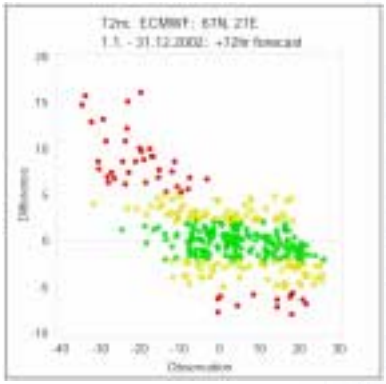


Pertti Nurmi

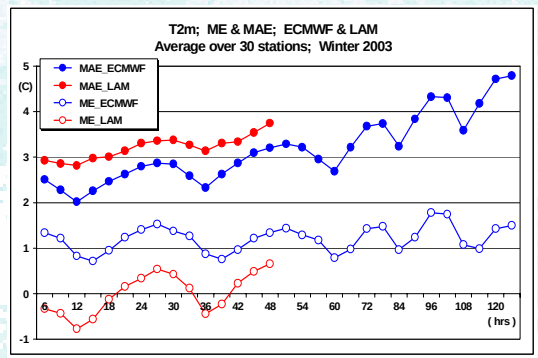
& Sigbritt Näsman

Finnish Meteorological Institute

Experimentation with the LEPS Score:

Comparison of local temperature forecast errors in probability and measurement space

Montreal
15.9.2004



NOT to be mixed with:

Limited-area

Ensemble

Prediction

System (LEPS)

Instead …

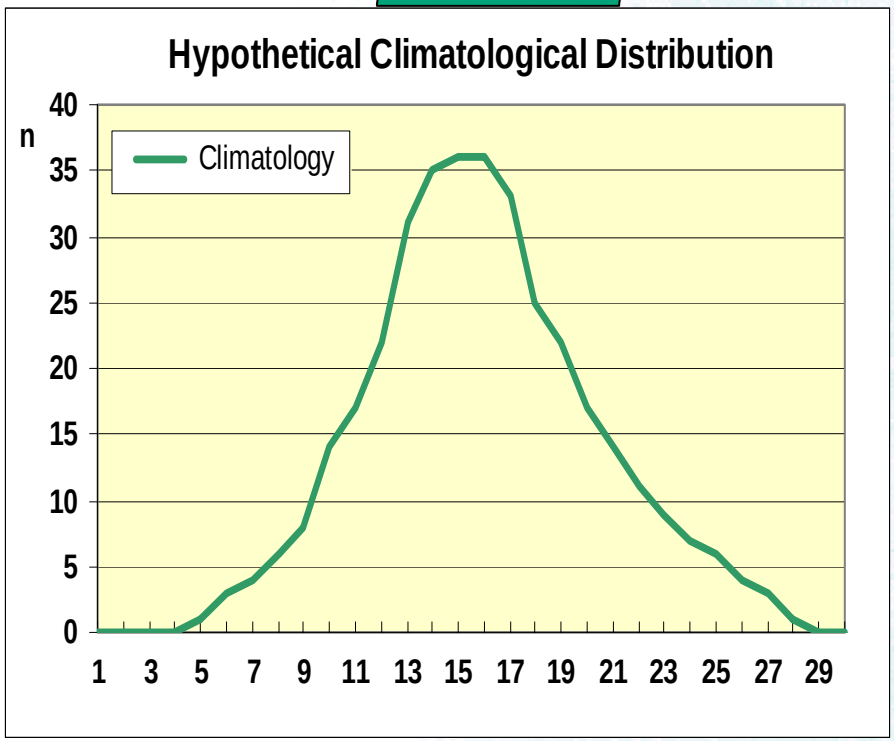
Linear Error in Probability Space

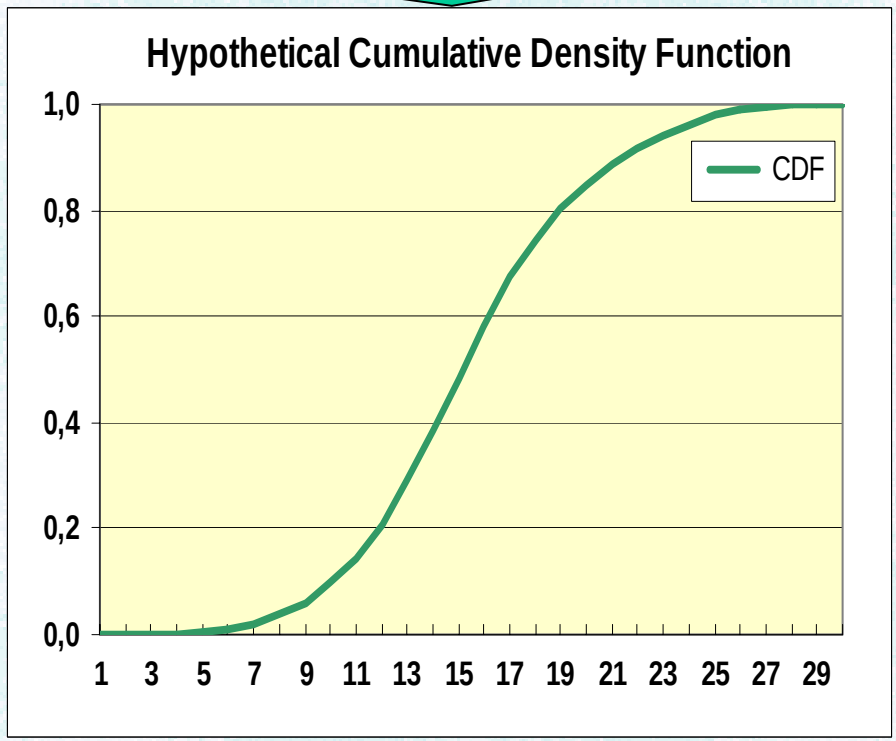
$$\mathbf{LEPS} = (1/n) \sum |CDF(f_i) - CDF(o_i)|$$

Range: 0 to 1
Perfect score = 0

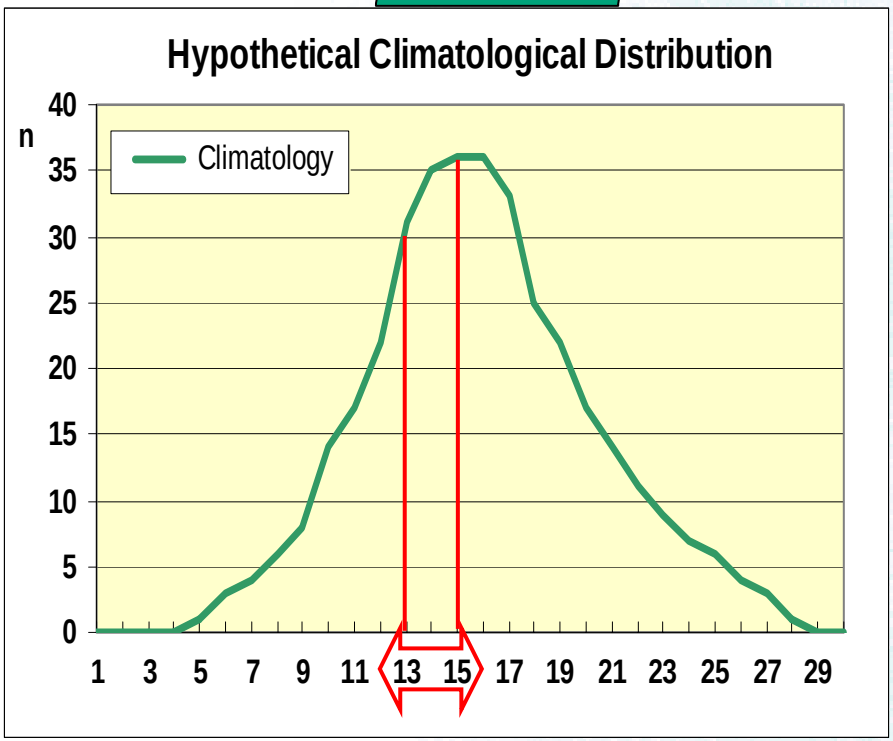
where **CDF** is the Cumulative probability Density Function of the forecasts (f_i) and observations (o_i), determined from a relevant climatology

