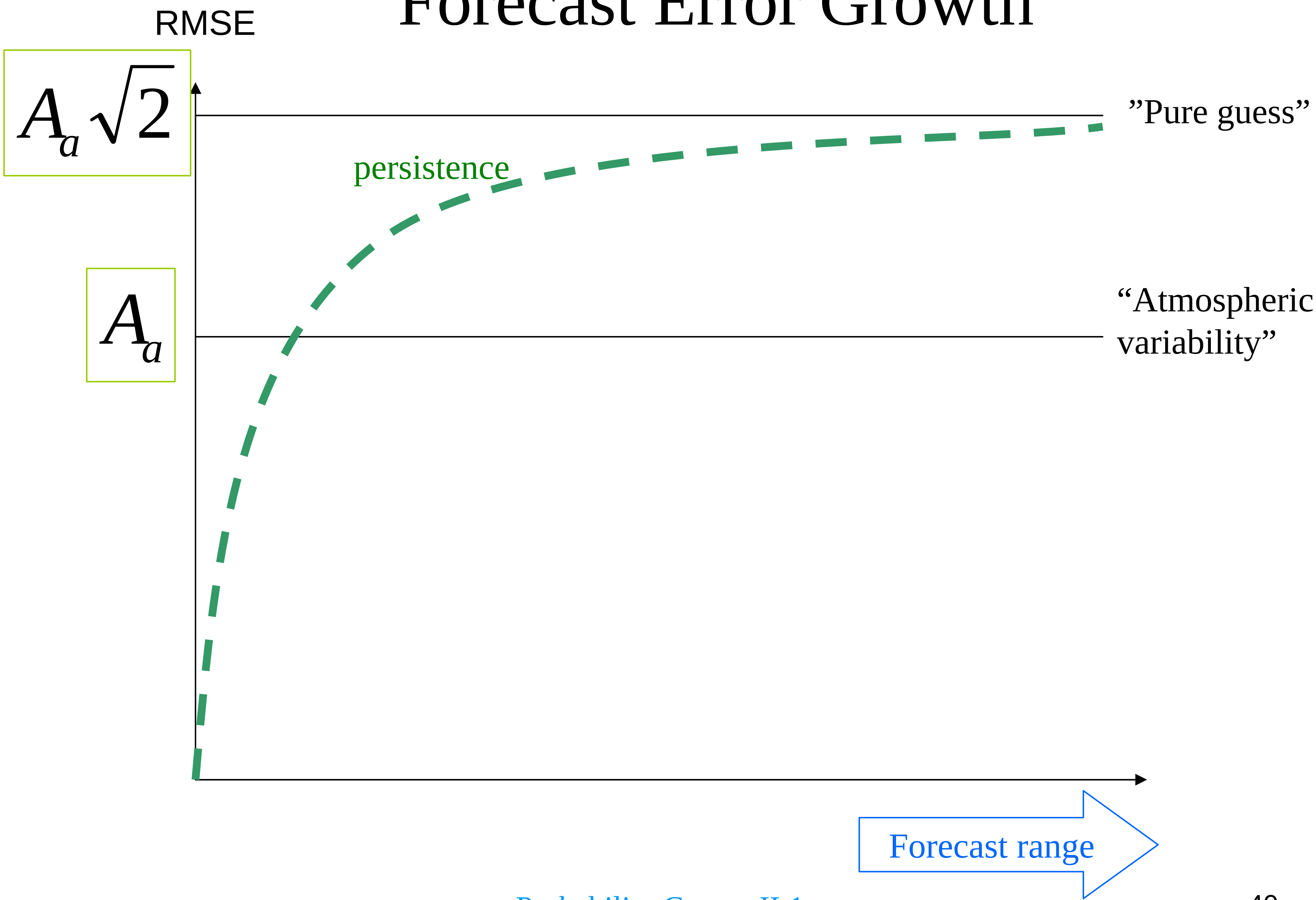
The RMS error saturation level



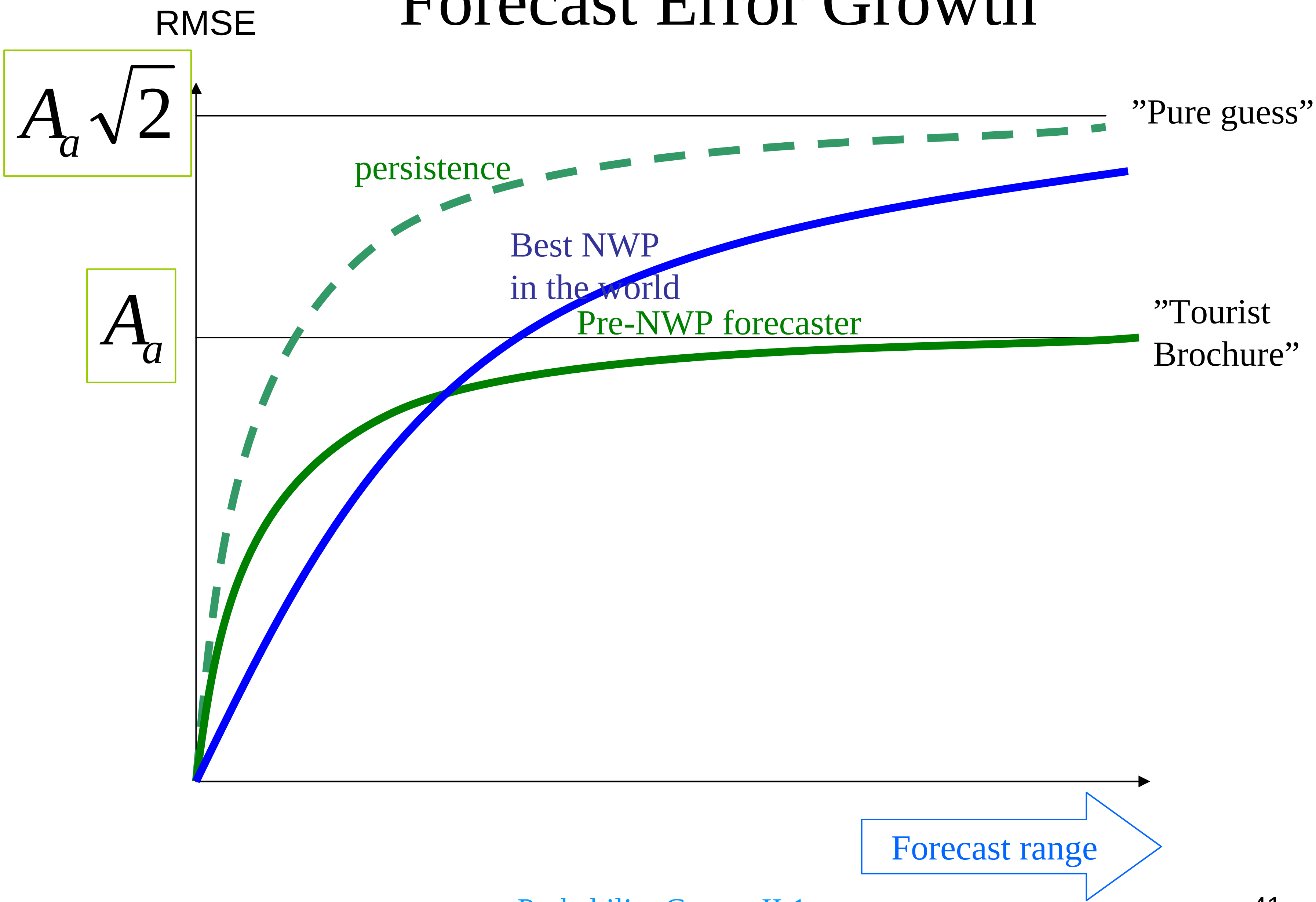


With decreasing forecast accuracy the angle β will increase.
The maximum RMSE equals the variability times √2

