Environment and Climate Change Canada | Environnement et Changement climatique Canada

Verification of Continuous Predictands

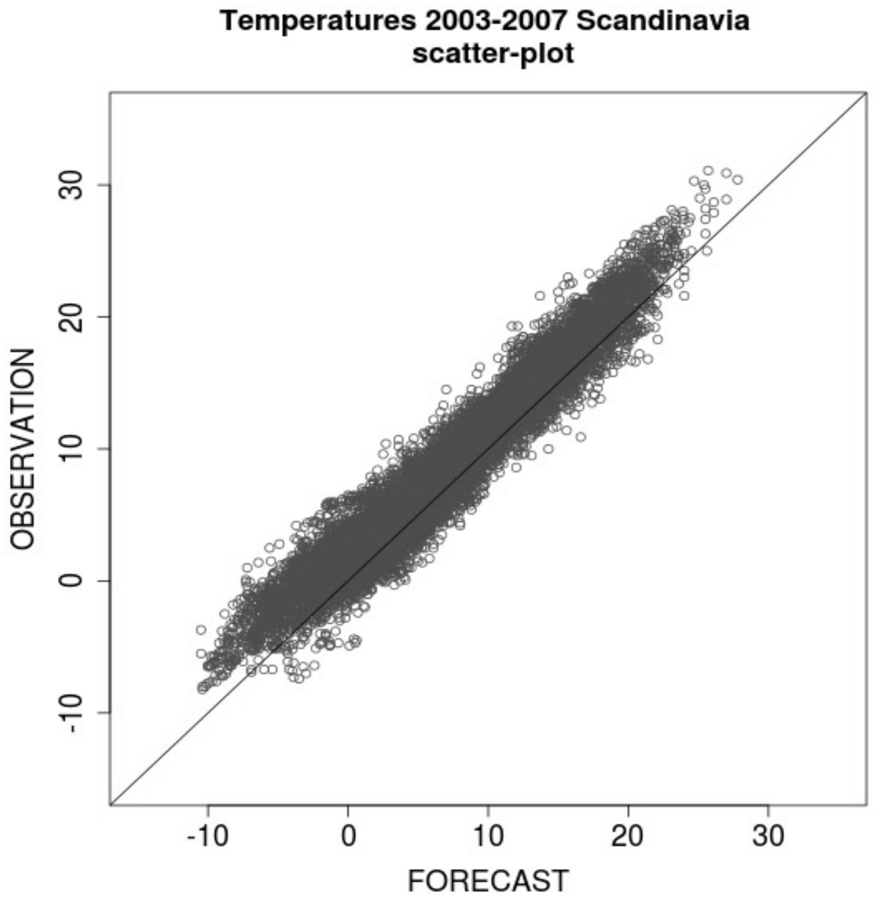
Barbara Casati, Meteorological Research Division, ECCC
MPE-CDT summer school on “Forecast Verification”
Online, 21-25 June 2021

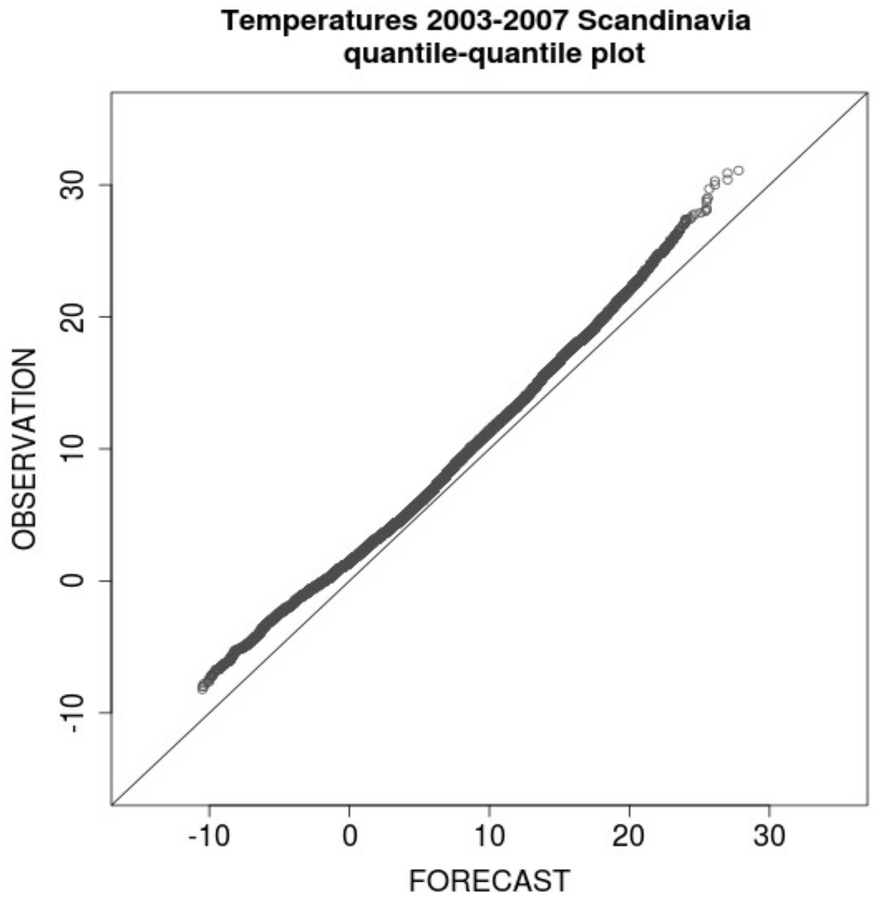
Outline:

- Exploratory methods
- Continuous scores
- Aggregation of continuous statistics

Canada

Graphical Exploratory Analysis





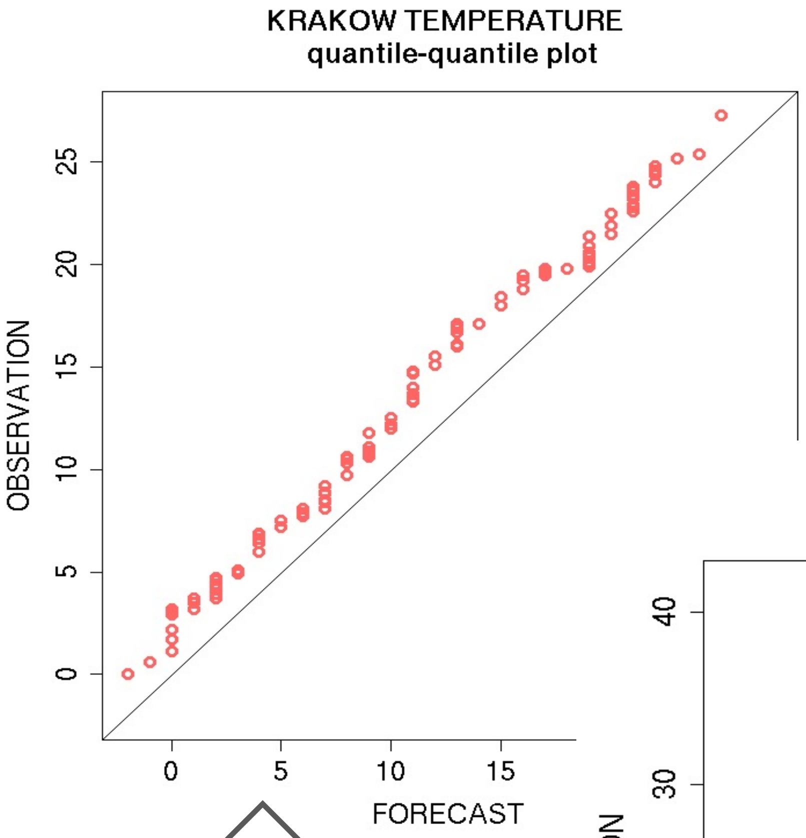
- Provides a glimpse at the overall forecast-obs characteristics and relationship:
- Helps identifying peculiar behaviours and outliers
- Informs on biases, error magnitude, linear association
- Helps establishing the most suitable verification approach!

Scatter-plot: plot observation versus forecast paired values. scatter-plots are a representation of the joint distribution, they assess the accuracy

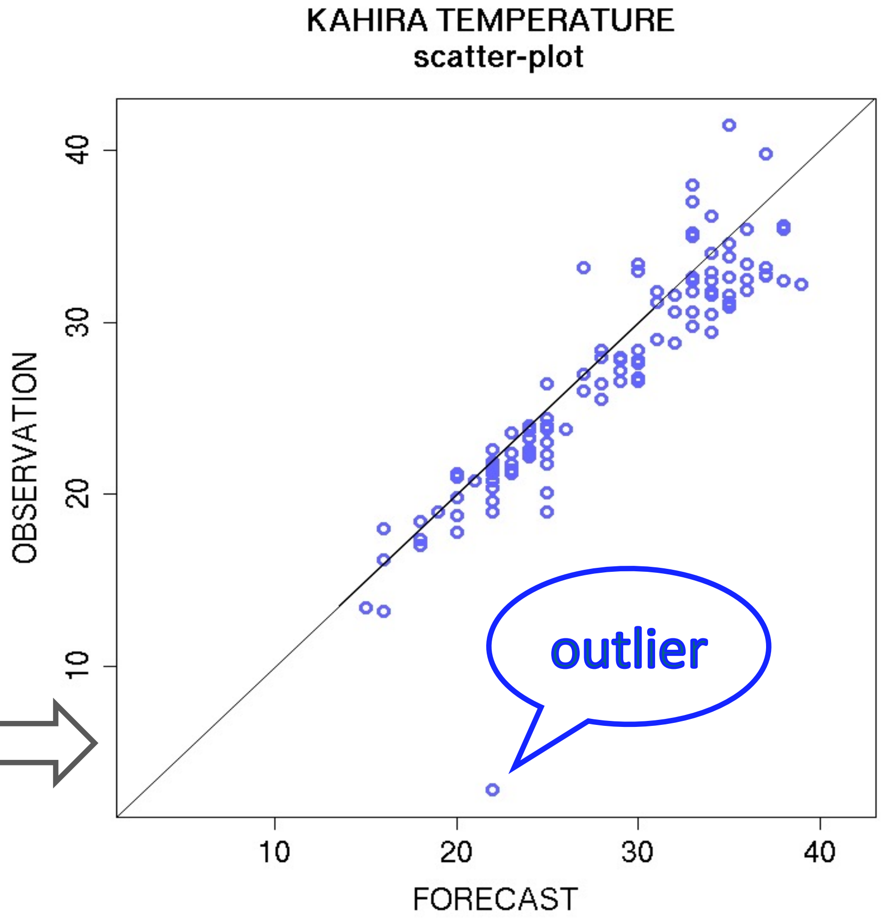
QQ-plots plot the obs versus forecast quantiles (ordered values): compare the forecast vs obs marginal distributions, regardless of the forecast-obs pairing: qq-plots assess the marginal bias