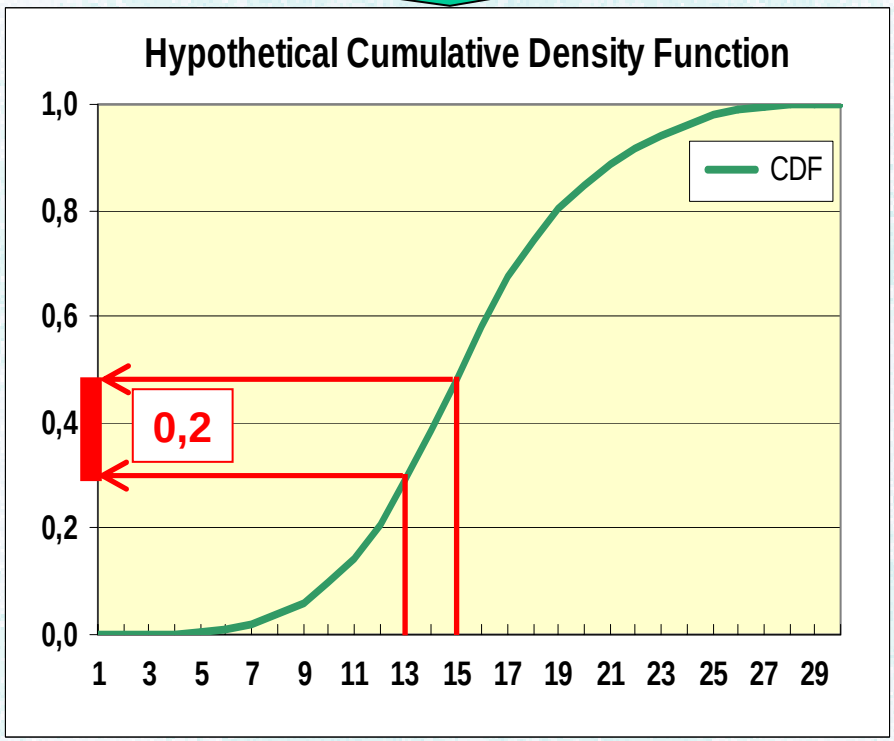
- Corresponds to **MAE** $[= (1/n) \sum |f_i - o_i|]$ of measurement space, transformed into probability space
- Does not depend on the scale of the variable
- Takes into account the variability of the weather element
- Can be used to evaluate forecasts at different locations
- Encourages forecasting in extreme tails of the climatological distribution
- Penalizes less than for similar sized errors in a more probable region of the distribution, i.e. opposite to (R)MSE
- Applicable for both continuous and categorical variables
- "Laborious": Computation requires definition of cumulative climatological distributions at each location





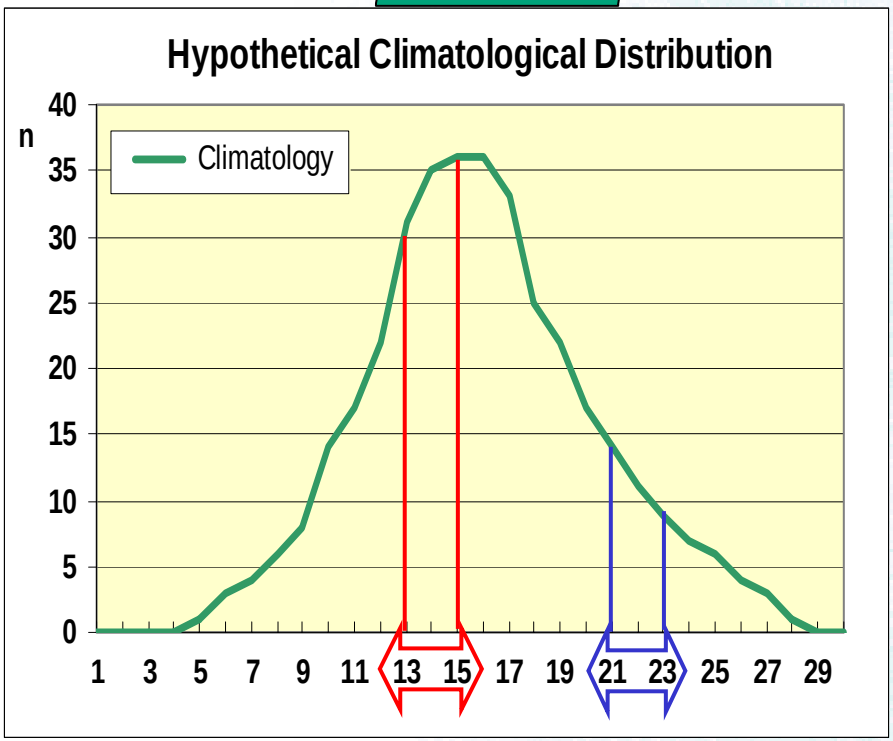
LEPS for a hypothetical distribution and location: The climatological frequency distribution (left) is transformed to a cumulative probability density distribution (right).

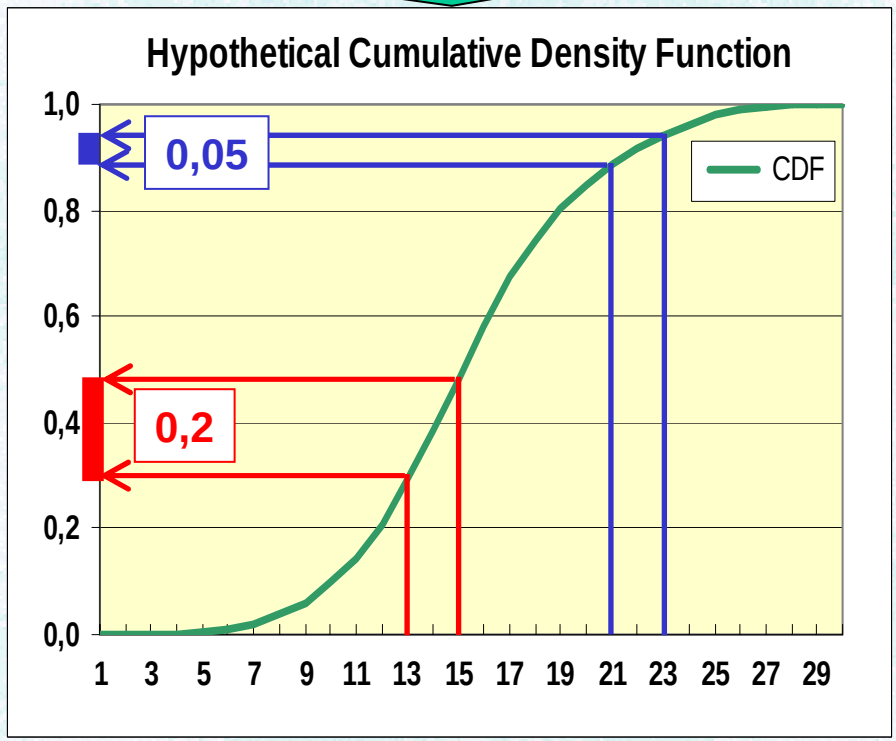




LEPS for a hypothetical distribution and location: The climatological frequency distribution (left) is transformed to a cumulative probability density distribution (right).