Forecast Error Growth

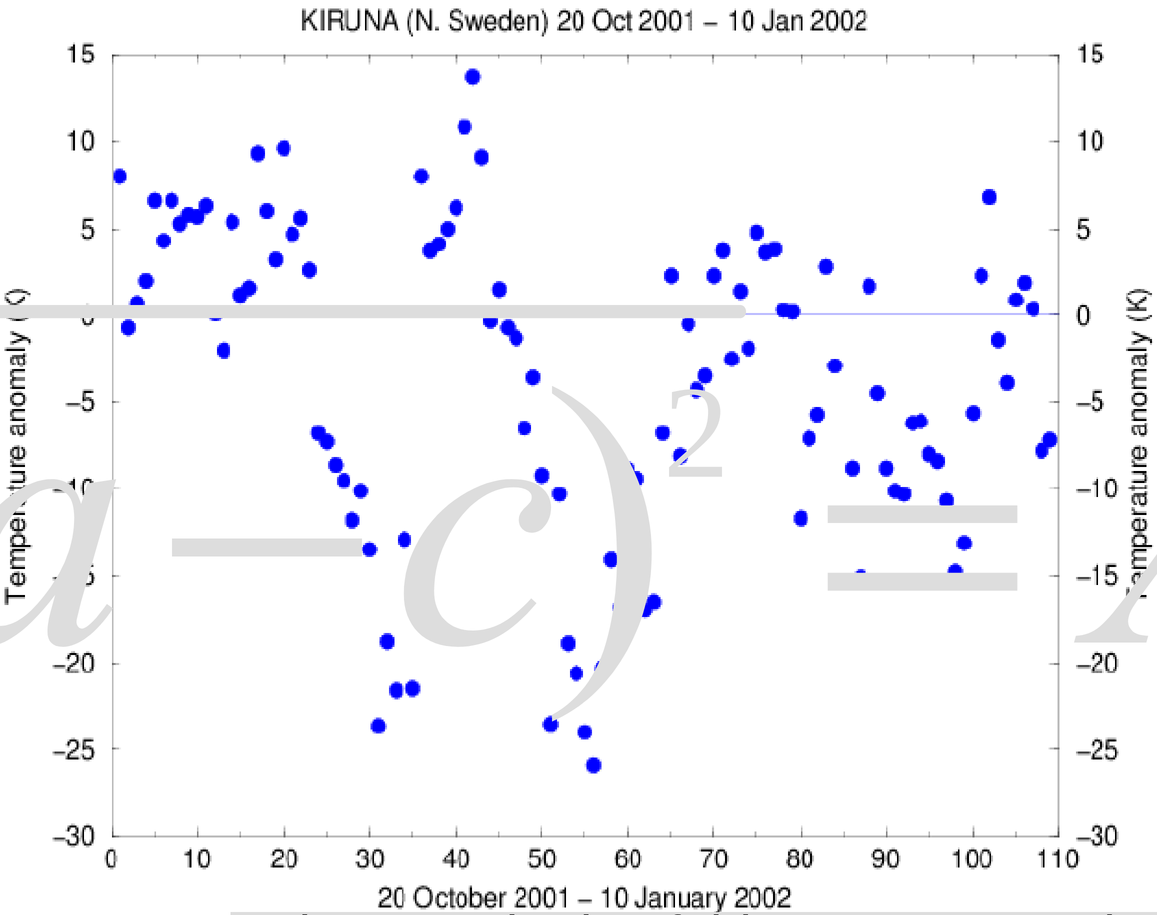


Forecast Error Growth



The interpretation of the $\overline{(a-c)}$ term

The *observed* variability around the climatological mean

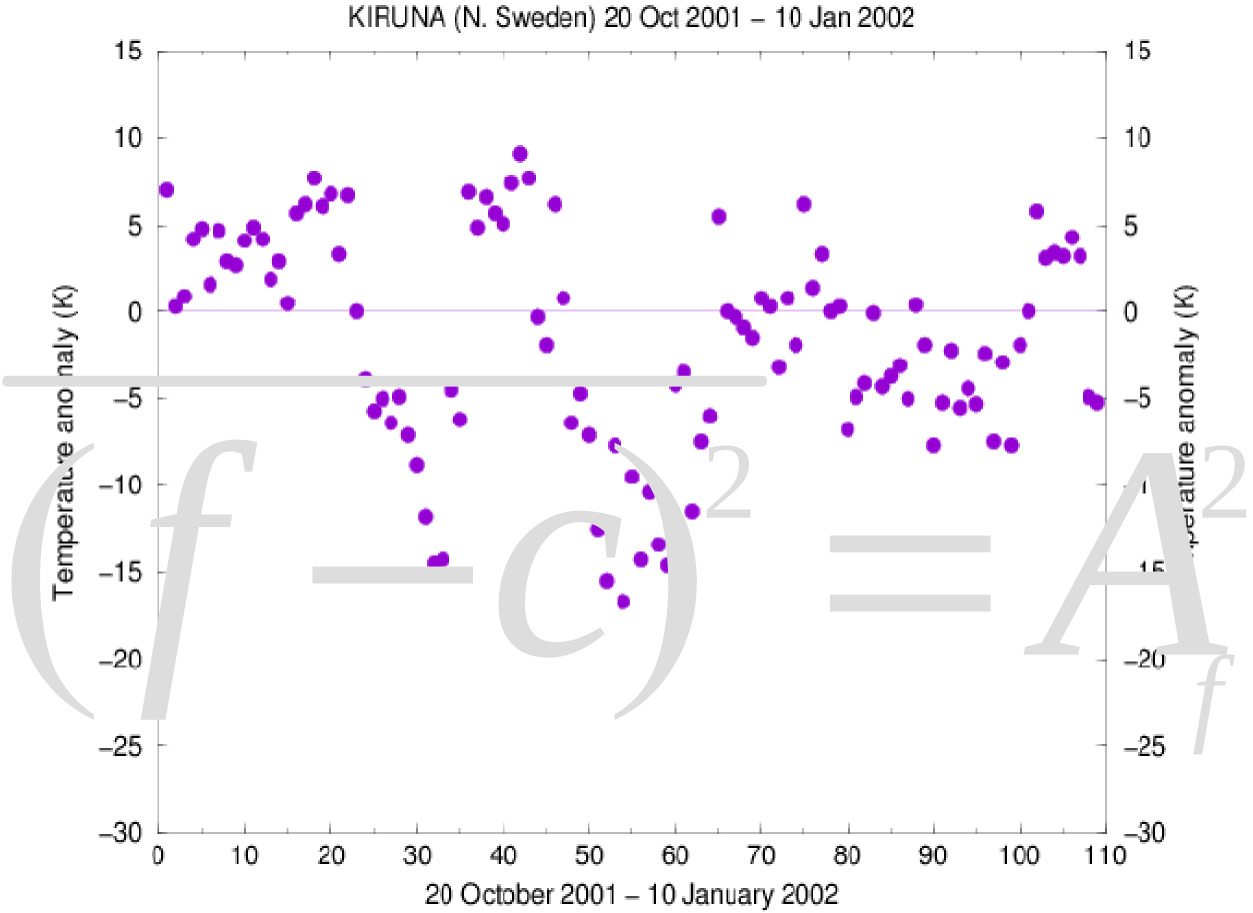


$$\overline{(a-c)^2} = A_a^2$$

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

The interpretation of the $\overline{(f-c)}$ term

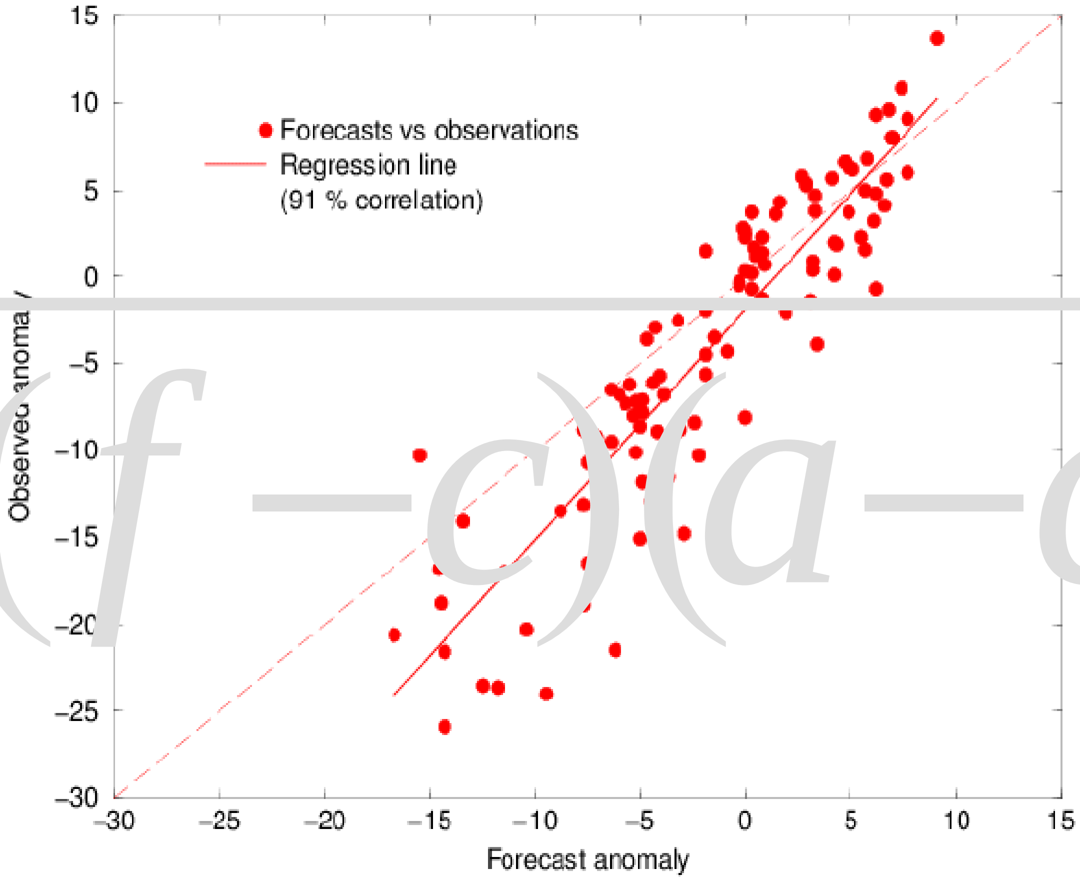
The *forecast* variability around the climatological mean



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

The interpretation of the "skill" term

The correspondence between $\overline{(f\text{-}c)}$ and $\overline{(a\text{-}c)}$

