Issues in the verification of EPS

Åke Johansson
SMHI
Spread-Skill T2M from HIRLAM_K for JAN 2016

Spread & Skill(RMSE) : T2m
Verification Period: 20160101-20160131 Cycle: All
ALL Stations

Score
- RMSE
- Spread

Model
- HirEPS_K

Lead Time (hours)

Spread, RMSE

num/Cases

3e+05
2e+05
1e+05
0e+00
CU

The Commonly Used

Spread-Skill Relationship
$E[\text{RMSE}_M]$
\[ E[\text{RMSE}_M] \]

\[ E[\text{Std. Dev.}_E] \]

\[ \sqrt{[o - \overrightarrow{f}]^2} \]

\[ \sqrt{[f - \overrightarrow{f}]^2} \]
CU SKILL – SPREAD RELATION

\[ E\left[ \text{RMSE}_M \right] = E\left[ \text{Std. Dev.}_E \right] \]
CU SKILL – SPREAD RELATION

\[ E \left[ \text{RMSE}_M \right] = E \left[ \text{Std. Dev.}_E \right] \]

\[ E \left[ \text{MSE}_M \right] = E \left[ \text{VARE} \right] \]
CU Skill-Spread relation is only valid for an Ideal ensemble
4 reasons why
it is inappropriate for
Real World EPS
CU SKILL – SPREAD RELATION

\[ E \left[ \text{MSE}_M \right] = E \left[ \text{VARE} \right] \]
Real World Models usually have a BIAS
SKILL of Perturbed Members
As measured by $\text{MSE}_p$ and $\text{SDE}_p^2$

- $\text{MSE}_p$
- $\text{SDE}_p^2$
- $\text{SDE}_p^2 / \text{MSE}_p$ Ratio

Forecast Lead Time [Hour] vs. $[K*K]$ vs. $\text{MSE}_p / \text{SDE}_p^2$
\[ E[\text{MSE}_M] = E[\text{VARE}] \]
\[ \downarrow \]
\[ E[\text{SDE}_M^2] = E[\text{VARE}] \]
Real World Models usually have members that are not statistically equal.
\[ E\left[ SDE_M^2 \right] = E\left[ VARE \right] \]
\[ \downarrow \]
\[ E\left[ SDE_{MP}^2 \right] = E\left[ VARE_P \right] \]
Using U-Statistics for estimating Skill and Spread gives more realistic estimates
SKILL

$e_o$ $e$

$F_k$, $O$
\[
SPRE_p = \frac{2}{K \cdot (K - 1)} \sum_{k=1}^{K-1} \sum_{\kappa=k+1}^{K} [f(t_o, t_f, k) - f(t_o, t_f, \kappa)]^2
\]
\[ SPRE_p = \frac{2}{K \cdot (K - 1)} \sum_{k=1}^{K-1} \sum_{\kappa=k+1}^{K} \left[ f(t_o, t_f, k) - f(t_o, t_f, \kappa) \right]^2 \]
\[ \text{SPRE}_p = \frac{2}{K \cdot (K - 1)} \sum_{k=1}^{K-1} \sum_{\kappa=k+1}^{K} \left[ f(t_o, t_f, k) - f(t_o, t_f, \kappa) \right]^2 \]

**U-statistic**

A U-statistic is defined as the average – across all combinatorial selections of the given size from the full set of observations – of the basic estimator applied to the sub-samples.

U-statistics, where the letter U stands for unbiased, arise naturally in producing **minimum-variance unbiased estimators**.
\[ \text{SPRE}_P = \frac{2}{K \cdot (K-1)} \sum_{k=1}^{K-1} \sum_{\kappa=k+1}^{K} \left[ f(t_o, t_f, k) - f(t_o, t_f, \kappa) \right]^2 \]

\[ \text{VARE}_P = \left[ f(t_o, t_f, k) - \bar{f}(t_o, t_f) \right]^2 \]

\[ \text{SPRE}_P = 2 \frac{K}{K-1} \text{VARE}_P \iff \mathbb{E} \left[ \text{SPRE}_P \right] = 2 \mathbb{E} \left[ \text{VARE}_P \right] \]
\[ SDE_p^2 = \text{MSE}_p - \text{ME}_p^2 \]