A 2 ”unit” forecast error around the median, 13 vs. 15 “units” (red arrows), would yield a LEPS value of c. 0.2 in the probability space (| 0.5 – 0.3 |, red arrows).





LEPS for a hypothetical distribution and location: The climatological frequency distribution (left) is transformed to a cumulative probability density distribution (right). A 2 ”unit” forecast error around the median, 13 vs. 15 “units” (red arrows), would yield a LEPS value of c. 0.2 in the probability space (| 0.5 – 0.3 |, red arrows).

An equal error in the measurement space close to the tail of the distribution, 21 vs. 23 ”units” (blue arrows), would result a LEPS value of c. 0.05 (| 0.95 – 0.9 |, blue arrows) => Fc errors of rare events are much less penalized using LEPS !

Revised, normalized, LEPS *(Potts et al., J. Clim., 1996, 34-53)*

$$LEPS_{rev} = 3 * (1 - | F_f - F_o | + F_f^2 - F_f + F_o^2 - F_o) - 1$$

where F_f and F_o are the CDFs of the forecasts and observations, respectively

Range: -1 to 2

- ”Is equitable”
- ”Normalization provided by the factor 3”
- ”Does **NOT** exhibit certain pathological behavior at the extremes (like `normal´ LEPS)”

General Skill Score

$$SS = (A - A_{ref}) / (A_{perf} - A_{ref})$$

where **A** = the applied measure of accuracy,

subscript "**ref**" refers to some reference forecast, "**perf**" to a perfect forecast

For negatively oriented accuracy measures like MAE, MSE, and **LEPS** :

Range: - oo to 1
Perfect score = 1

$$LEPS\_SS = 1 - LEPS / LEPS_{ref}$$

with respect to the climatological median, becomes:

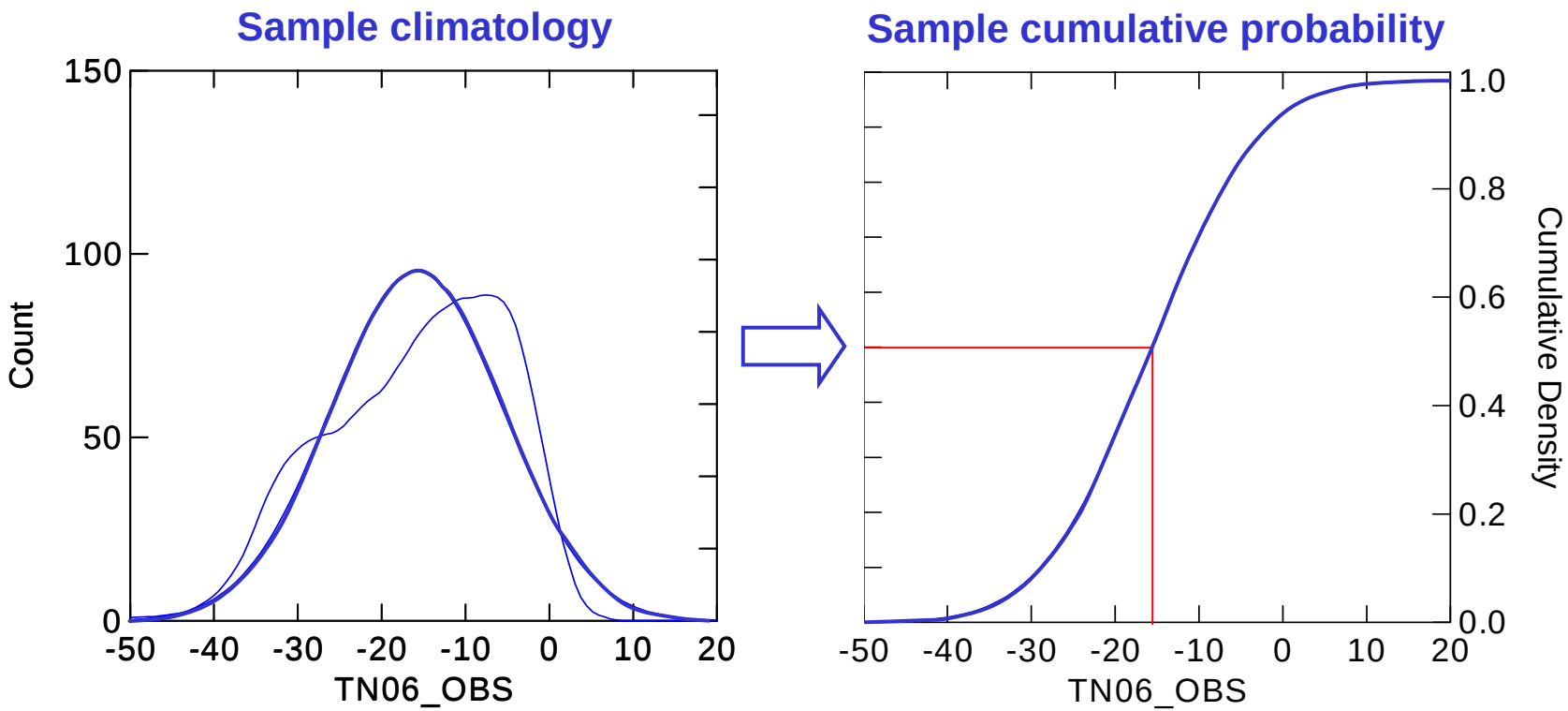
$$LEPS\_SS = 1 - [LEPS / (1/n) \Sigma | 0.5 - CDF(o_i) |]$$

Note: $LEPS_{rev}\_SS$, computed with respect to "a perfect forecast"