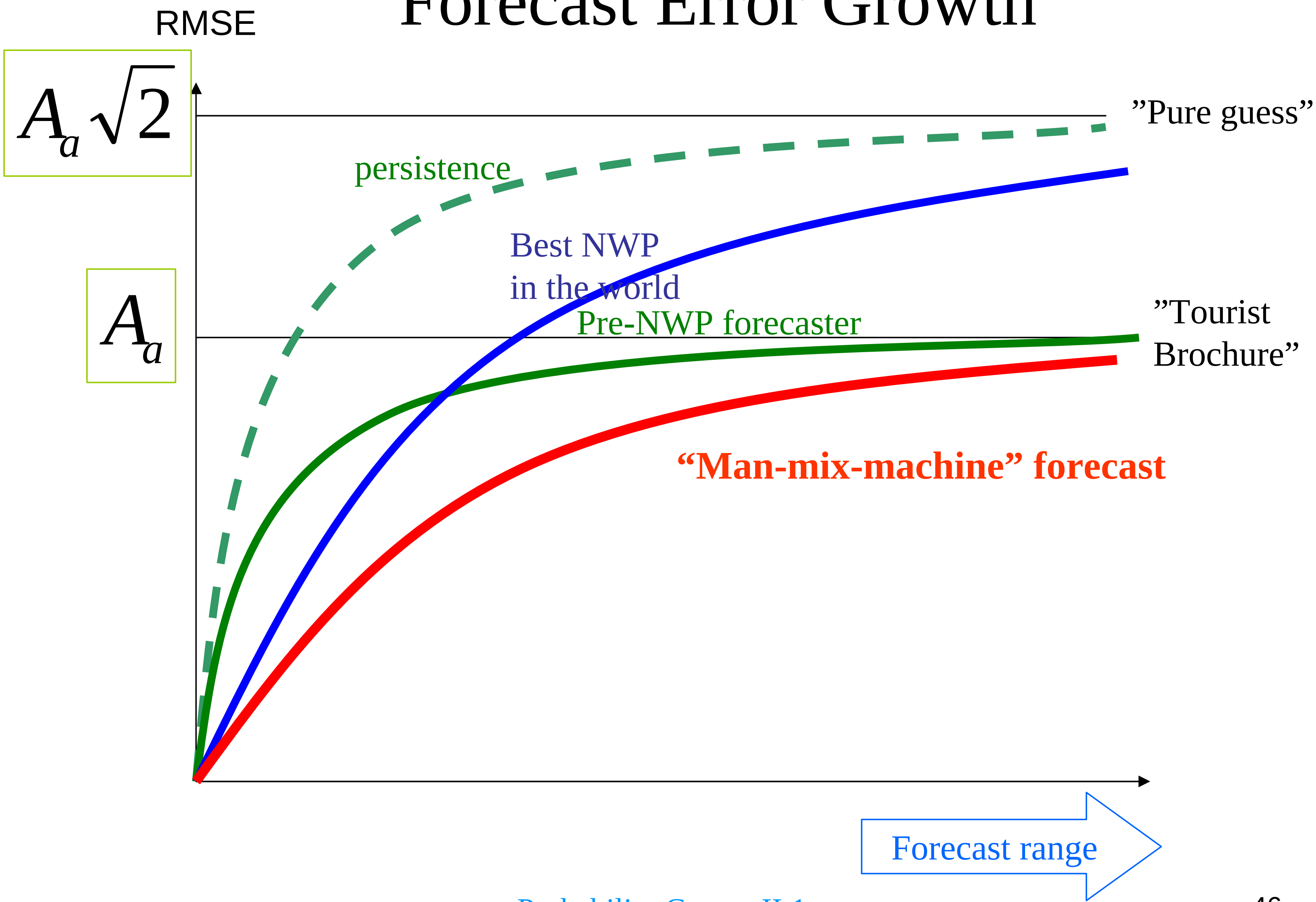


$$\overline{2(f-c)(a-c)}$$

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

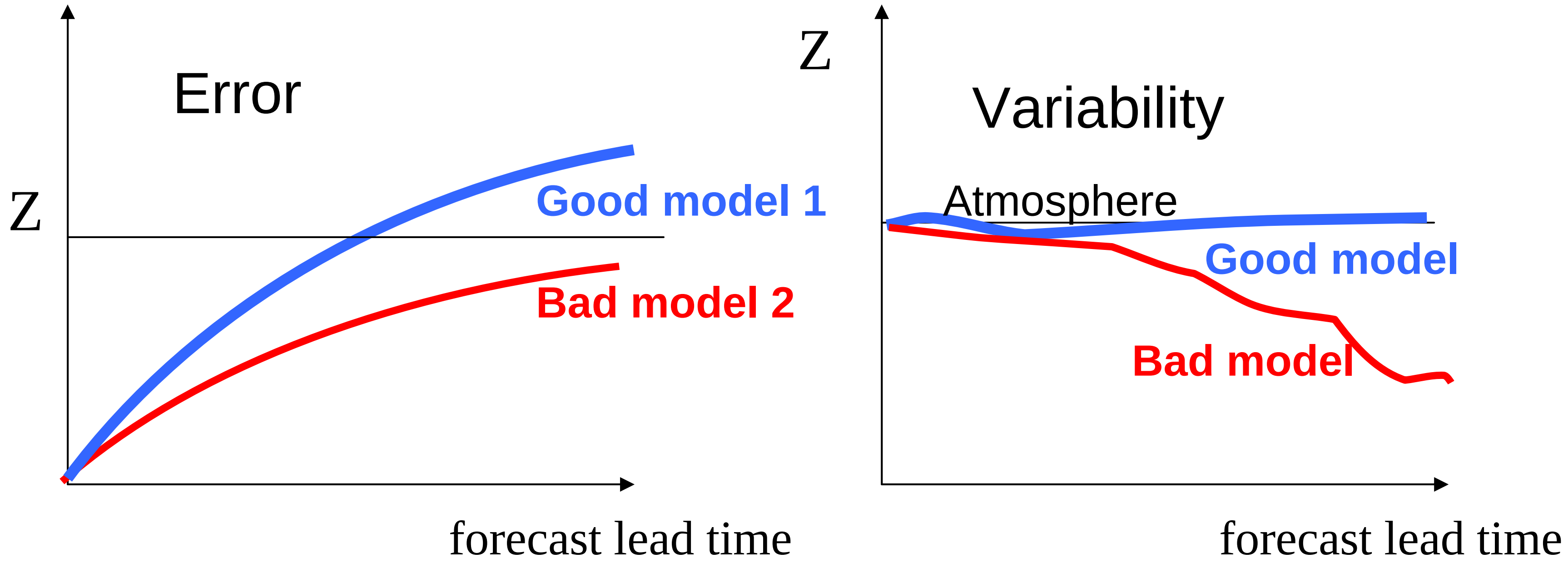
II.1.7 What looks good . . .

Forecast Error Growth



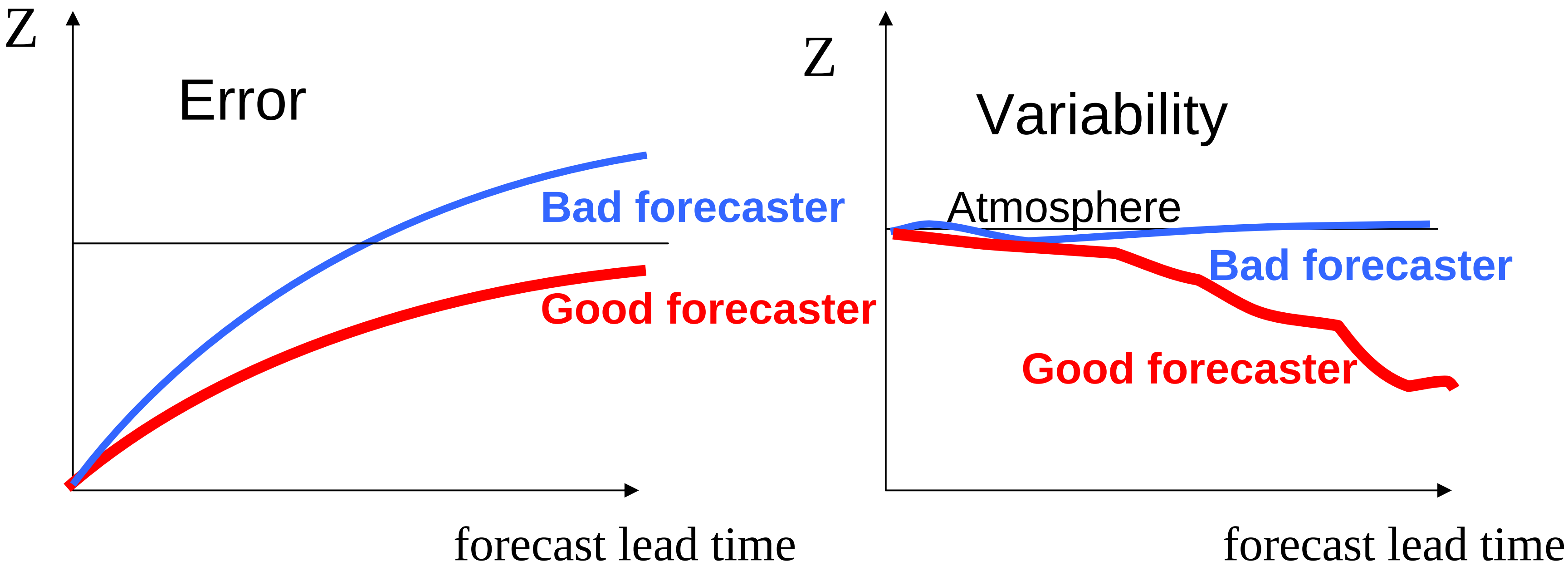
It is not trivial to compare a human forecaster with a NWP system since they strive for different objectives

The good versus the bad NWP model



The decrease in variability, and thus ability to simulate the atmospheric motions, may give low (good) RMSE verifications