Perfect forecast = obs, points are on the 45° diagonal

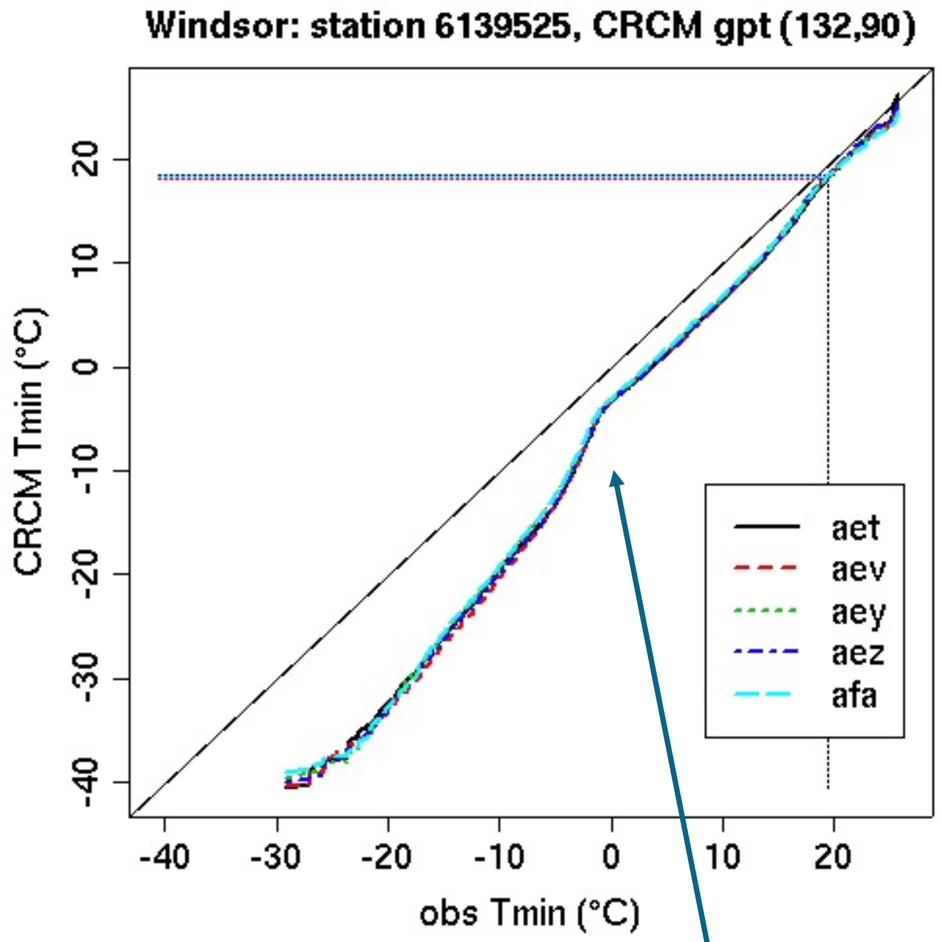
Examples



constant positive bias



error growing with temperature



The qq-plot mix behaviour possibly indicate a mis-representation of a process below freezing point

Comparing the marginal distributions

Graphical tools: qq-plots, histograms, box-plots, …

Summary statistics, describing the forecast (Y) and observed (X) marginal distributions:

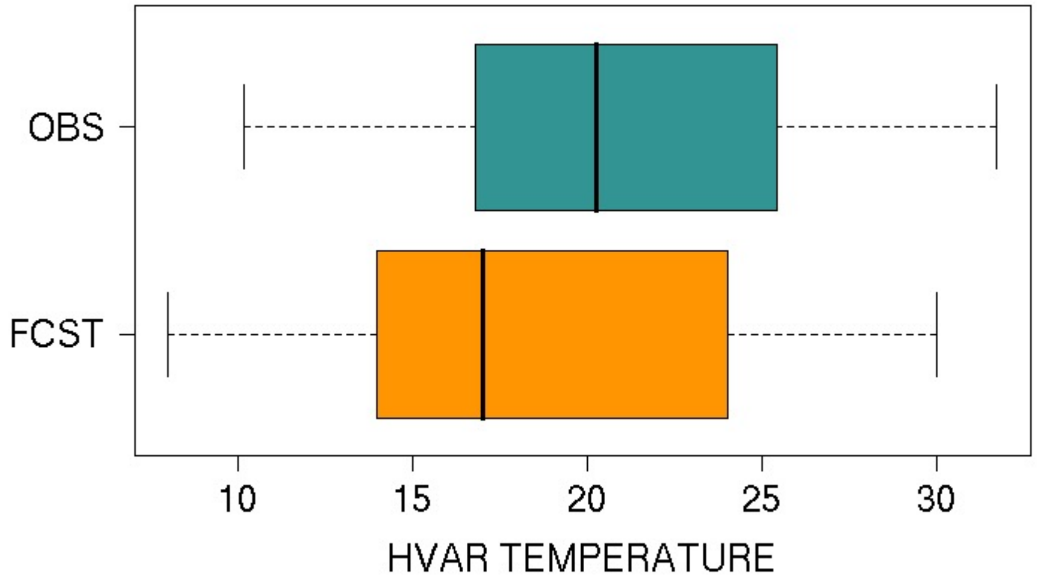
Location: $mean = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$

$median = q_{0.50}$

Spread:

standard deviation $s_X = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$

Inter Quartile Range $IQR = q_{0.75} - q_{0.25}$



	MEAN	$q_{0.50}$	STDEV	IQR
OBS	20.71	20.25	5.18	8.52
FCST	18.62	17.00	5.99	9.72

Most Common Continuous Scores

$$\text{linear bias} = ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) = \bar{Y} - \bar{X}$$

bias

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \qquad MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

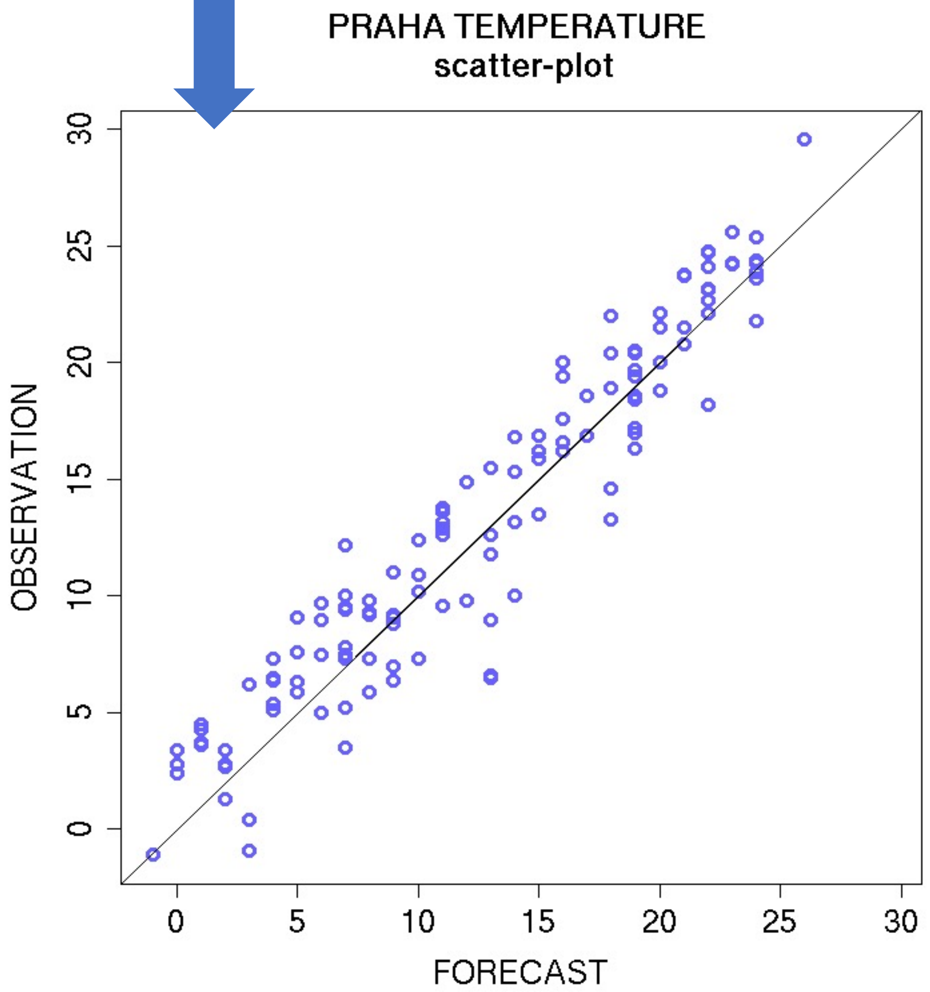
accuracy

$$r_{XY} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{\mathrm{cov}(Y,X)}{s_Y s_X}$$

linear association

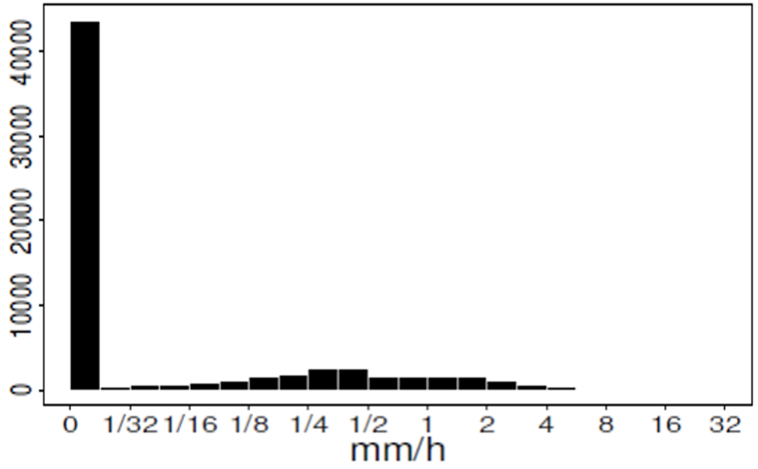
Continuous scores are **summary statistics** to evaluate forecasts of continuous variables (e.g. temperature). These are appealing for e.g. administrative or operational monitoring. However, **a single number** cannot possibly describe in full the whole (often complex) relationship between forecast and observation: **use a few + exploratory methods**