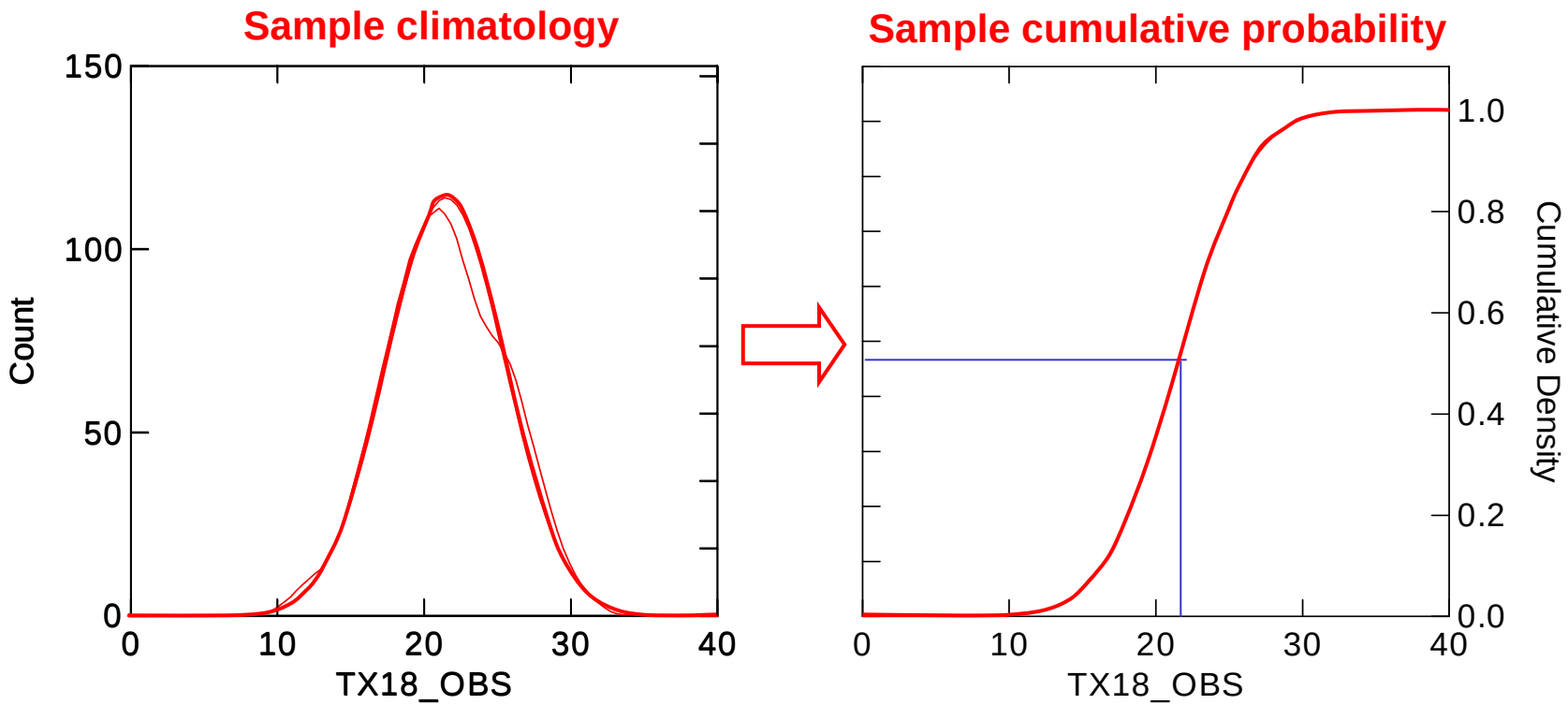
Data in the experimentation

- **Ten years, 1994 – 2003**
- **Winter (cold season) vs. Summer (warm season)**
- **T_{min} vs. T_{max}**
- **Two Finnish stations :**
  - Helsinki-Vantaa airport, in the south - Summer T_{max}
  - Sodankylä observatory, up north – Winter T_{min}

T_{min} – Winter (Sodankylä)

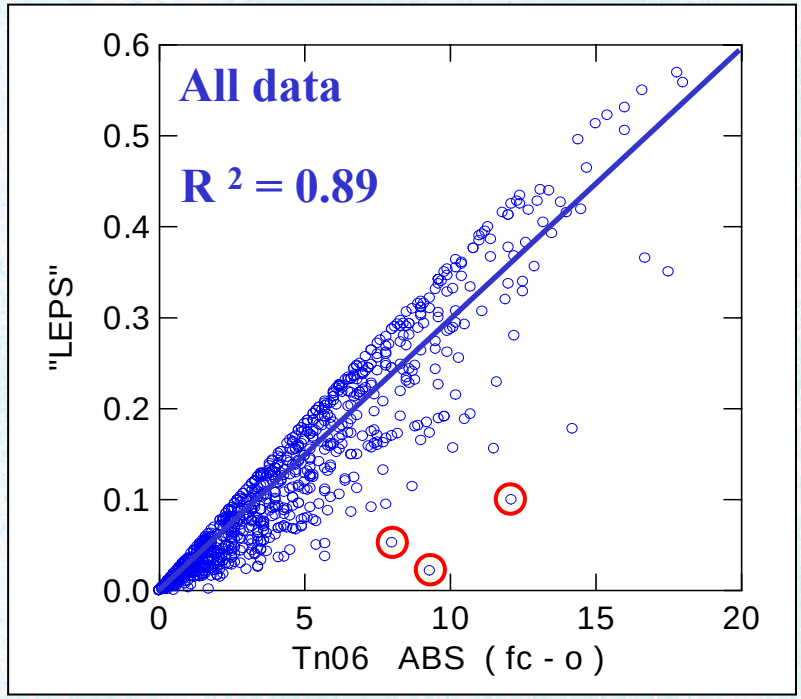


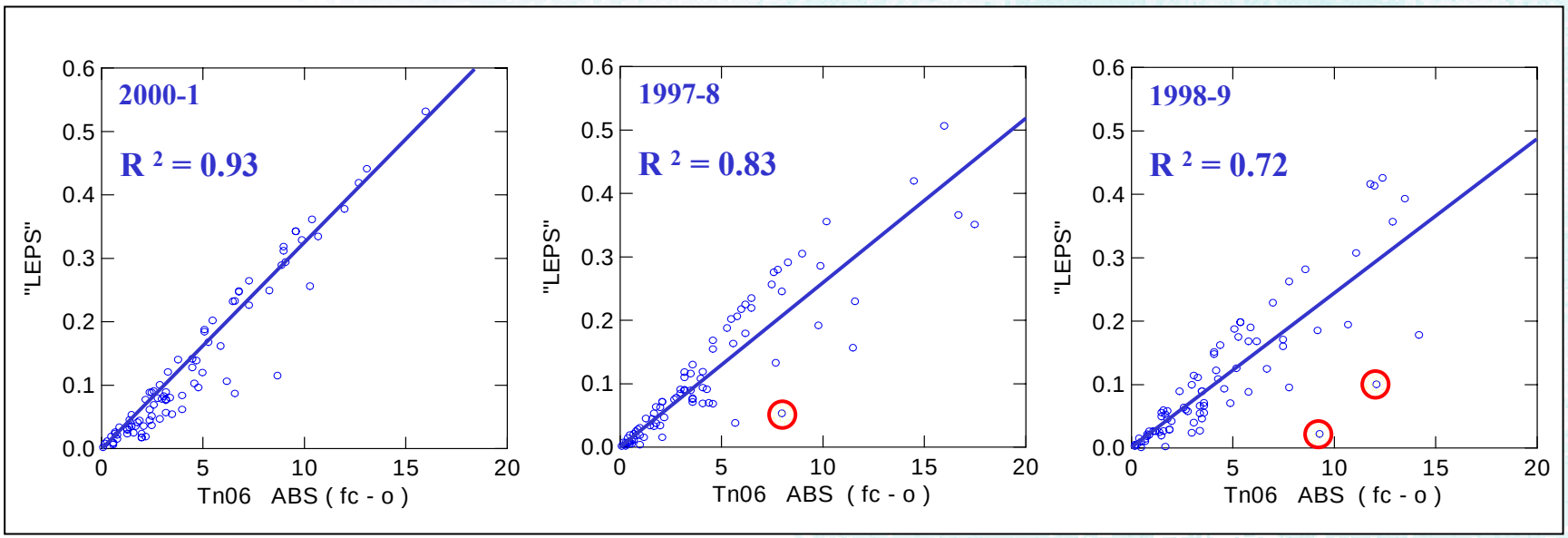
T_{max} – Summer (Helsinki-Vantaa)



T_{min} – Winter error:

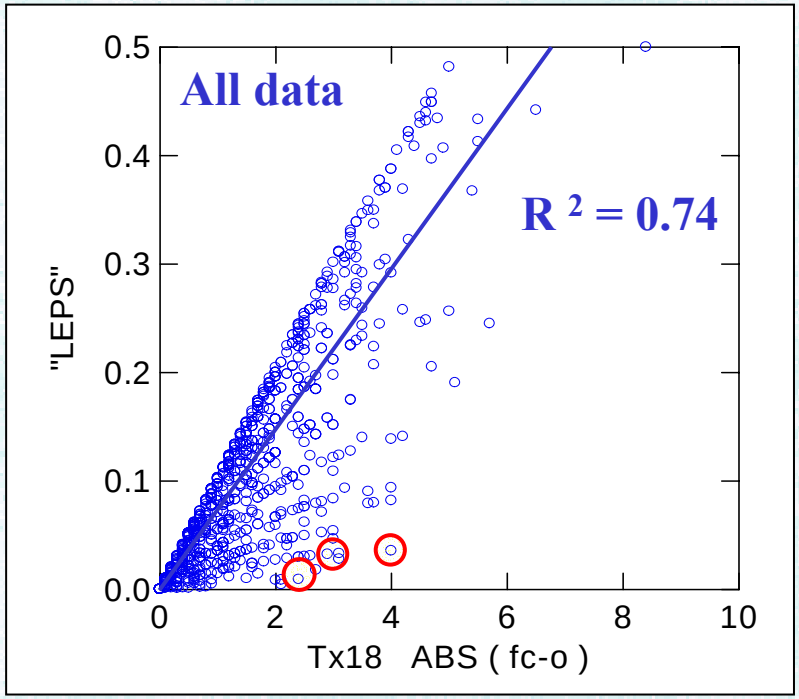
Measured absolute errors vs. "LEPS" scatterplot