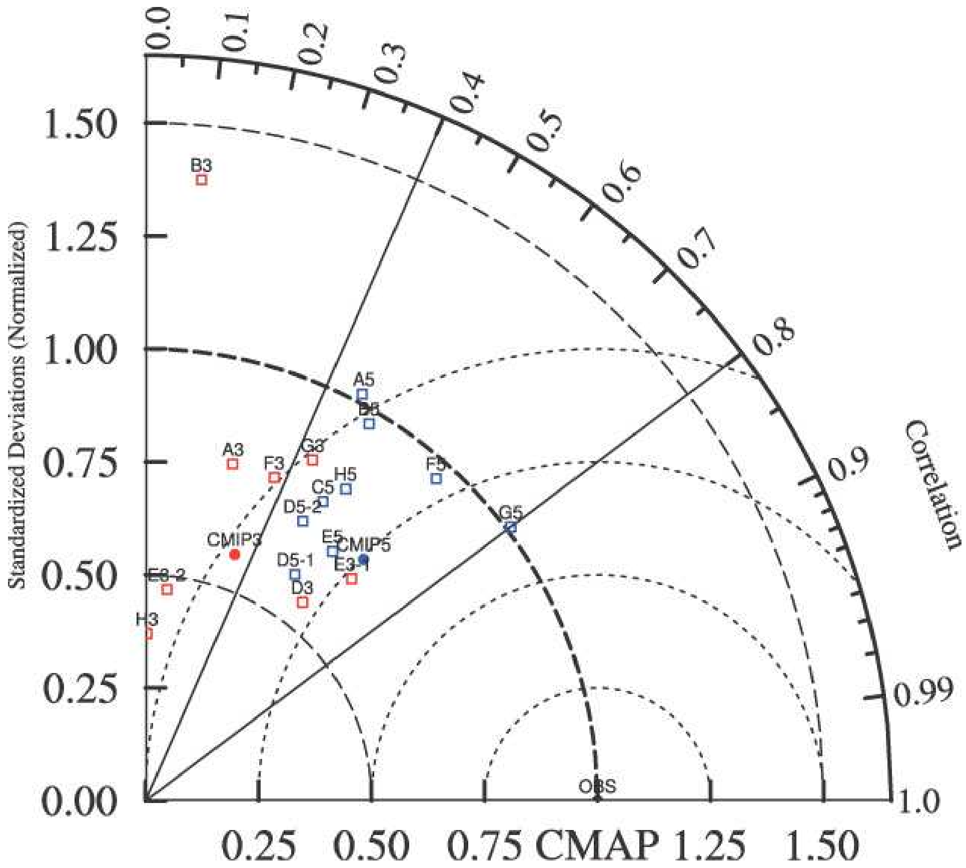
The good versus the bad forecaster

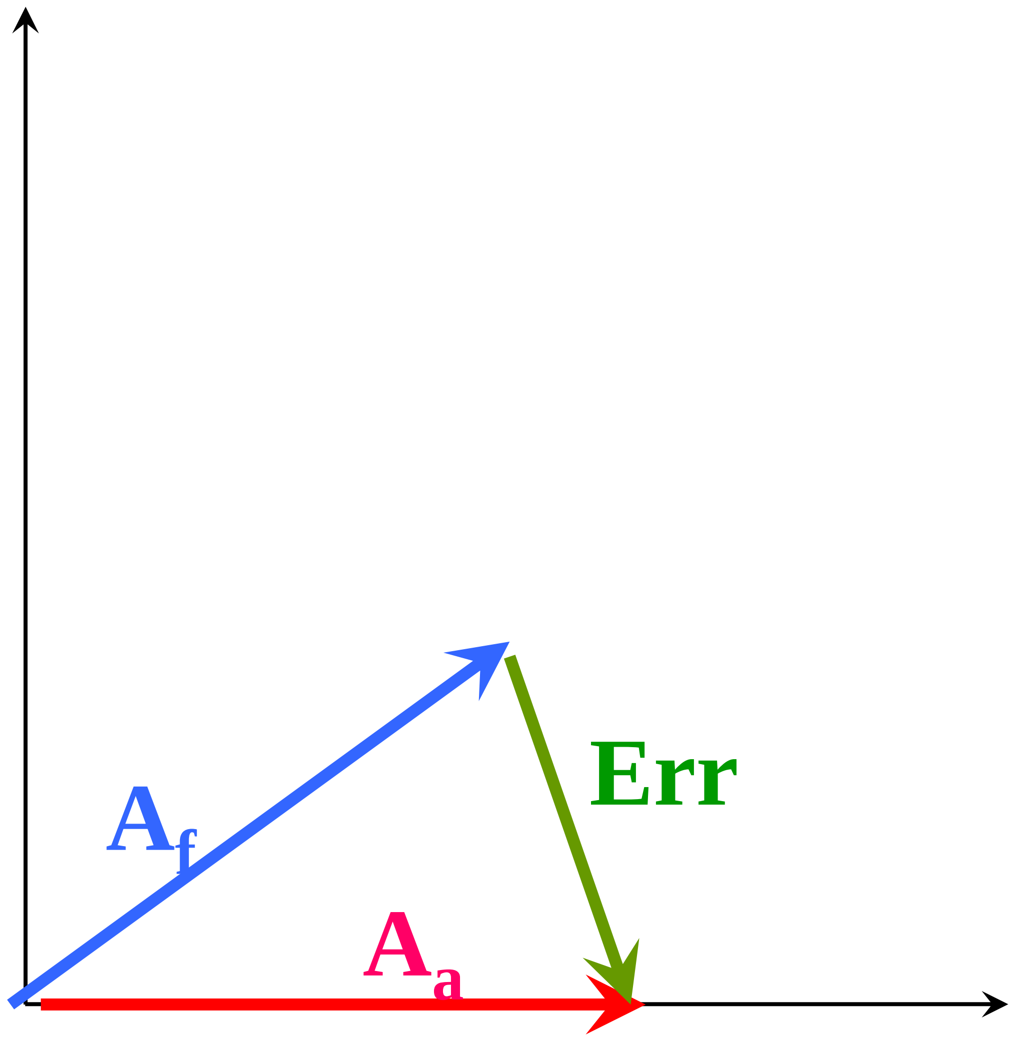


The decrease in variability, due to a skilful filtering of non-predictable atmospheric features, may yield low (good) RMSE verifications