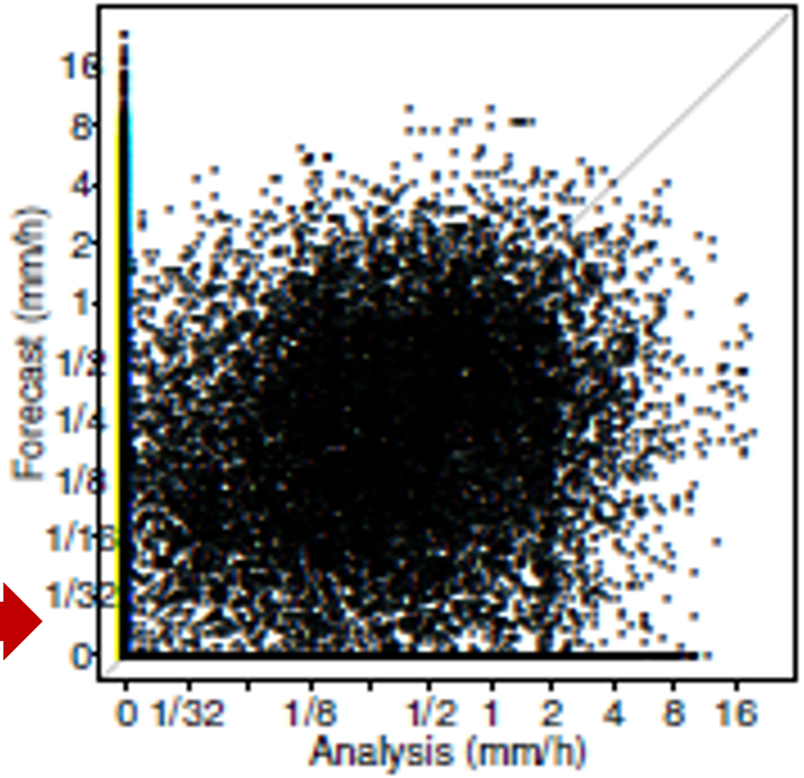
Continuous scores are suitable for continuous Gaussian variables (e.g. temperature)



Continuous scores are not suitable for

- Mixed, bimodal or right-skewed distributions
- Binary or categorical variables (e.g. cloud vs clear sky; sea-ice vs water)
- Episodic spatially discontinuous on-and-off variables (e.g. precipitation)





The additive (linear) bias

ATTRIBUTE: BIAS

$$bias = ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) = \bar{Y} - \bar{X}$$

Additive (linear) bias = Mean Error = average of the errors = difference between the forecast (Y) and observation (X) means (overbar).

- It indicates the average direction of error: positive bias indicates over-forecast, negative bias indicates under-forecast.
- It does not indicate the magnitude of the error (in fact positive and negative error can cancel outs)

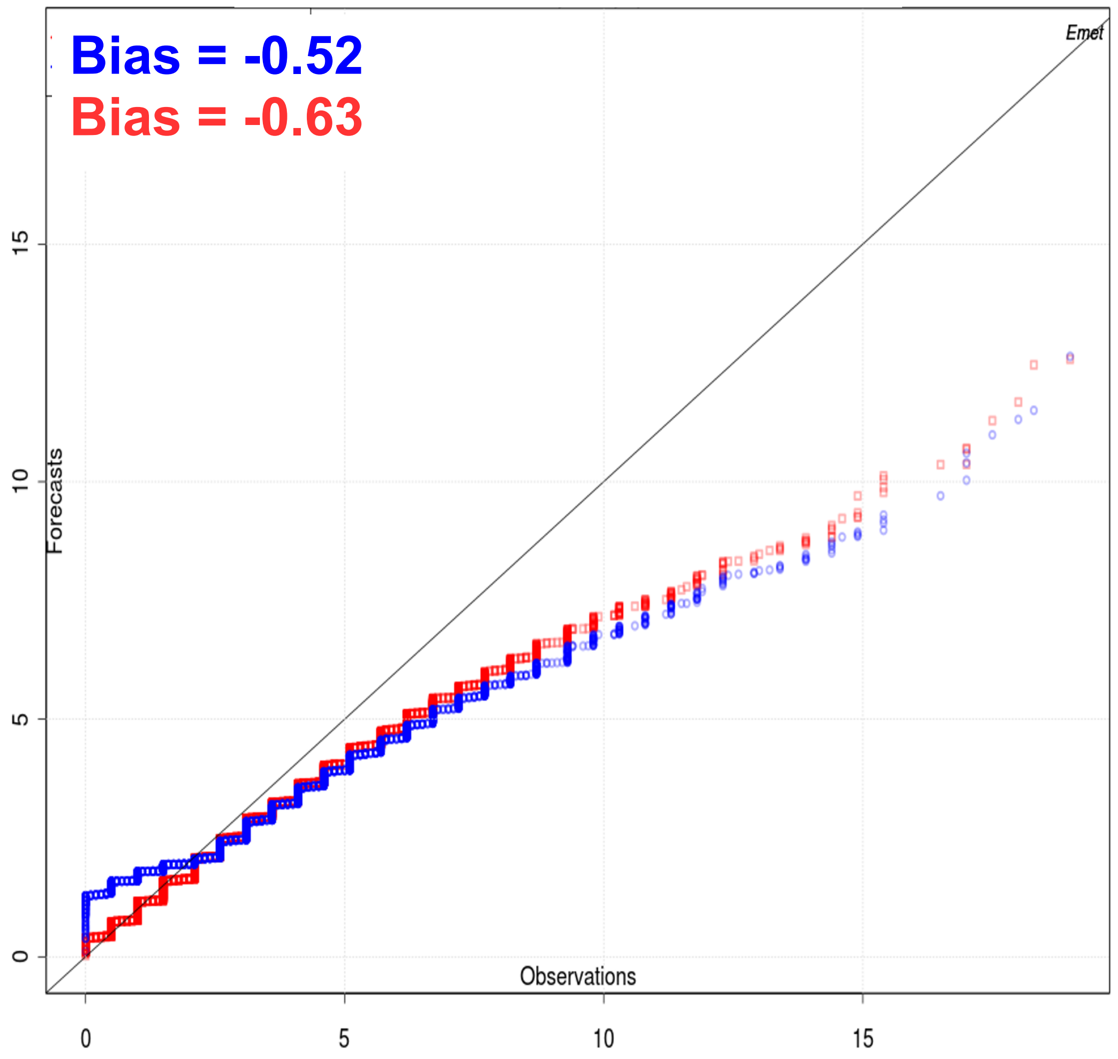
The bias: so simple, so misleading …

The bias (ME) is often affected by compensating errors.

Example: the red curve overall better represents the wind speeds (is nearer to the diagonal), however for the blue curve positive and negative errors cancel out, leading to an overall better bias. The qq-plot is more informative!

Diagnose: if MAE >> ME, then positive and negative biases have cancelled out!

Wind Speed qq-plot



The MAE and MSE

ATTRIBUTE: ACCURACY

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

MAE = average of the magnitude of the errors. Linear score = each error has same weight

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2$$

MSE = average of the squared errors: it measures the magnitude of the error weighted on the squares of the errors

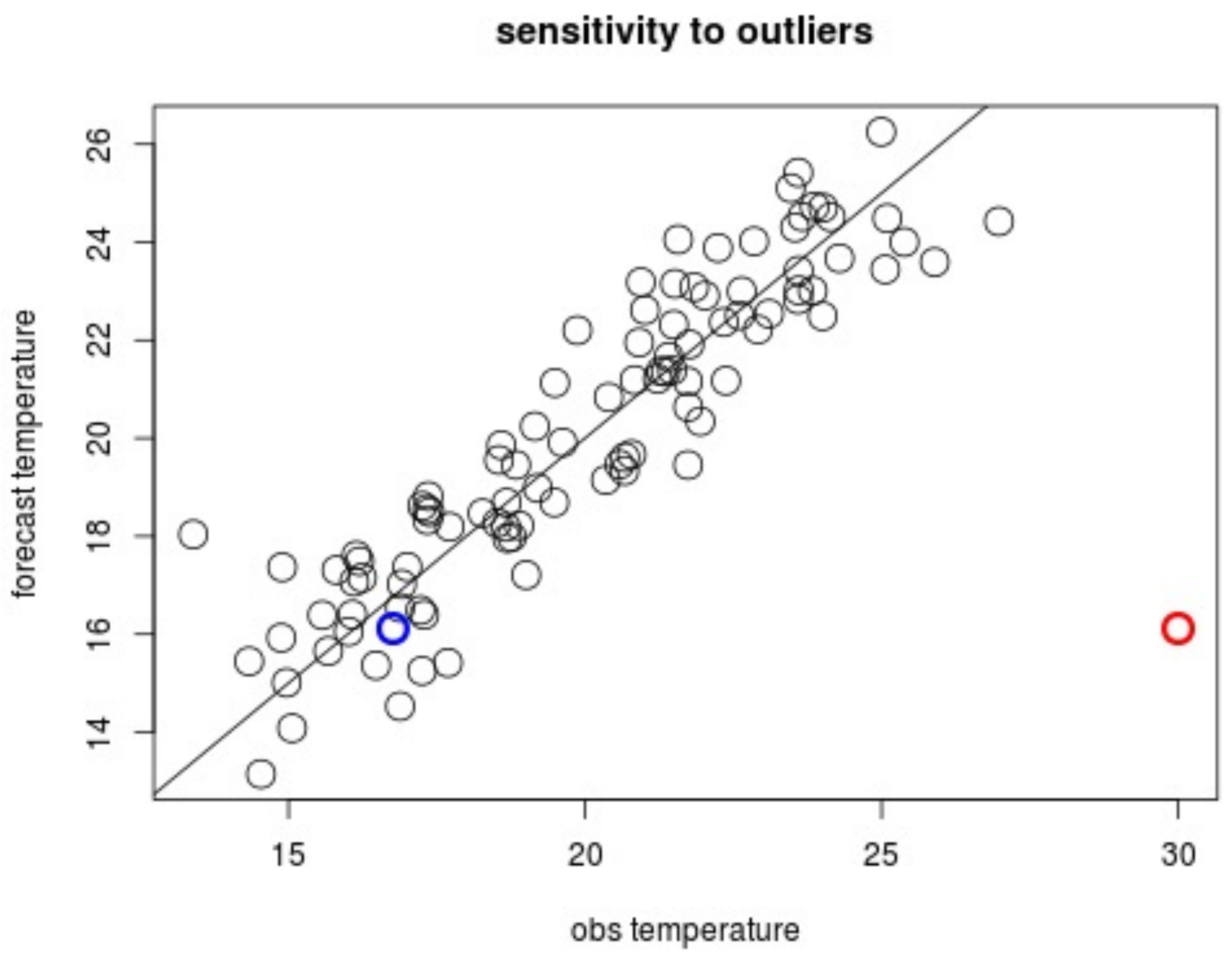
MSE is defined by a quadratic rule, therefore large weight on large errors:

- good if you wish to penalize large error
- sensitive to large values and outliers; sensitive to large variance (high resolution models); encourage conservative forecasts (e.g. climatology)

Diagnose: RMSE >> MAE indicate the presence of outliers, right skewed distribution

Example

	No outlier	Outlier = 30	Outlier = 50
Mean obs	19.98	20.11	20.31
Median obs	20.39	20.58	20.58
stdev obs	3.13	3.27	4.31
IQR obs	5.05	5.04	5.04
RMSE	1.26	1.87	3.59
MAE	1.01	1.14	1.34
correlation	0.92	0.83	0.56



Correlation, MSE, stdev:

Not resistant: sensitive to large values & outliers

Not robust: not suitable for right skewed and/or not Gaussian distributed values

Pearson linear correlation coefficient

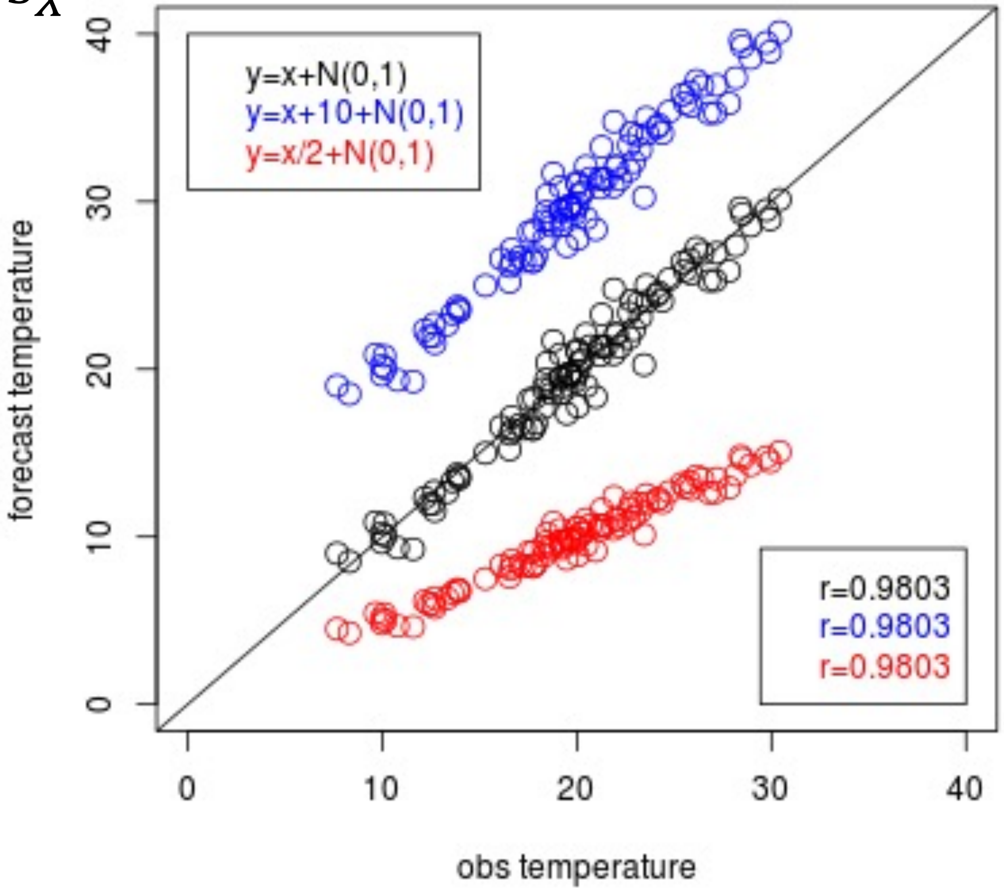
ATTRIBUTE: LINEAR ASSOCIATION

$$r_{YX} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{cov(Y,X)}{s_Y s_X}$$

Is the rescaled (non-dimentional) covariance, ranges in [-1,1]

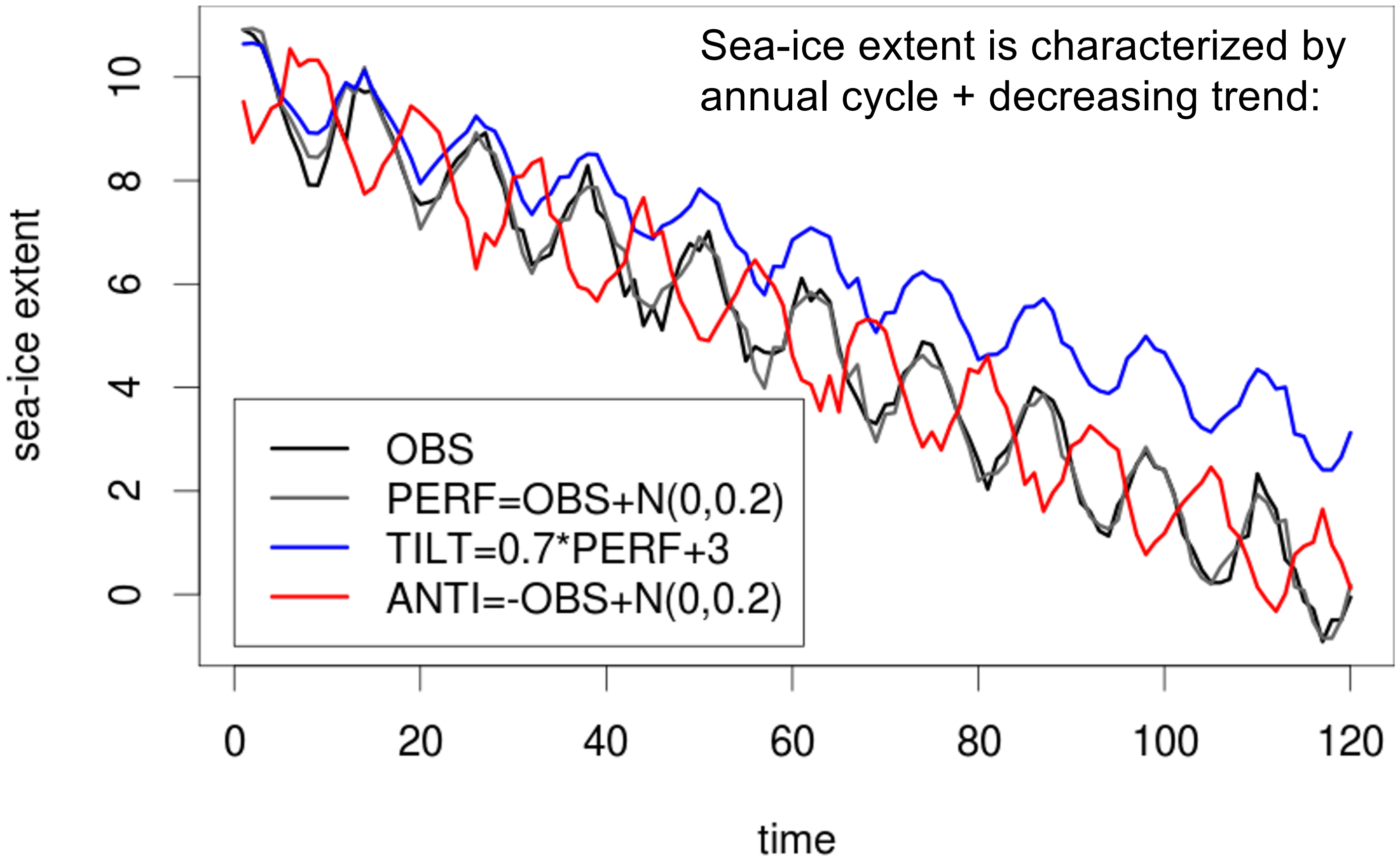
- is not sensitive to the bias
- is not sensitive to the variances
- is not the inclination of the regression line which is $b = (s_X/s_Y)r_{XY}$

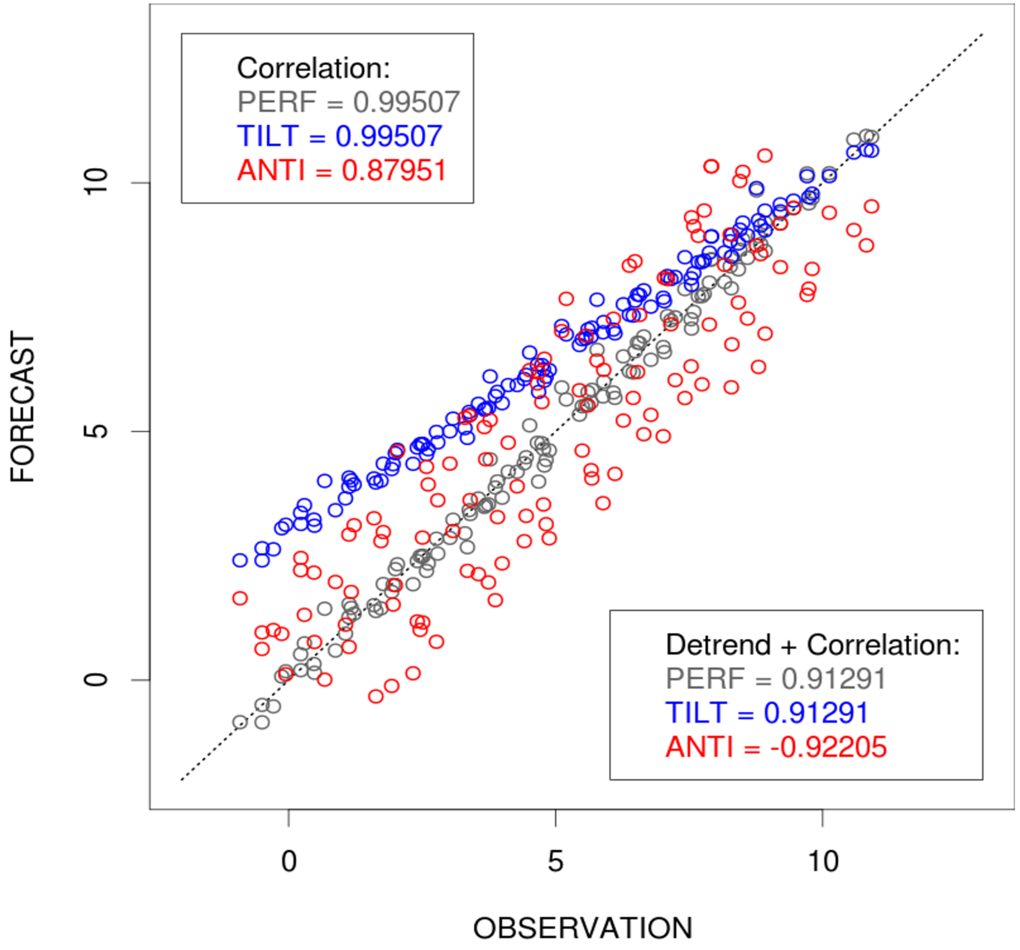
Measures the linear association between forecast and observation



Correlation and False Skill

Correlation is artificially inflated for variables with a (climate) trend!