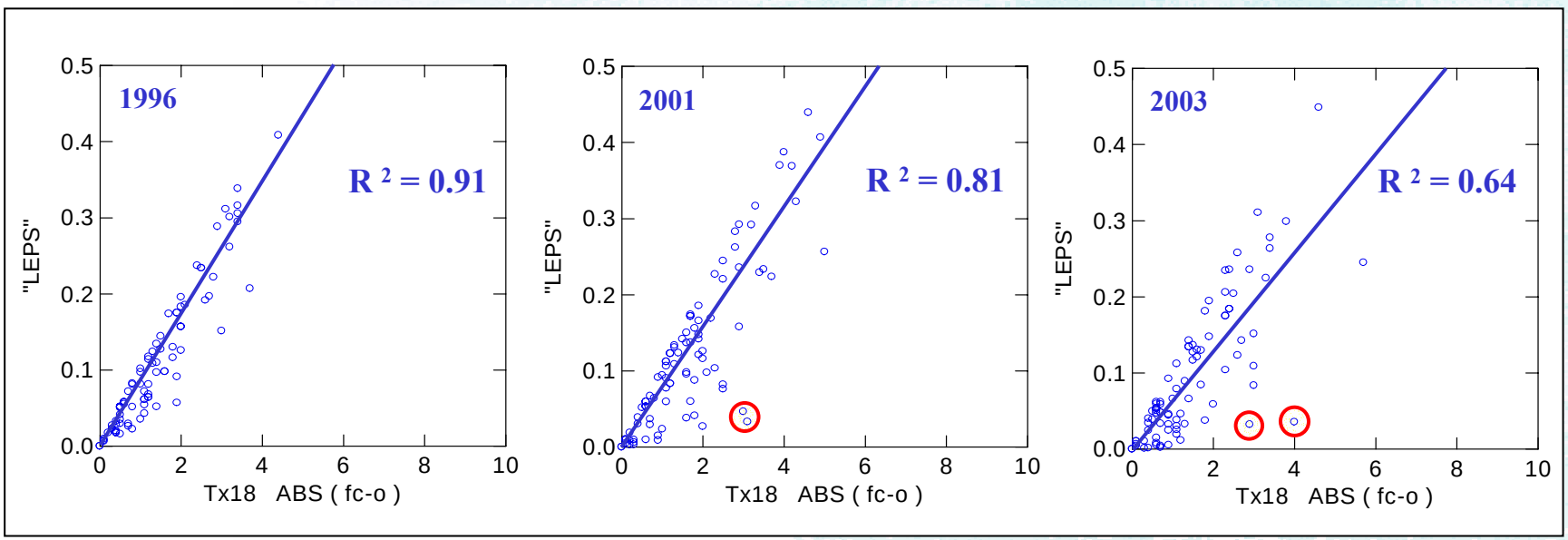




T_{max} – Summer error:

Measured absolute error vs. "LEPS" scatterplot



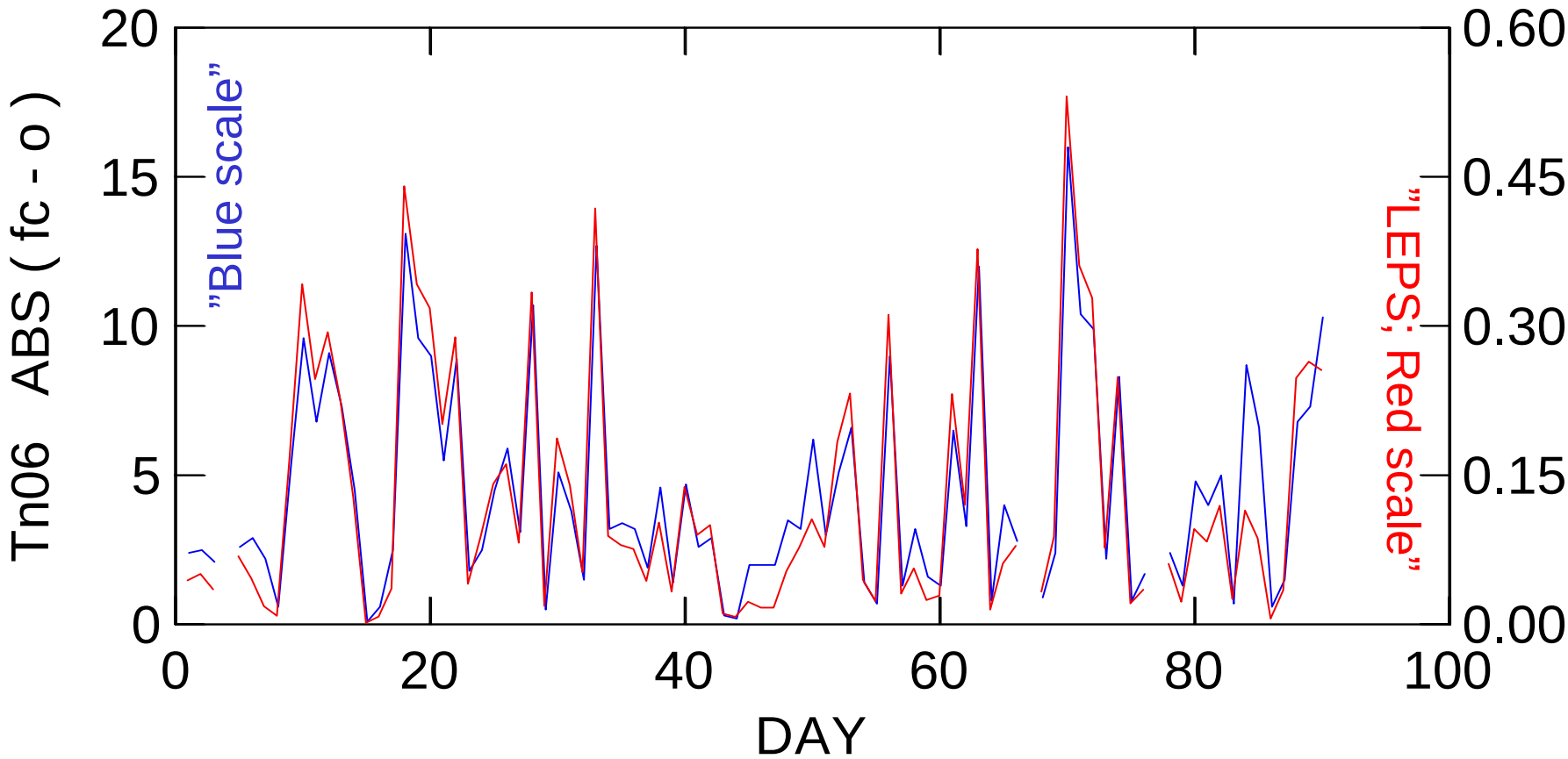


Coefficient of Determination, R^2

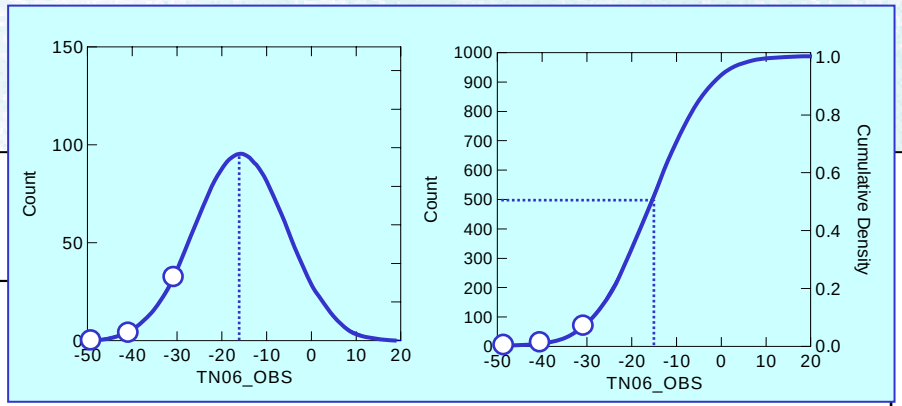
(proportion of the variation of LEPS "described" by the regression)

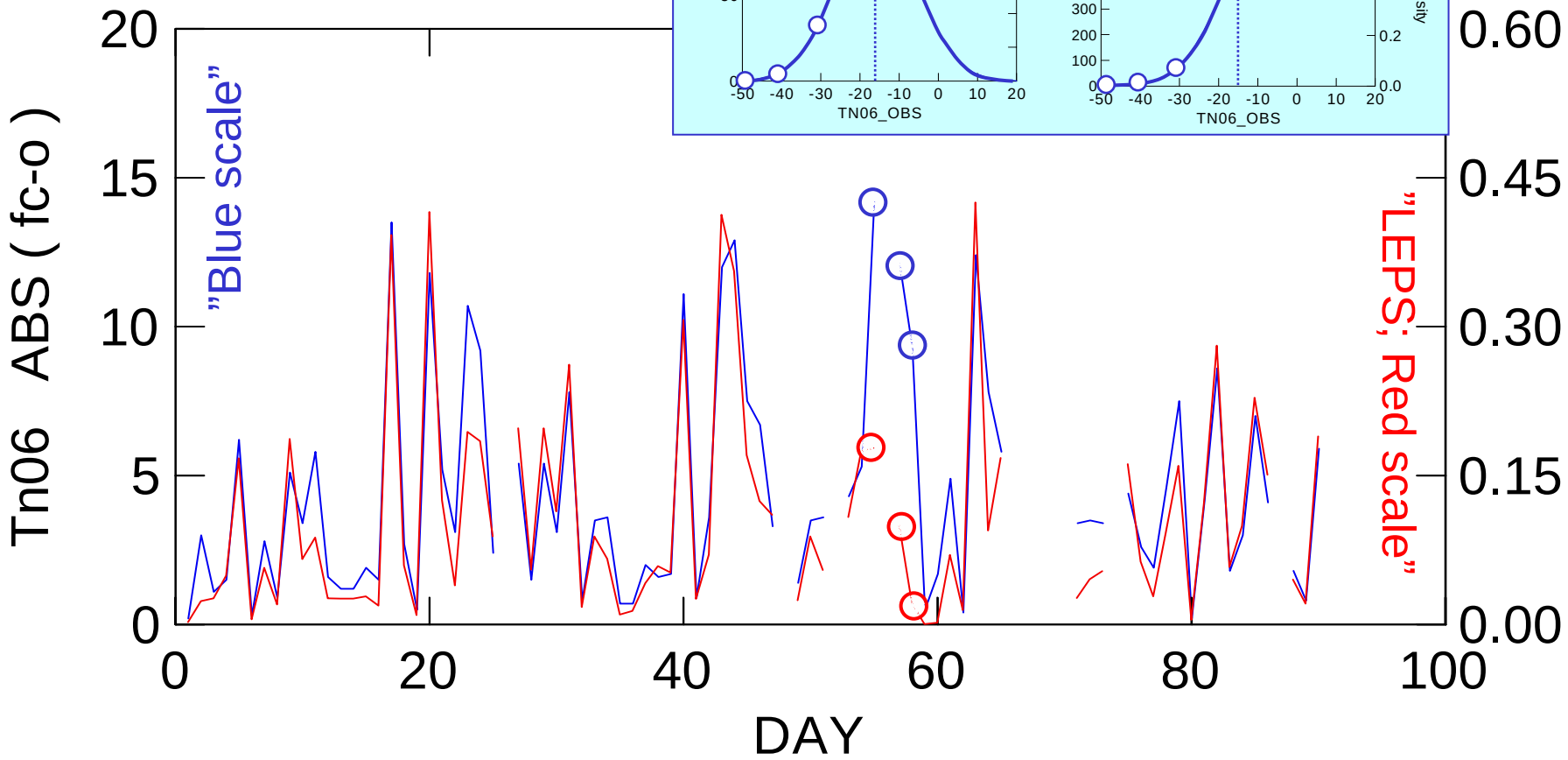
Year	T_{max} - Summer	T_{min} - Winter
1994	0.73	0.91
1995	0.69	0.95
1996	0.91	0.87
1997	0.65	0.83
1998	0.64	0.72
1999	0.63	0.93
2000	0.88	0.93
2001	0.81	0.94
2002	0.80	0.90
2003	0.64	0.92
1993-2003	*0.74*	*0.89*