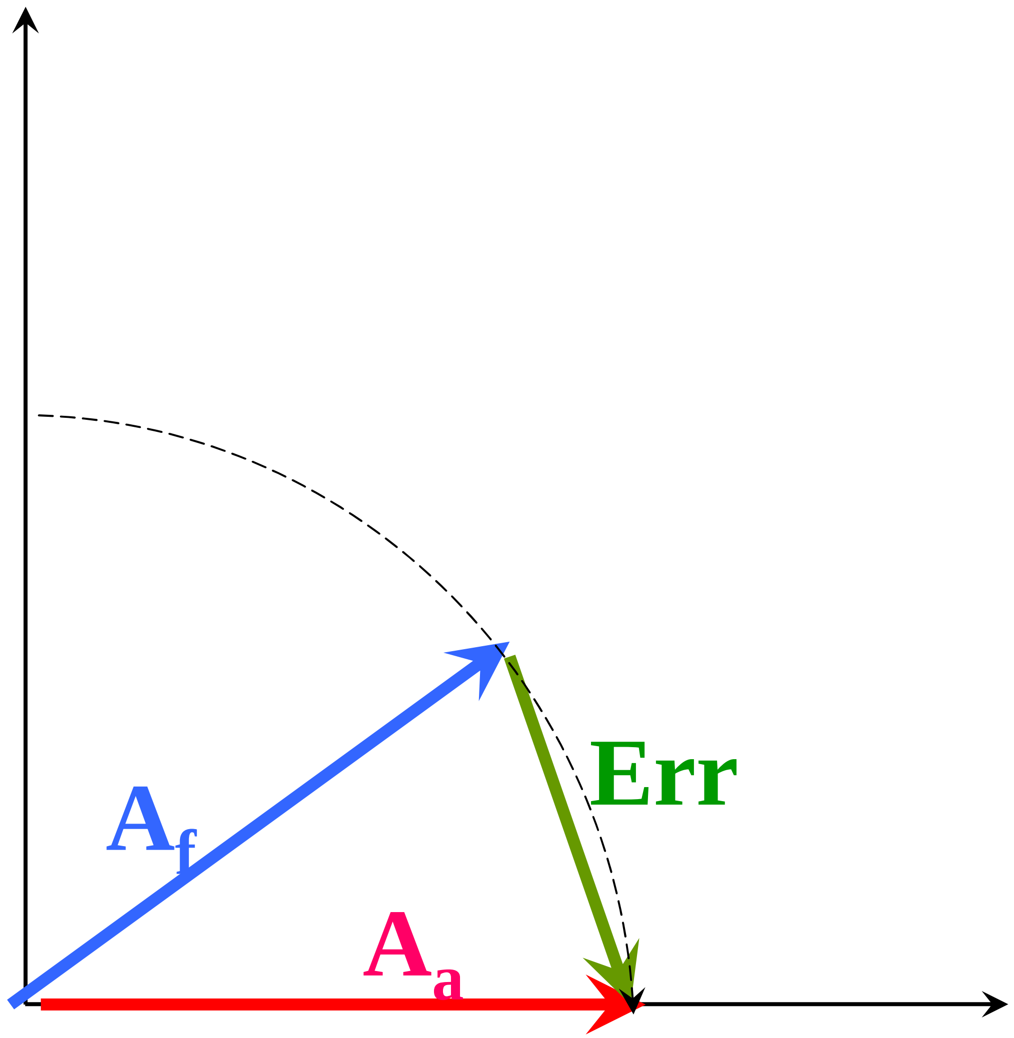
II.1.8 The Taylor diagram

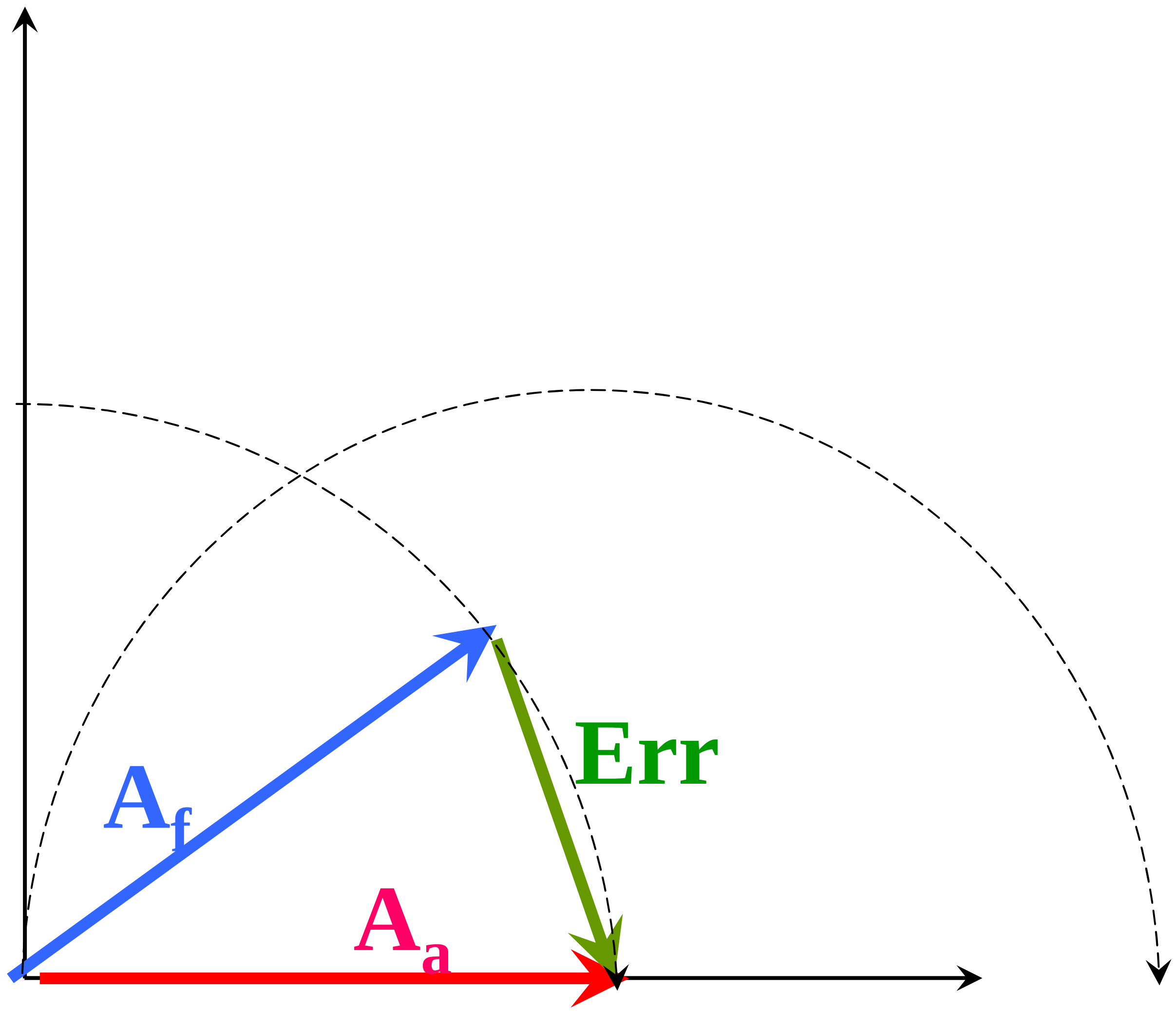
The Taylor Diagram

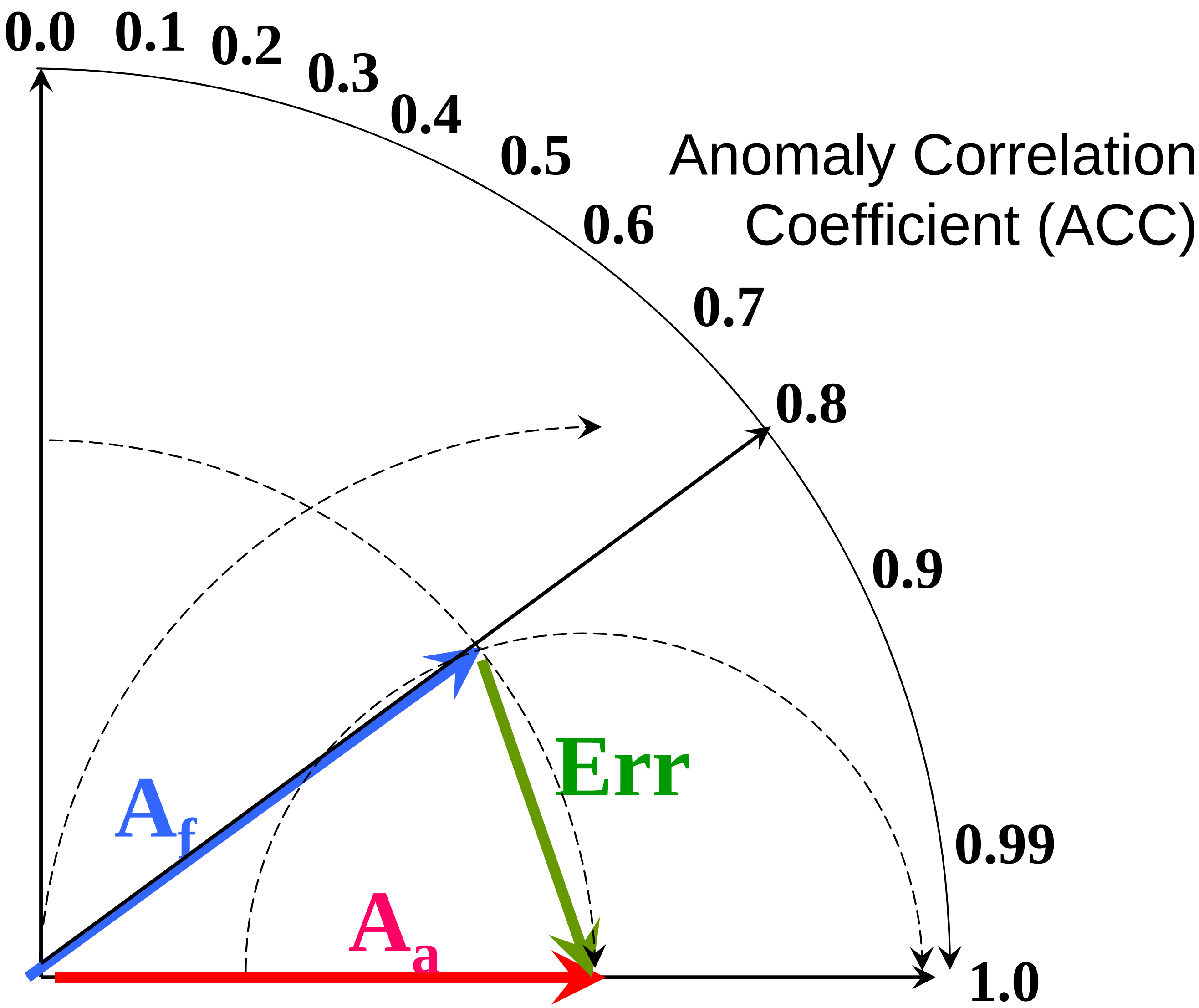




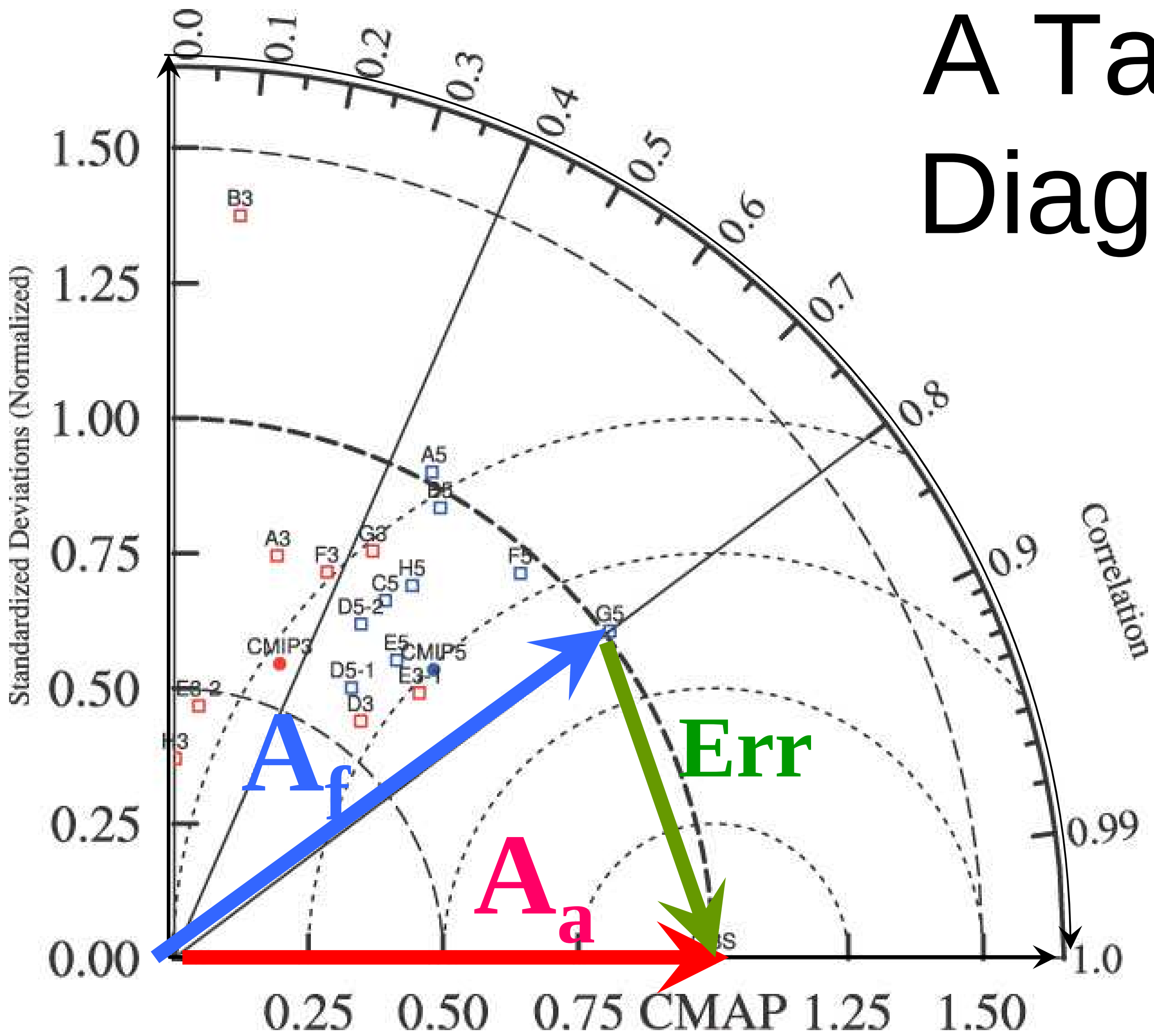