Similarly, map correlation is artificially inflated if the domain includes different climatologies; **Hamill and Juras (2006)**, show that part of the skill is due to reproducing the local climatologies!

Anomaly Correlation Coefficient

$\forall\, m \in MAP \qquad y'_m = y_m - c_m$

$x'_m = y_m - c_m$

Forecast and observation anomalies: to evaluate forecast quality not accounting for correct forecast of climatology (e.g. driven by topography)

$$ACC_{unc} = \frac{\sum_{m \in MAP} (y'_m)(x'_m)}{\sqrt{\sum_{m \in MAP} (y'_m)^2}\sqrt{\sum_{m \in MAP} (x'_m)^2}}$$

$$ACC_{cent} = \frac{\sum_{m \in MAP} \left(y'_m - \overline{y'}\right)\left(x'_m - \overline{x'}\right)}{\sqrt{\sum_{m \in MAP} \left(y'_m - \overline{y'}\right)^2}\sqrt{\sum_{m \in MAP} \left(x'_m - \overline{x'}\right)^2}}$$

Centred and uncentred ACC for weather variables defined over a spatial domain, where c_m is the climatology at the grid-point m, and the over-bar denotes averaging over the field

Relationship between continuous statistics (MSE, bias, variances, correlation, error stdev)

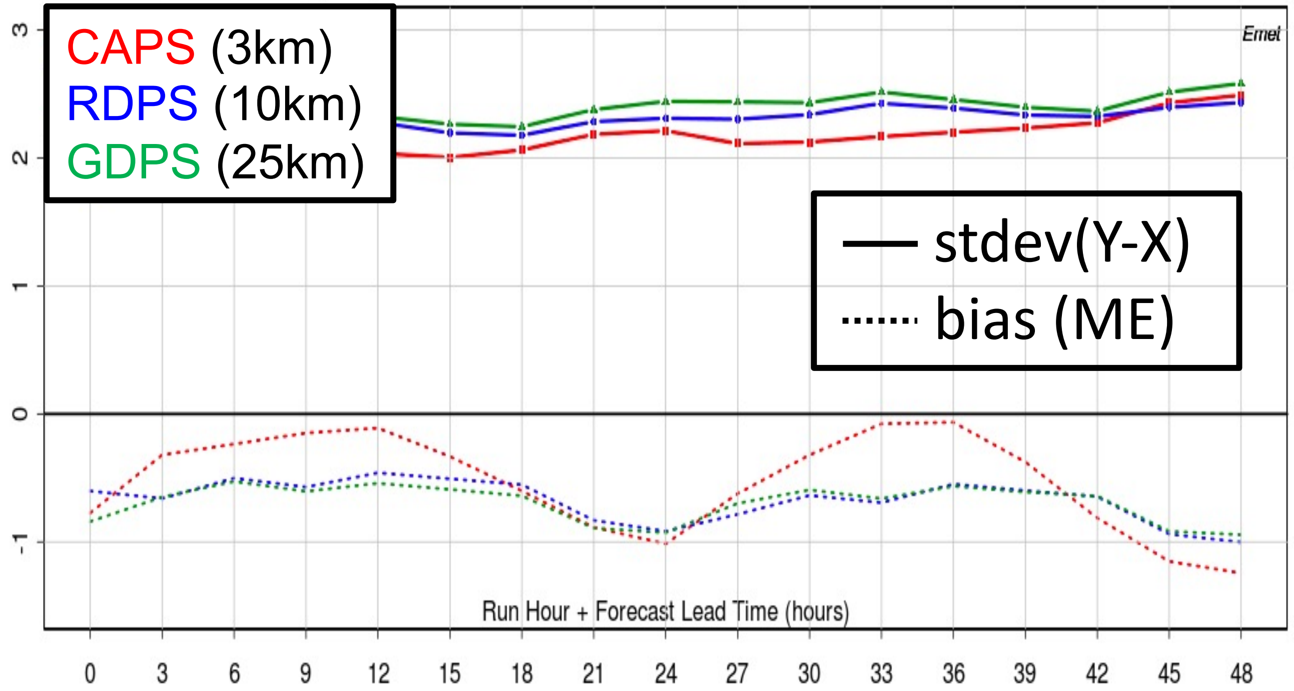
Continuous scores are related:

$$MSE = (\bar{Y} - \bar{X})^2 + s_Y^2 + s_X^2 - 2 s_Y s_X r_{YX}$$

$$MSE = \boxed{ME^2} + \boxed{var(Y - X)}$$

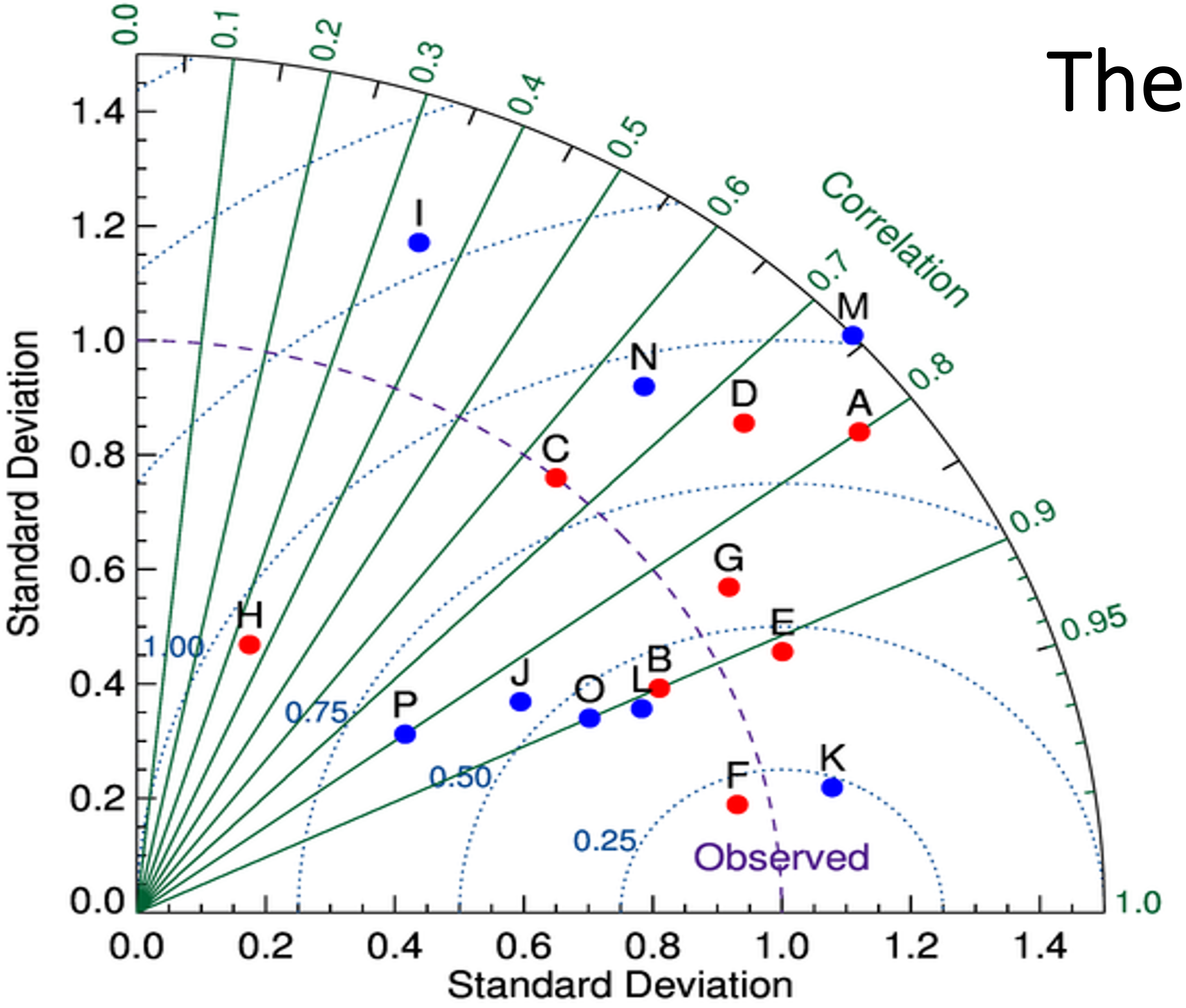
Bias + error standard deviation are often used to summarize performance

2m temperature, JAS 2018, North America >60N



High resolution forecasts exhibit larger variance (s_Y^2), which leads to larger MSE; smoother forecasts (e.g. ensemble mean) exhibit smaller variance, hence smaller MSE ... this is expected / synthetic

The Taylor diagram



the mathematical relation between continuous statistics can be exploited in order to display multiple scores in a single summary diagram (Taylor, 2001, JGR)

MSE Skill scores

A skill score compares the accuracy of the forecast versus that of a reference (often trivial) forecast (e.g. persistence, climatology, random).

Example: **Reduction of Variance**
skill score associated to the MSE with reference = sample climatology = $\bar{X}$

$$SS = \frac{MSE - MSE_{clim}}{MSE_{best} - MSE_{clim}} = 1 - \frac{MSE}{\overline{\overline{(\bar{X} - X)^2}}} = 1 - \frac{MSE}{s_X^2}$$

The reduction of Variance in Hydrology is known as the **Nash-Sutcliffe Efficiency** (Nash and Sutcliffe, 1970)

Nash-Sutcliffe and Kling-Gupta Efficiency

$$NSE = 1 - \frac{MSE}{s_X^2} = r^2 - \left(r - \frac{s_Y}{s_X}\right)^2 - \boxed{\frac{(\bar{Y} - \bar{X})^2}{s_X^2}}$$

Gupta et al (2009) J.Hydrology:

- NSE assess the skill in three components: correlation, ratio of variances, additive bias
- The NSE is maximized for $r = s_Y/s_X$, since r ≤1 it encourages smoother forecasts

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{s_Y}{s_X} - 1\right)^2 + \left(\frac{\bar{Y}}{\bar{X}} - 1\right)^2}$$

Three distinct components; bias as a ratio; Euclidian distance from the perfect forecast

Continuous Scores of Ranks

Continuous scores sensitive to large values or non robust (e.g. MSE or correlation coefficient) are some-times evaluated by using the ranks of the variable, rather than its actual values

Temp ºC	27.4	21.7	24.2	23.1	19.8	25.5	24.6	22.3	x_i
ordered	19.8	21.7	22.3	23.1	24.2	24.6	25.5	27.4	$\hat{x}_i$
ranks	1	2	3	4	5	6	7	8	i