T_{min} - Winter 2000-01
Hi correlation; $R^2 = 0.93$

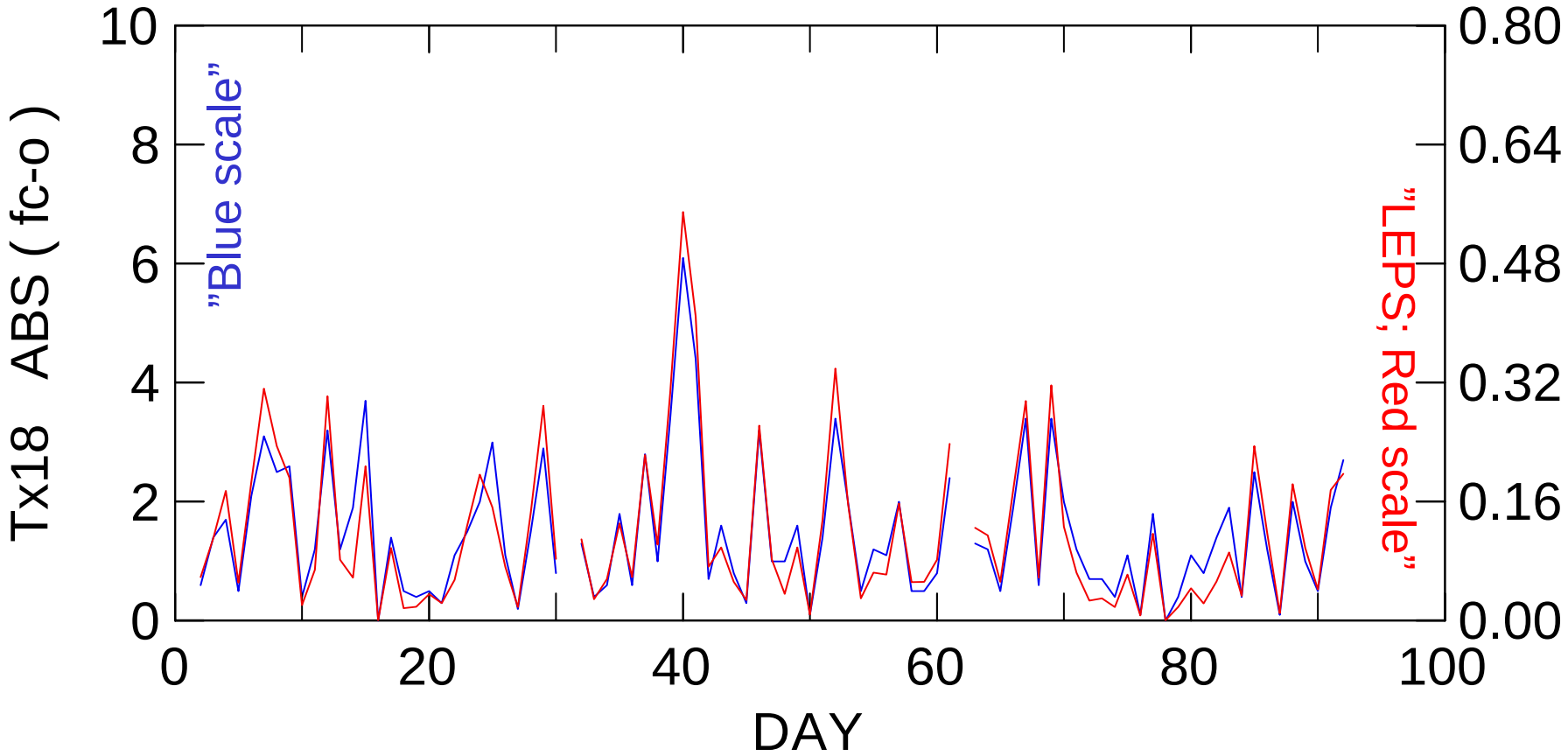


T_{min} - Winter 1998-99
Lo correlation ; $R^2 = 0.72$

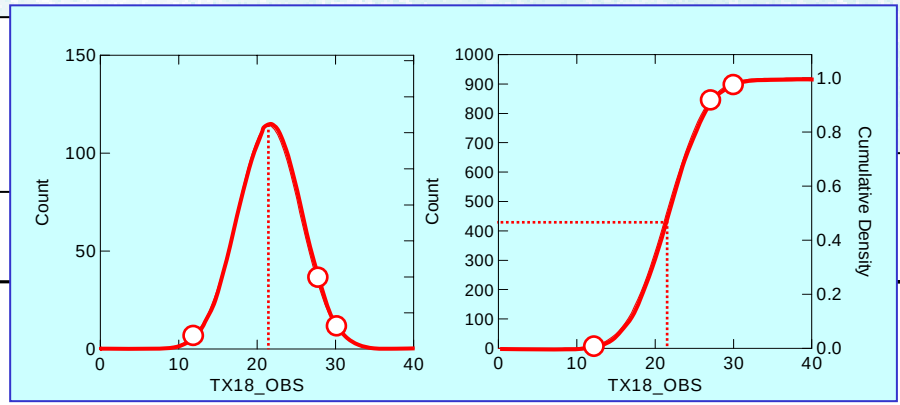


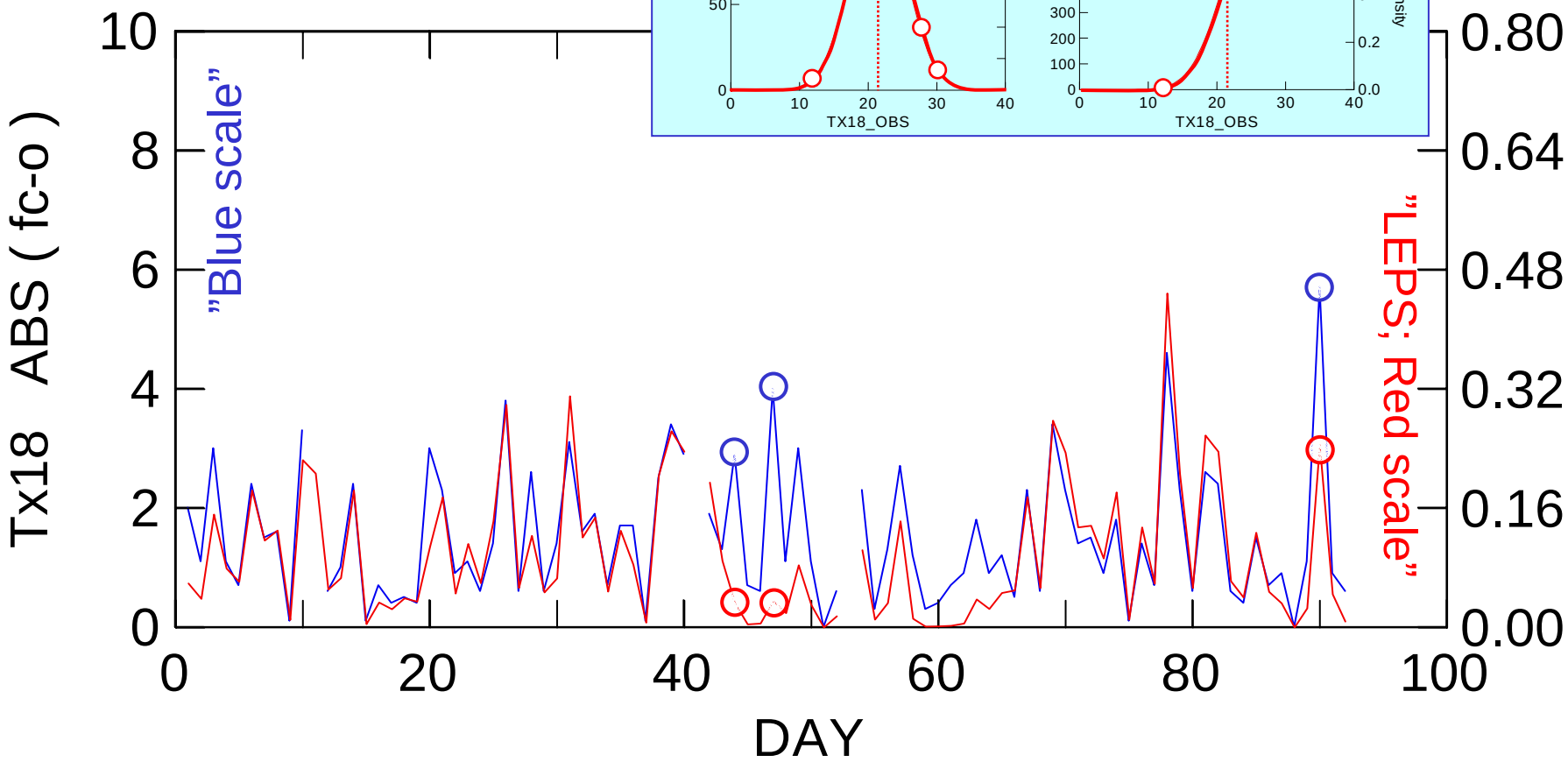


T_{max} - Summer 1996
Hi correlation ; $R^2 = 0.91$

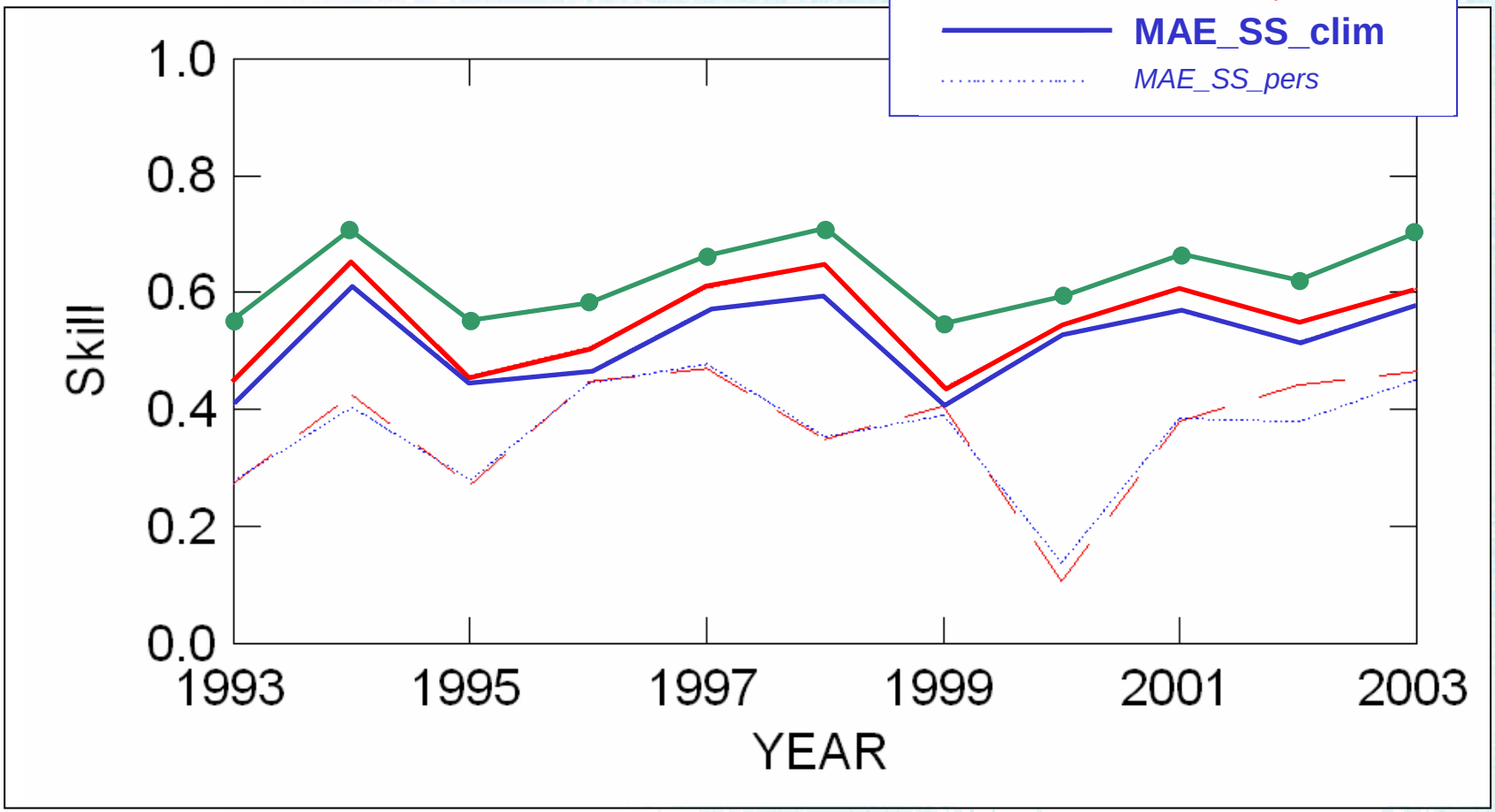


T_{max} - Summer 2003
Lo correlation ; $R^2 = 0.64$

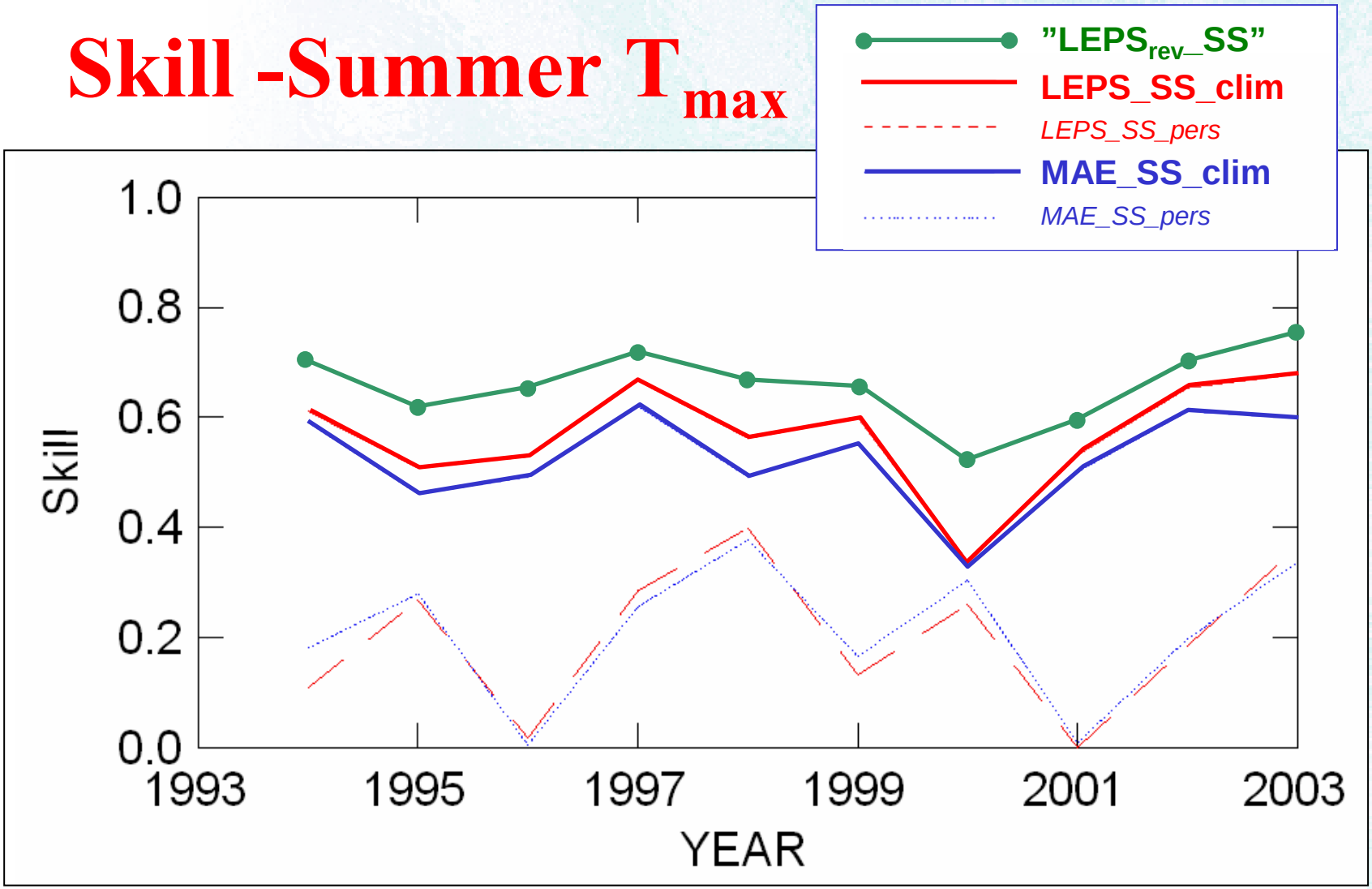




Skill -Winter T_{min}



Skill -Summer T_{max}



Conclusions:

- ✓ Reasoning for this study: New methodologies in forecast verification need be ”vigorously explored”
- ✓ LEPS appears to be a fairer score than e.g. (R)MSE or MAE when comparing climatologically different locations
- ✓ Skill: LEPS_SS produces slightly higher absolute values than MAE_SS
- ✓ $LEPS_{rev}$_SS qualitatively Similar to LEPS_SS
- ✓ Applicability to other parameters like precipitation, wind, etc. to be explored next