U-statistic
$U.UI \text{ SKILL} - \text{SPREAD CONDITION}$

$E \left[ SDE^2_P \right] = E \left[ SPRE_P \right]$
UUI SKILL – SPREAD CONDITION

\[ E \left[ SDE_P^2 \right] = E \left[ SPRE_P \right] \]

\[ \frac{1}{2} E \left[ SDE_P^2 \right] = E \left[ VARE_P \right] \]
\[ CU \rightarrow U/UI \]

\[ E[MSE_M] = E[VARE] \]

\[ \downarrow \]

\[ E[SDE_M^2] = E[VARE] \]

\[ E[SDE_{MP}^2] = E[VARE_P] \]

\[ \downarrow \]

\[ \frac{1}{2} E[SDE_P^2] = E[VARE_P] \]
Example for T2M from HIRLAM_K for JAN 2016

Spread & Skill (RMSE) : T2m
Verification Period: 20160101-20160131 Cycle: All
ALL Stations

Score
- RMSE
- Spread

Model
- HirEPS_K

Graph showing spread and skill (RMSE) over lead time (hours) with all stations included.
Example for T2M from HIRLAM_K for JAN 2016

UUI Spread & Skill : T2m
Verification Period: 20160101-20160131 Cycle: All

- Commonly used (CU) spread-skill
- Bias corrected CU spread-skill
- Bias corrected CU spread skill (perturbed members only)
- U.UI spread-skill

Score
- Skill
- Spread

Model
- HirEPS_K
Why U-Statistics
Theoretical justification - Spread

\[ E[F] = a \]

\[ E[(F - a)^2] = E[F'^2] \]
Theoretical justification - Spread

\[ E[F] = a \]

\[ E[(F - a)^2] = E[F'^2] \]

\[ E[(F_i - F_j)^2] = E[(a + F_i' - a - F_j')^2] = \]

\[ = E[(F_i' - F_j')^2] = \]

\[ = 2 \cdot E[F'^2] \]
Theoretical justification - Spread

\[ E[(F - a)^2] = E[F'^2] \]

\[ E[(F_i - F_j)^2] = 2 \cdot E[F'^2] \]
Theoretical justification - Skill

\[ E[O] = 0 \quad , \quad E[F] = a \]

\[ E[(O - a)^2] = E[O'^2] + a^2 \]
Theoretical justification - Skill

\[ \text{E}[O] = 0, \quad \text{E}[F] = a \]

\[ \text{E}\left[(O - a)^2\right] = \text{E}[O'^2] + a^2 \]

\[ \text{E}\left[(F - O)^2\right] = \text{E}\left[(\text{E}(F) + F' - \text{E}(O) - O')^2\right] = \]

\[ = \text{E}\left[(a + F' - O')^2\right] = \]

\[ = \text{E}[F'^2] + \text{E}[O'^2] + a^2 \]
Theoretical justification - Skill

\[ E[(O - a)^2] = E[F'^2] + a^2 \]

\[ E[(F - a)^2] = 2 \cdot \left[ E[F'^2] + \frac{1}{2}a^2 \right] \]
\[
\begin{align*}
\text{Spread} & : & E[F'^2] & \quad & E[F'^2] & + a^2 \\
2 \cdot E[F'^2] & & 2 \cdot \left[E[F'^2] + \frac{1}{2} a^2\right]
\end{align*}
\]
Real World Models are usually not verified against the "Truth"
T850 EC HRES 00 UTC FEB 2015

Against observations
Against "truth" from observations
Against "truth" from analysis
Against analysis
\[ \vec{d}_{bi}^o = \vec{d}_{ai}^o + \vec{d}_{bi}^a \]
\[ \vec{d}_{bi}^o = \vec{d}_{ai}^o + \vec{d}_{bi}^a \]
\[ \vec{d}_{bi}^o = \vec{d}_{ai}^o + \vec{d}_{bi}^a \]
\[ \vec{d}_{bi}^o = \vec{d}_{ai}^o + \vec{d}_{bi}^a \]
\[ \overrightarrow{d}_{bi}^o = \overrightarrow{d}_{ai}^o + \overrightarrow{d}_{bi}^a \]
\[ \vec{d}_{bi}^o = \vec{d}_{ai}^o + \vec{d}_{bi}^a \]
Inflation = 1.0
Inflation = 1.1
Inflation = 1.16
Example for T2M from HIRLAM_K for JAN 2016
Conclusions

Over-dispersiveness – not under-dispersiveness – seems to be the problem with EPS

Therefore one can wonder whether the following really is necessary

- Parameter perturbations
- Stochastic physics
- Stochastic kinetic energy backscatter