The value-to-rank transformation:

- diminish effects due to large values
- remove bias
- transform marginal distribution to a Uniform distribution
- **is equivalent to score in probability space**

How do we calculate the cumulative probabilities from the ranks?

$$\cancel{p_i = \frac{i}{n}} \qquad \cancel{p_i = \frac{i-1}{n}}$$

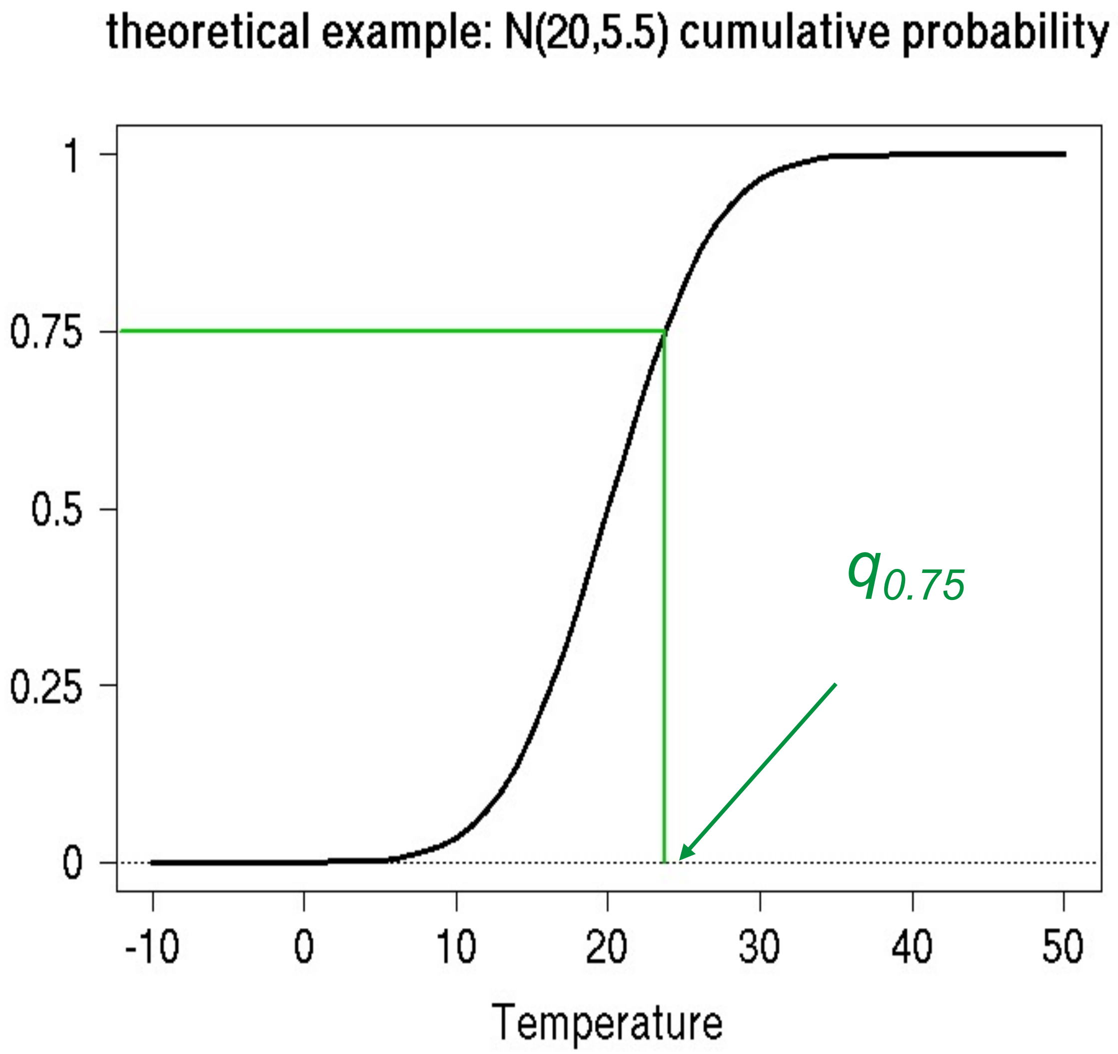
$$p_i = \boxed{\frac{i-1}{n-1}} \qquad p_i = \boxed{\frac{i}{n+1}}$$

Linear Error in Probability Space

$$LEPS = \frac{1}{n}\sum_{i=1}^{n}\left|F_X\left(y_i\right) - F_X\left(x_i\right)\right|$$

The LEPS is a MAE evaluated by using the cumulative frequencies of the observation

Errors in the tail of the distribution are penalized less than errors in the centre of the distribution



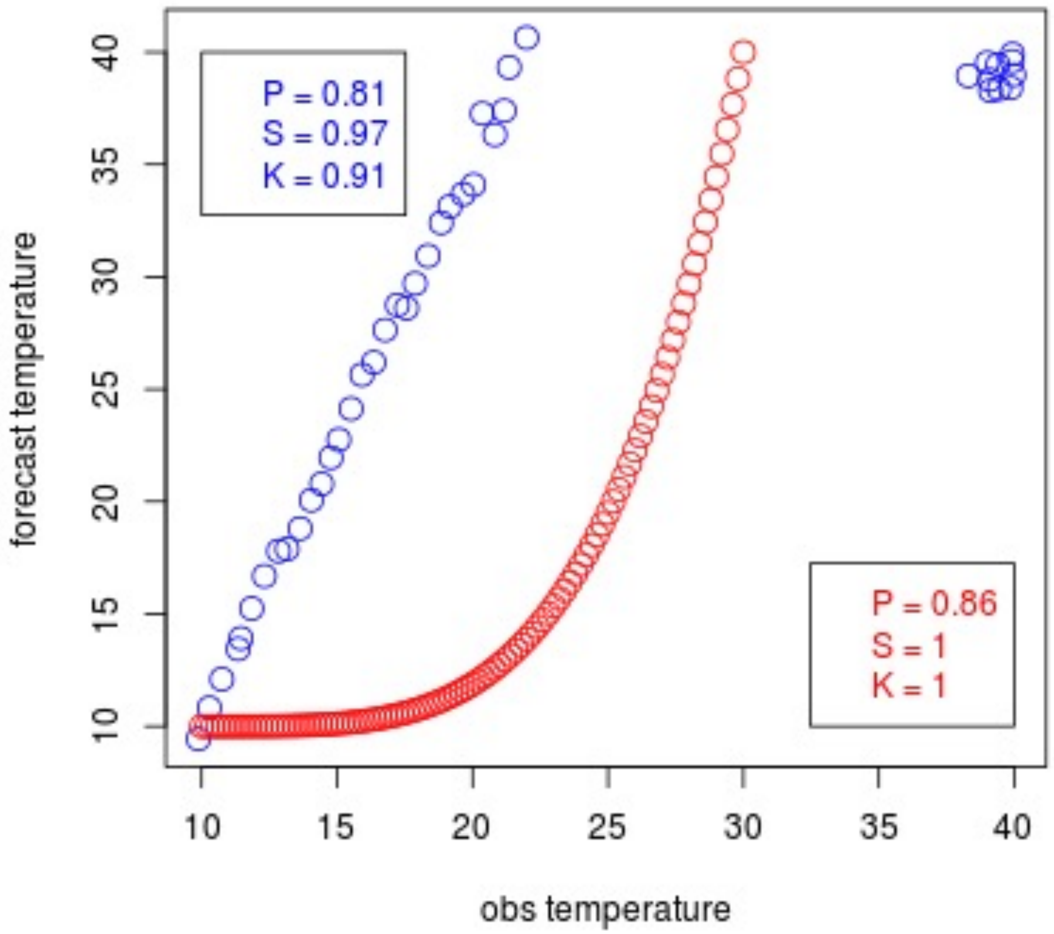
Spearman and Kendall rank correlation

The **Spearman rank correlation** is the Pearson **correlation between the ranks of X and Y**. It assess how well the relationship between two variables can be described using a **monotonic function**

Similarly, the **Kendall rank correlation**

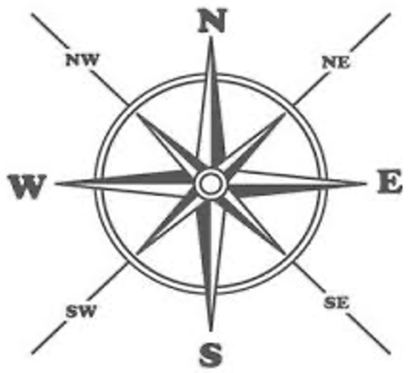
$$\tau = \frac{n_{concordant} - n_{discordant}}{n(n-1)/2}$$

measures the **ordinal association** between two variables.



Spearman and Kendall correlations are **more resistant and robust** than Pearson correlation

Circular statistics

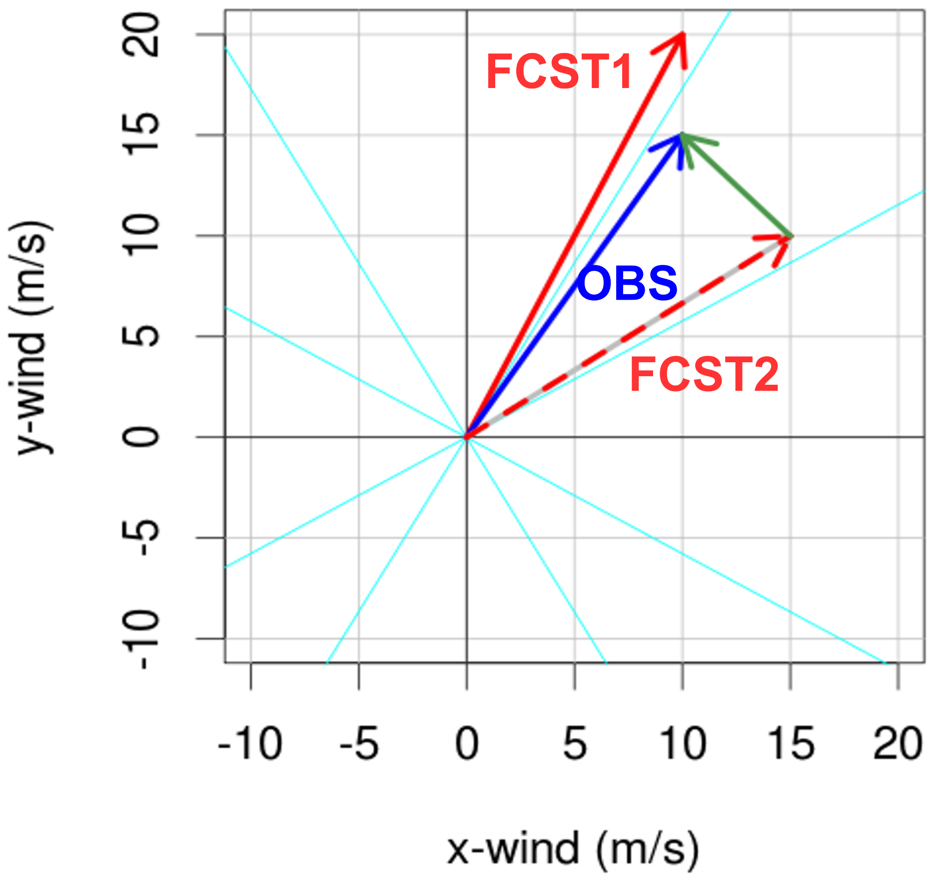


Wind is a circular / vectorial variables.
How do we verify wind?