Err
A_f
A_a


A
f
Err
A
a


Err
Af
Aa

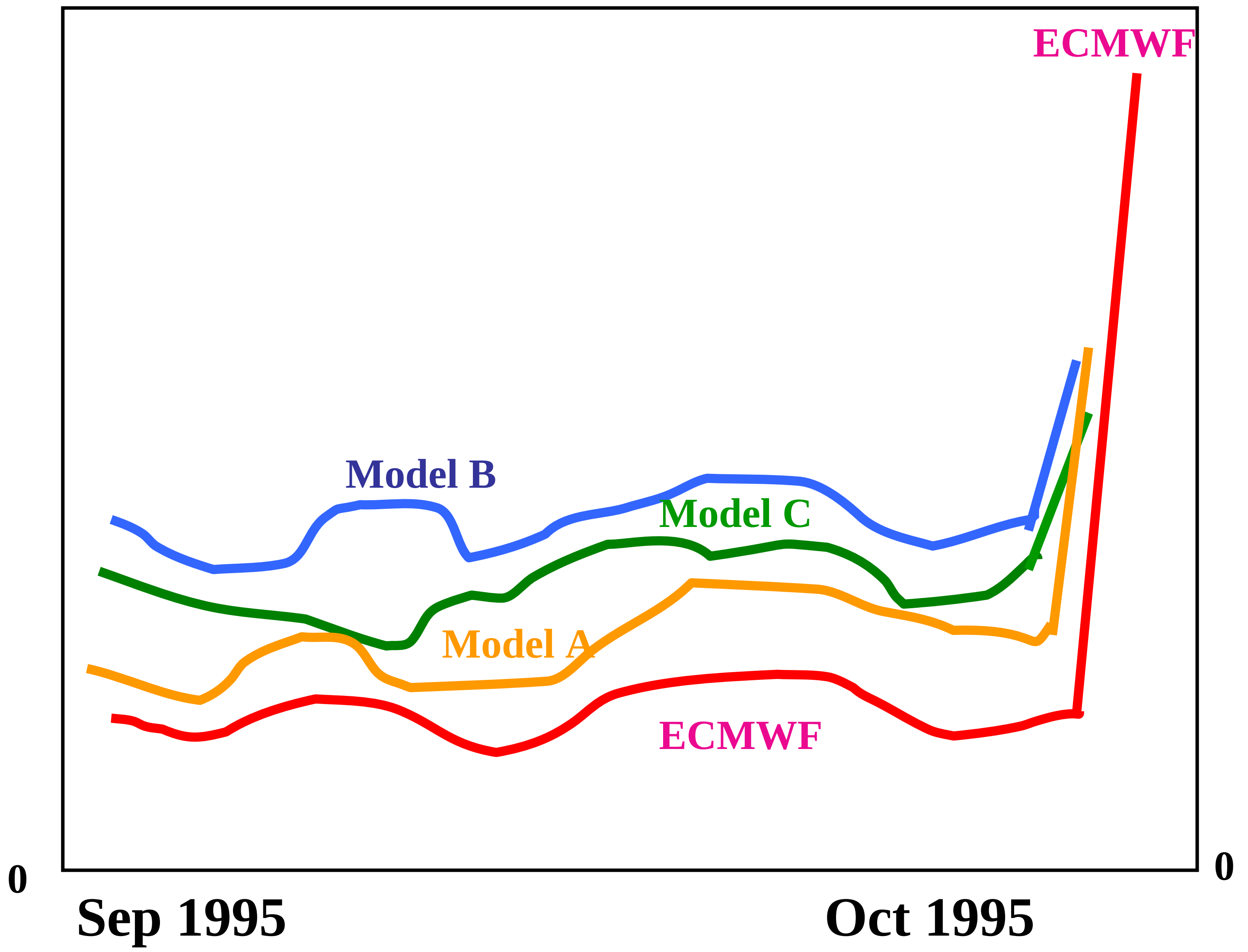

0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.99
1.0
Anomaly Correlation
Coefficient (ACC)
Err
A_f
A_a