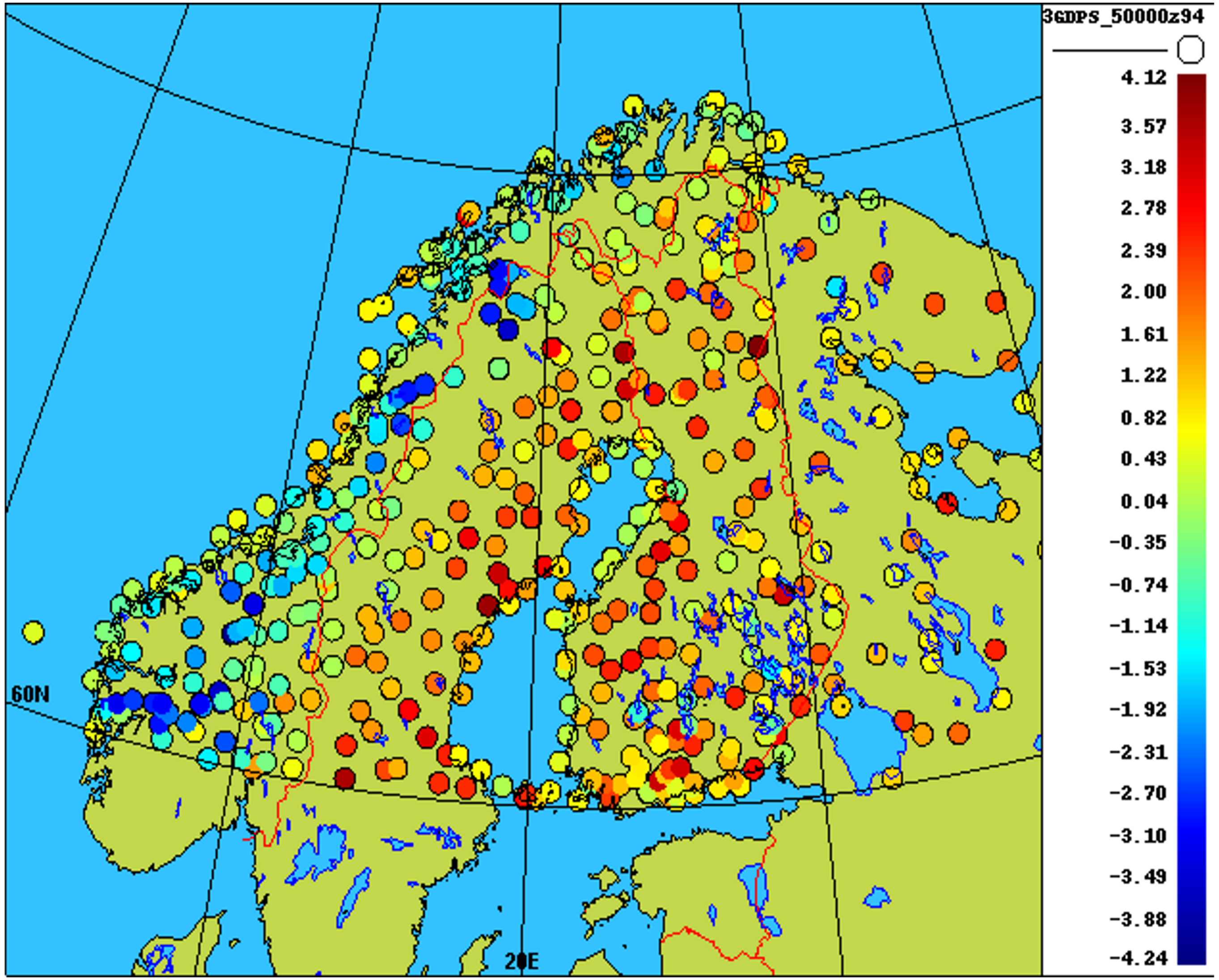
1. Score **x (u) and y (v) components**

2. Decompose into modulo+angle:
   - **wind speed** bias and stdev
   - **wind direction**: multi-categorical scores (30^o sectors, winds speeds > 3m/s)

3. **Vectorial difference:** vector-RMSE is the modulo of vectorial difference, measures the overall accuracy
   - weak winds => small errors
   - strong winds => large errors



Aggregation versus Stratification

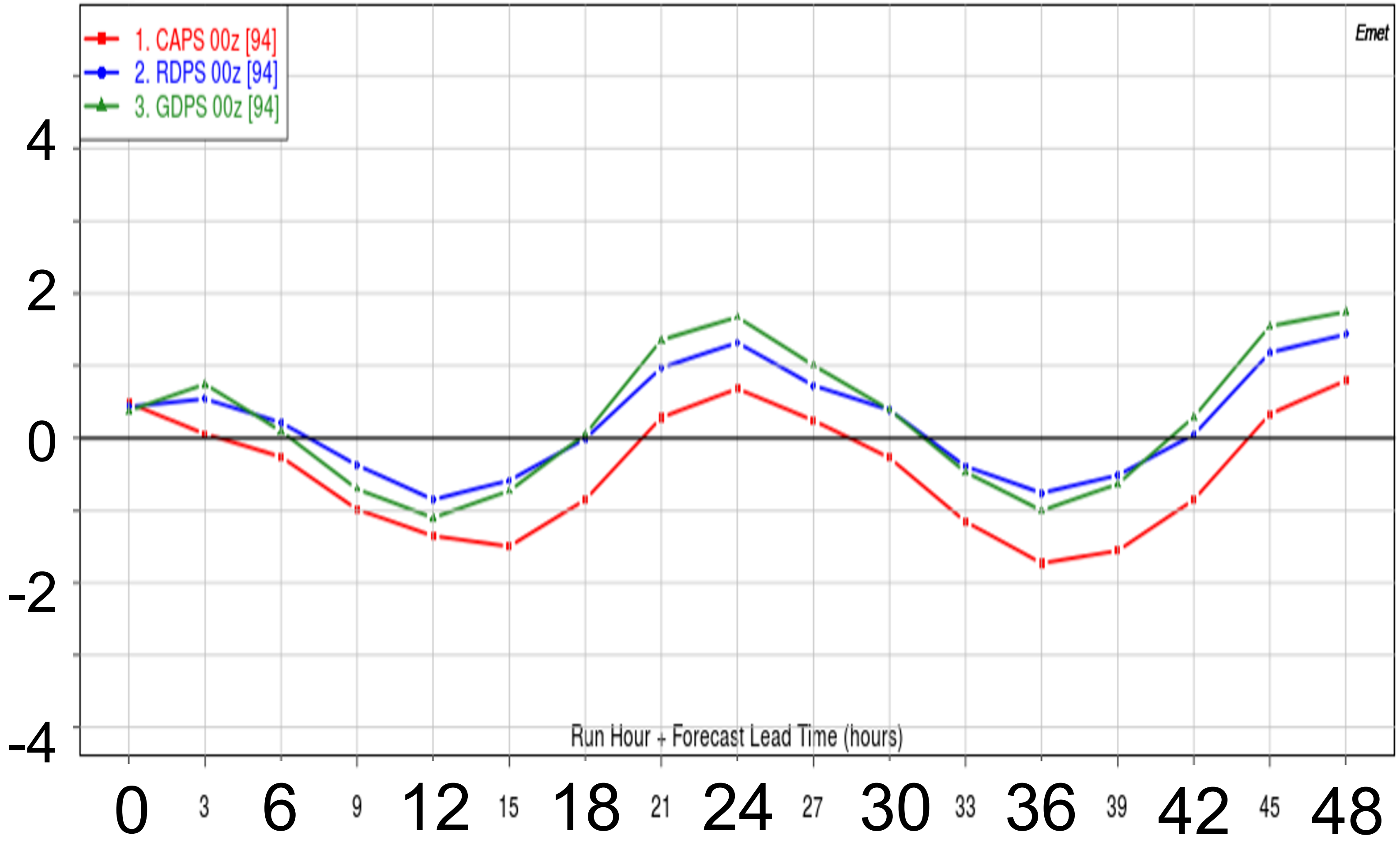


Aggregation leads to more **robust verification results**, however we need to **stratify** our dataset into **homogeneous sub-samples**

Score maps help characterizing regional behaviours and define appropriate sub-domains for aggregating verification scores

Aggregation versus Stratification

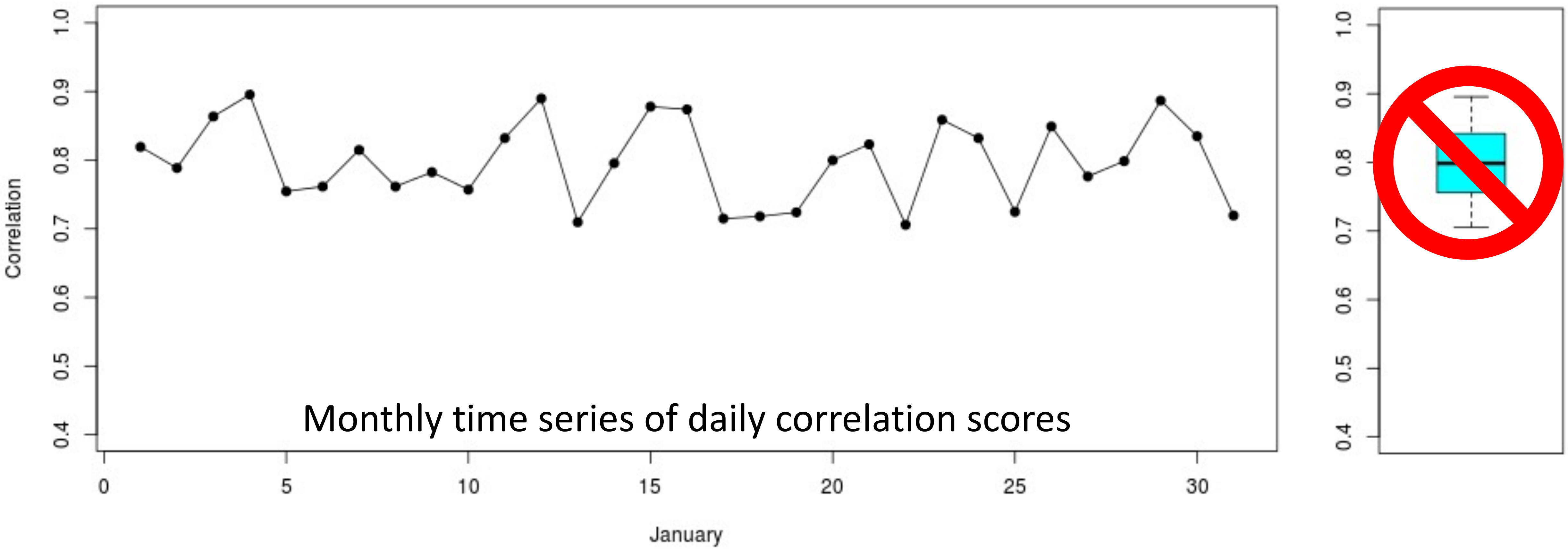
Surface Temperature Bias Sodankyla, JAS 2018



Aggregation leads to more **robust verification results**, however we need to **stratify** our dataset into **homogeneous sub-samples**

Separate aggregation for **origin + leadtime** reveals **diurnal cycle**

Aggregation is not averaging



Q: What is the aggregated score?

A: The score obtained by pulling together all data.

Aggregation on continuous statistics: forecast (y) and obs (x) averages and bias

sample size

day 1	nn_{d1}	$\overline{x_{d1}}$	$\overline{y_{d1}}$
day 2	nn_{d2}	$\overline{x_{d2}}$	$\overline{y_{d2}}$
day 3	nn_{d3}	$\overline{x_{d3}}$	$\overline{y_{d3}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
day j	nn_{dj}	$\overline{x_{dj}}$	$\overline{y_{dj}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
day n	nn_{dn}	$\overline{x_{dn}}$	$\overline{y_{dn}}$
	$\boldsymbol{nn_{aggr}}$	$\boldsymbol{\overline{x_{aggr}}}$	$\boldsymbol{\overline{y_{aggr}}}$

$$\overline{x_{d1}} = \frac{1}{nn_{d1}} \sum_{i}^{I} x_{d1,i} \qquad \overline{x_{d2}} = \frac{1}{nn_{d2}} \sum_{k}^{K} x_{d2,k}$$

$$\overline{x_{d1\cup d2}} = \frac{1}{nn_{d1} + nn_{d2}} \left[\sum_{i}^{I} x_{d1,i} + \sum_{k}^{K} x_{d2,k} \right]$$

$$\overline{x_{d1\cup d2}} = \frac{nn_{d1}}{nn_{d1} + nn_{d2}} \overline{x_{d1}} + \frac{nn_{d2}}{nn_{d1} + nn_{d2}} \overline{x_{d2}}$$

$$\overline{x_{d1\cup d2}} = w_{d1}\overline{x_{d1}} + w_{d2}\overline{x_{d2}}$$

$$\overline{x_{aggr}} = \sum w_{dj}\overline{x_{dj}}$$

$$nn_{aggr} = \sum nn_{dj}$$

$$w_{dj} = nn_{dj}/nn_{aggr}$$

Aggregation on continuous statistics: forecast and obs variances (s_Y^2 , s_X^2)

sample size

day 1	nn_{d1}	$\overline{x_{d1}}$	$\overline{y_{d1}}$	$\overline{x_{d1}^2}$	$\overline{y_{d1}^2}$
day 2	nn_{d2}	$\overline{x_{d2}}$	$\overline{y_{d2}}$	$\overline{x_{d2}^2}$	$\overline{y_{d2}^2}$
day 3	nn_{d3}	$\overline{x_{d3}}$	$\overline{y_{d3}}$	$\overline{x_{d3}^2}$	$\overline{y_{d3}^2}$
⋮	⋮	⋮	⋮	⋮	⋮
day j	nn_{dj}	$\overline{x_{dj}}$	$\overline{y_{dj}}$	$\overline{x_{dj}^2}$	$\overline{y_{dj}^2}$
⋮	⋮	⋮	⋮	⋮	⋮
day n	nn_{dn}	$\overline{x_{dn}}$	$\overline{y_{dn}}$	$\overline{x_{dn}^2}$	$\overline{y_{dn}^2}$
	$\boldsymbol{nn_{aggr}}$	$\overline{\boldsymbol{x_{aggr}}}$	$\overline{\boldsymbol{y_{aggr}}}$	$\overline{\boldsymbol{x_{aggr}^2}}$	$\overline{\boldsymbol{y_{aggr}^2}}$

$$\mathrm{var}\ (\mathrm{X}) = s_X^2 = \frac{1}{nn}\sum_i^I (x_i - \bar{x})^2 =$$

$$= \frac{1}{nn}\sum_i^I \left(x_i^2 - 2\bar{x}x_i + \bar{x}\right) = \overline{x^2} - \bar{x}^2$$

$$\overline{x_{aggr}} = \sum w_{dj}\overline{x_{dj}}$$

$$\overline{x_{aggr}^2} = \sum w_{dj}\overline{x_{dj}^2}$$

$$s_{Xaggr}^2 = \overline{x_{aggr}^2} - \overline{x_{aggr}}^2$$

Aggregation on continuous statistics: forecast and obs covariance and correlation

sample size

day 1	nn_{d1}	$\overline{x_{d1}}$	$\overline{y_{d1}}$	$\overline{x^2_{d1}}$	$\overline{y^2_{d1}}$	$\overline{xy_{d1}}$
day 2	nn_{d2}	$\overline{x_{d2}}$	$\overline{y_{d2}}$	$\overline{x^2_{d2}}$	$\overline{y^2_{d2}}$	$\overline{xy_{d2}}$
day 3	nn_{d3}	$\overline{x_{d3}}$	$\overline{y_{d3}}$	$\overline{x^2_{d3}}$	$\overline{y^2_{d3}}$	$\overline{xy_{d3}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
day j	nn_{dj}	$\overline{x_{dj}}$	$\overline{y_{dj}}$	$\overline{x^2_{dj}}$	$\overline{y^2_{dj}}$	$\overline{xy_{dj}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
day n	nn_{dn}	$\overline{x_{dn}}$	$\overline{y_{dn}}$	$\overline{x^2_{dn}}$	$\overline{y^2_{dn}}$	$\overline{xy_{dn}}$
	$\boldsymbol{nn_{aggr}}$	$\overline{\boldsymbol{x_{aggr}}}$	$\overline{\boldsymbol{y_{aggr}}}$	$\overline{\boldsymbol{x^2_{aggr}}}$	$\overline{\boldsymbol{y^2_{aggr}}}$	$\overline{\boldsymbol{xy_{aggr}}}$

$$\mathrm{cov}\,(X,Y) = \frac{1}{nn}\sum_i^I (x_i - \bar{x})(y_i - \bar{y}) =$$

$$= \frac{1}{nn}\sum_i^I (x_i y_i - \bar{x} y_i - \bar{y} x_i + \bar{x}\bar{y}) =$$

$$= \frac{1}{nn}\sum_i^I x_i y_i - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y}$$

$$\overline{xy_{aggr}} = \sum w_{dj}\overline{xy_{dj}}$$

$$\mathrm{cov}\,(X,Y)_{aggr} = \overline{xy_{aggr}} - \overline{x_{aggr}} \cdot \overline{y_{aggr}}$$

$$r_{(X,Y)aggr} = \frac{\mathrm{cov}\,(X,Y)_{aggr}}{s_{Xaggr} s_{Yaggr}}$$

Aggregation on continuous statistics: mean squared and absolute error

sample size

day 1	nn_{d1}	$\overline{x}_{d1}$	$\overline{y}_{d1}$	$\overline{x^2_{d1}}$	$\overline{y^2_{d1}}$	$\overline{xy}_{d1}$	MSE_{d1}	MAE_{d1}
day 2	nn_{d2}	$\overline{x}_{d2}$	$\overline{y}_{d2}$	$\overline{x^2_{d2}}$	$\overline{y^2_{d2}}$	$\overline{xy}_{d2}$	MSE_{d2}	MAE_{d2}
day 3	nn_{d3}	$\overline{x}_{d3}$	$\overline{y}_{d3}$	$\overline{x^2_{d3}}$	$\overline{y^2_{d3}}$	$\overline{xy}_{d3}$	MSE_{d3}	MAE_{d3}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
day j	nn_{dj}	$\overline{x}_{dj}$	$\overline{y}_{dj}$	$\overline{x^2_{dj}}$	$\overline{y^2_{dj}}$	$\overline{xy}_{dj}$	MSE_{dj}	MAE_{dj}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
day n	nn_{dn}	$\overline{x}_{dn}$	$\overline{y}_{dn}$	$\overline{x^2_{dn}}$	$\overline{y^2_{dn}}$	$\overline{xy}_{dn}$	MSE_{dj}	MAE_{dj}
	nn_{aggr}	$\boldsymbol{\overline{x}_{aggr}}$	$\boldsymbol{\overline{y}_{aggr}}$	$\boldsymbol{\overline{x^2_{aggr}}}$	$\boldsymbol{\overline{y^2_{aggr}}}$	$\boldsymbol{\overline{xy}_{aggr}}$	MSE_{aggr}	MAE_{aggr}

The MSE_{aggr} is evaluated from the correlation and the forecast and obs means and variances. However, good to buddy check:

$$MSE_{aggr} = \sum w_{dj} MSE_{dj}$$

$$MAE_{aggr} = \sum w_{dj} MAE_{dj}$$

Aggregated continuous scores can be evaluated from few **atomic statistics**

	$\bar{x}$	$\bar{y}$	s_X	s_Y	$r_{X,Y}$
January	15	20	1	1	0.8
February	17	21	1	1	0.8
March	19	22	1.5	2	0.7
April	21	23	1.5	2	0.7
May	23	24	1.5	2	0.7
June	25	25	2	3	0.6
July	27	26	2	3	0.6
August	25	25	2	3	0.6
September	23	24	1.5	2	0.7
October	21	23	1.5	2	0.7
November	19	22	1.5	2	0.7
December	17	21	1	1	0.8

Assignment

Evaluate the aggregated annual bias, MSE and correlation, from the given monthly statistics

Alternative:
Evaluate the aggregated summer (JJA) or winter (DJF) correlation, from the given monthly statistics

Environment and Climate Change Canada
Environnement et Changement climatique Canada

Mathematics of Planet Earth

JWGFVR

Thank you!

b.casati@gmail.com
barbara.casati@ec.gc.ca

Canada