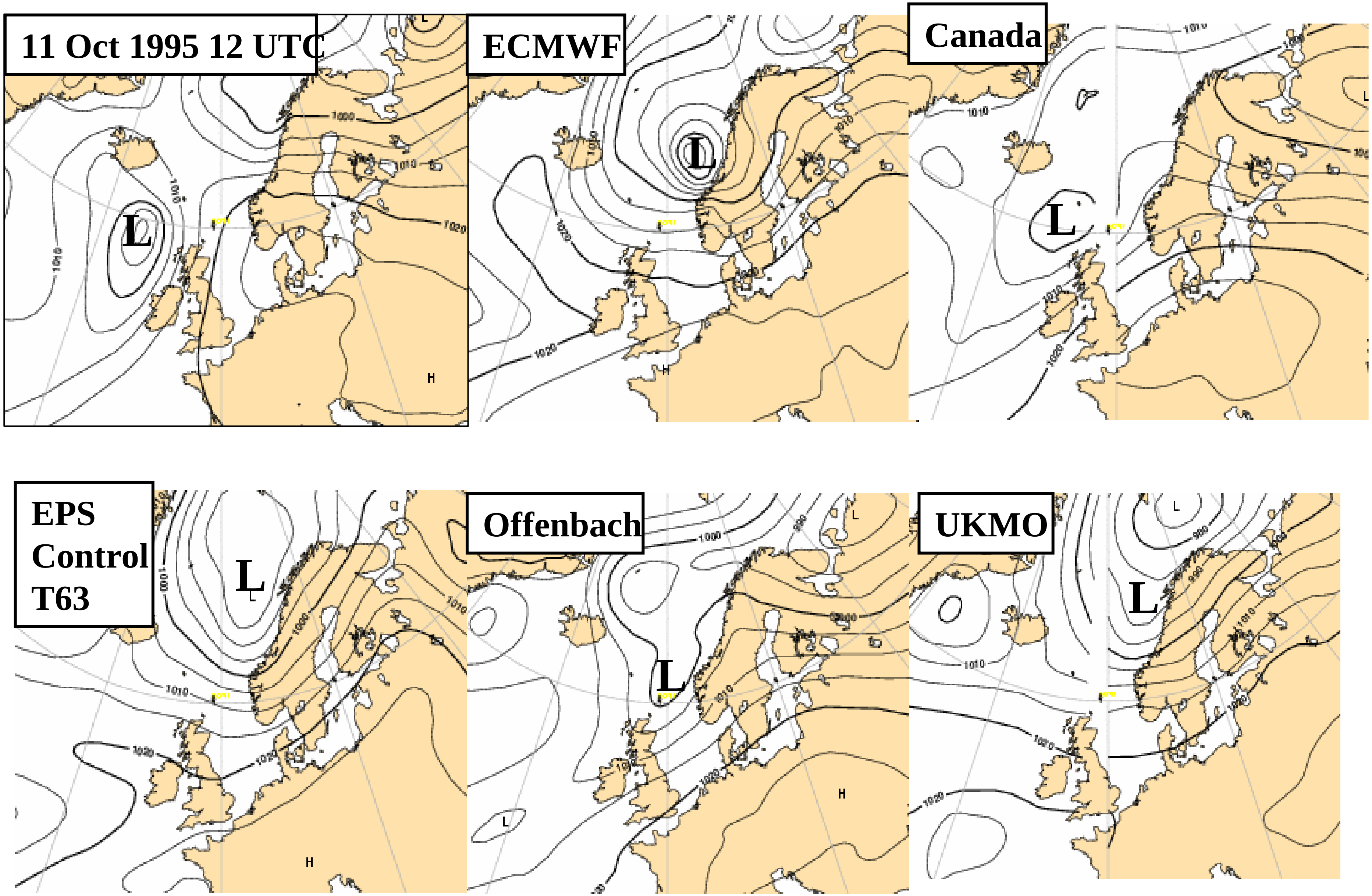
A Taylor Diagram

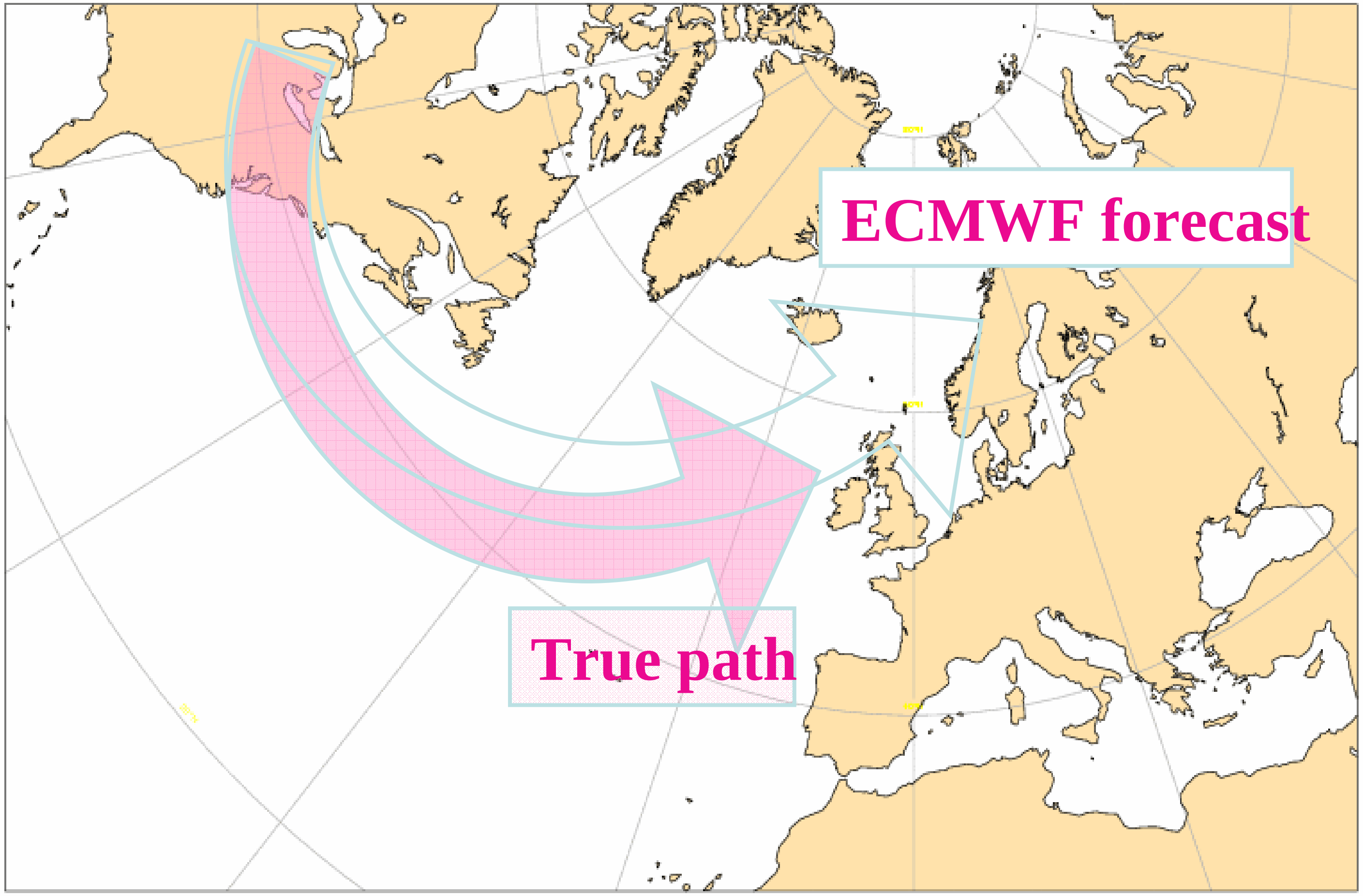


II.1.9. The double penalty effect

Root Mean Square Error (RMSE)




11 Oct 1995 12 UTC
ECMWF
Canada
EPS
Control
T63
Offenbach
UKMO

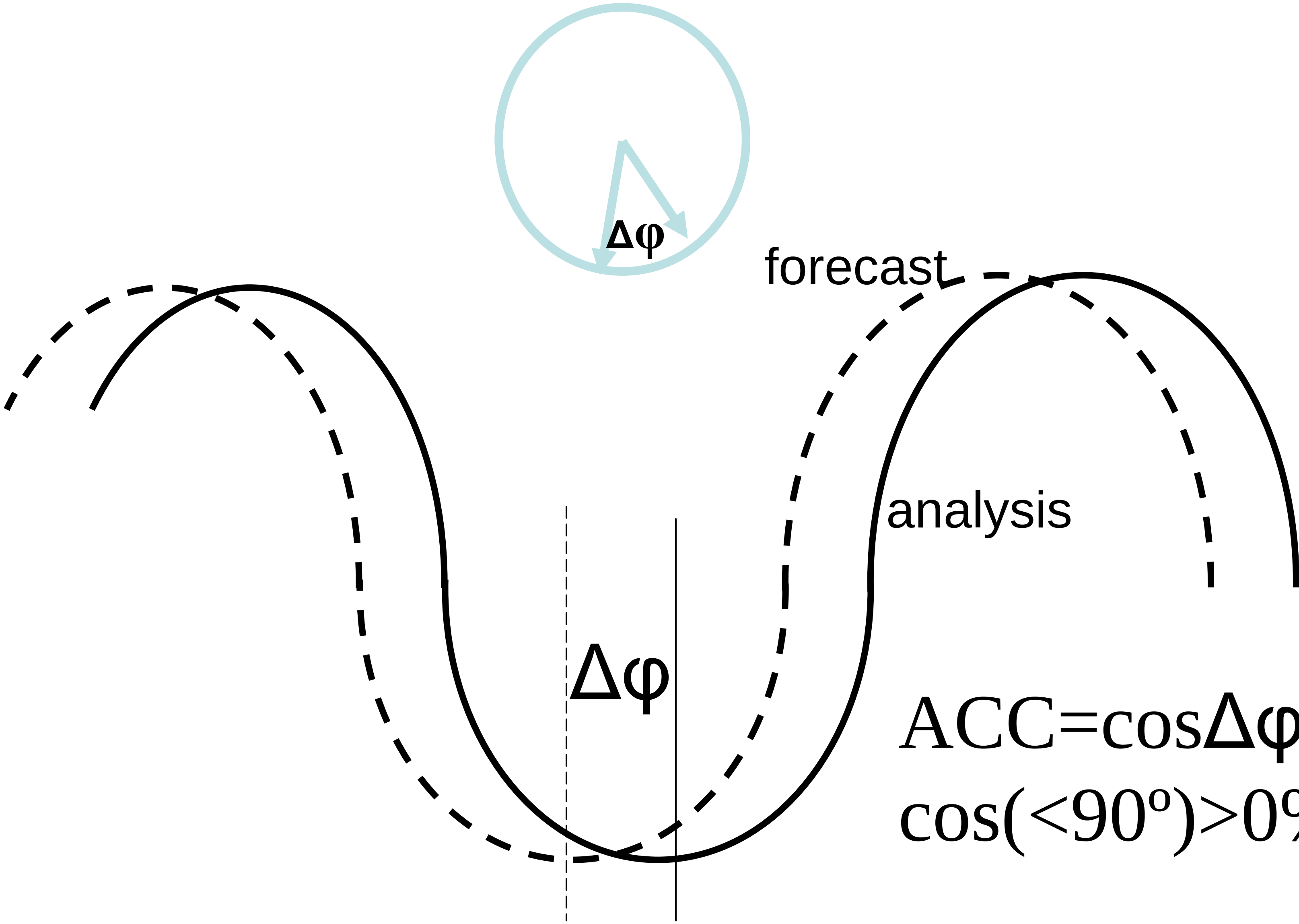

ECMWF forecast
True path

The ECMWF forecast scored worst because of

The “Double Penalty Effect”

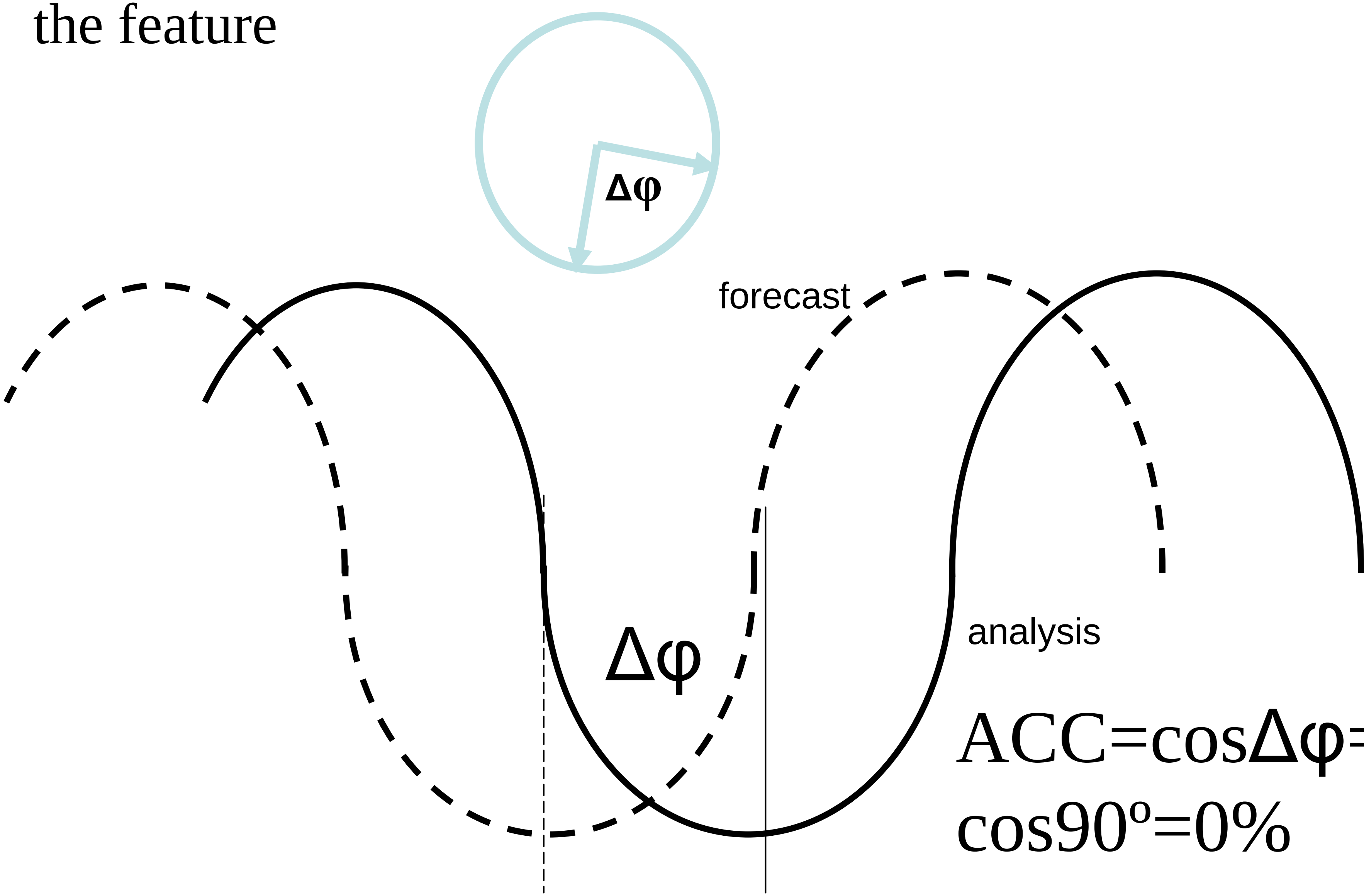
The forecast is punished both for having an anomaly where there isn’t one and not having an anomaly where there is one

If the phase error < ½ wave length there is still positive skill



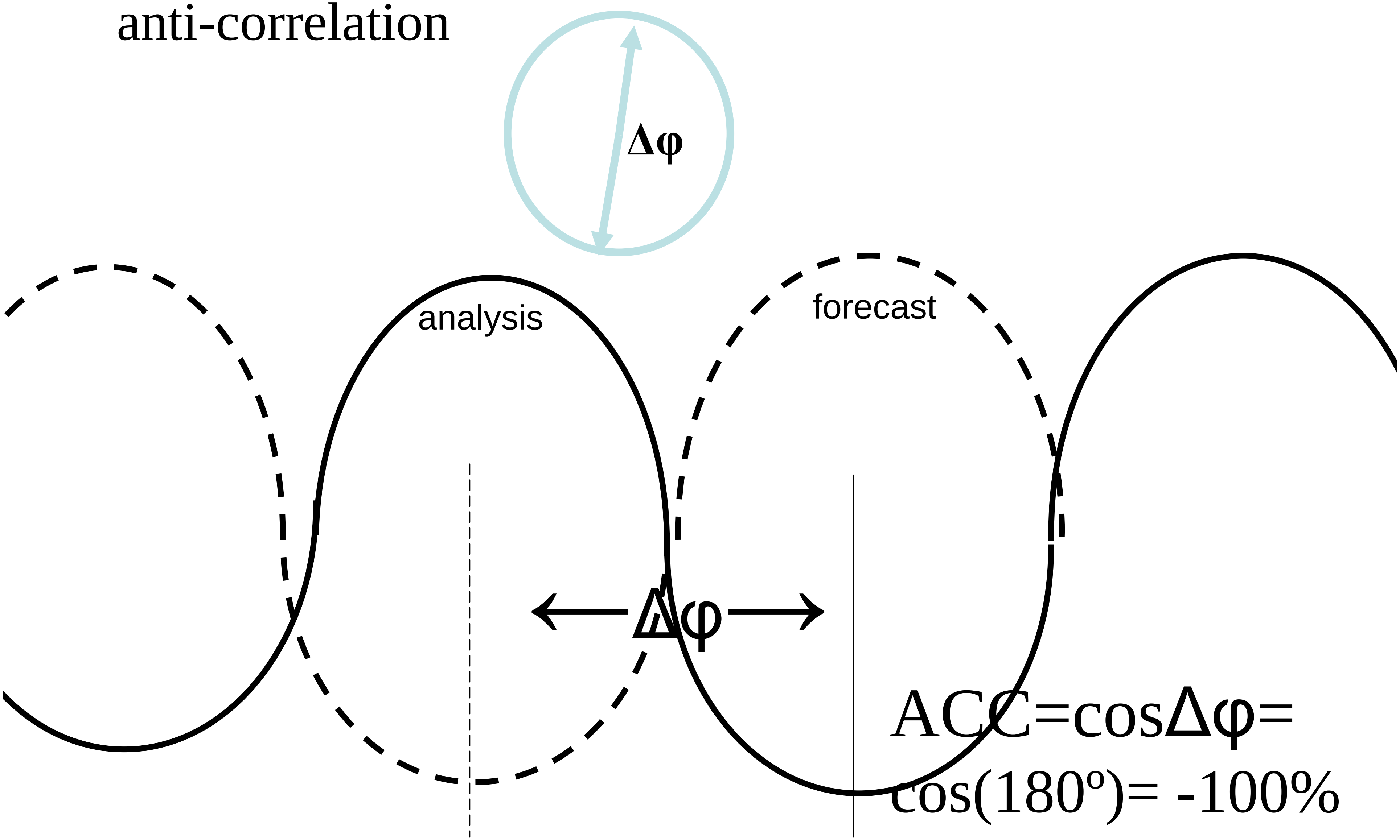
$$ACC = \cos\Delta\varphi = \cos(<90^{o}) > 0\%$$

If the phase error > ½ wave length it is better not to have the feature



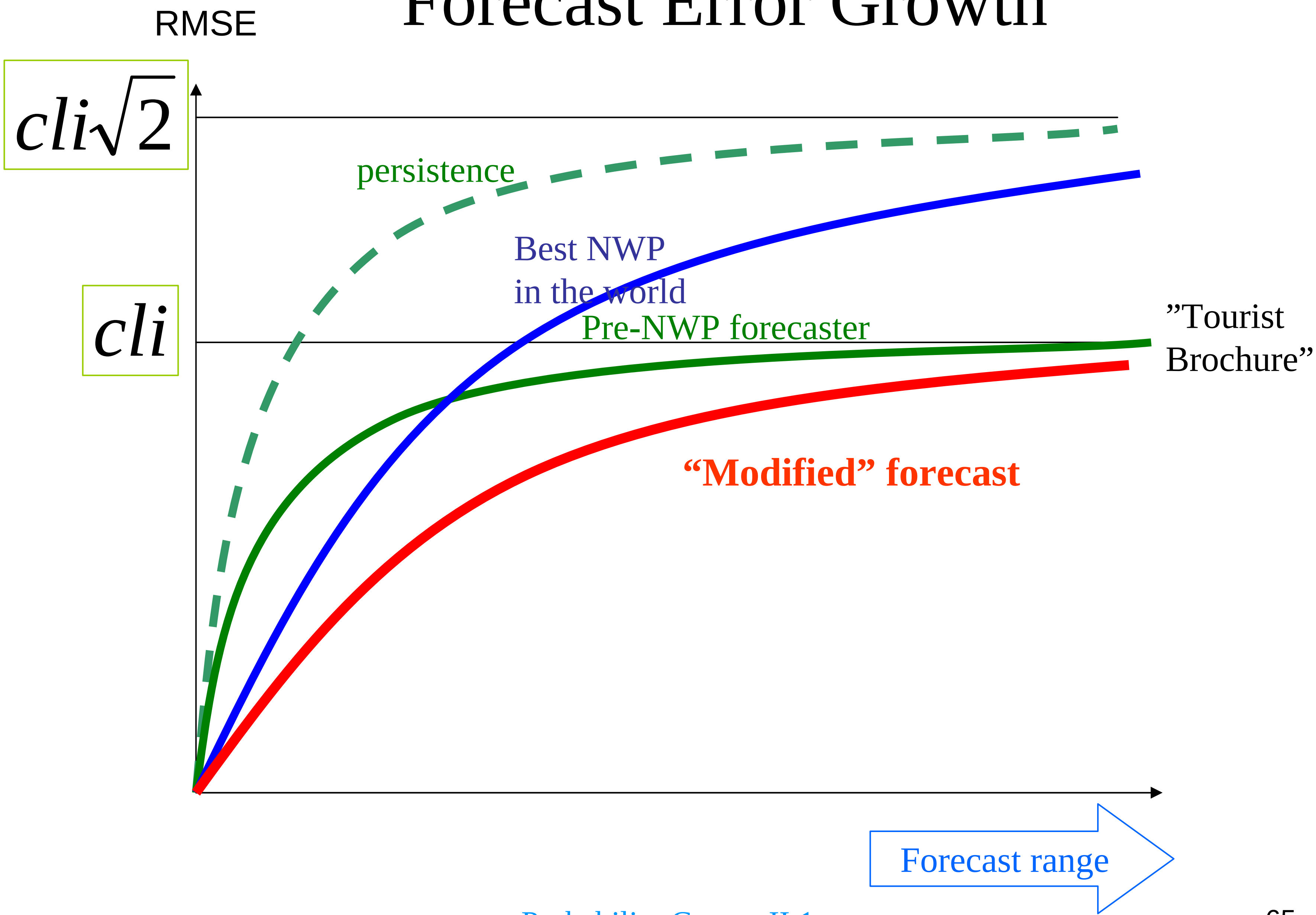
ACC=cosΔφ= cos90º=0%

If the phase error is one wave length there is complete anti-correlation



$$ACC = \cos\Delta\varphi = \cos(180^o) = -100\%$$

Forecast Error Growth



END