Comparison of multivariate ensemble post-processing methods

Mária Lakatos

Faculty of Informatics, University of Debrecen, Hungary

25th Young Statisticians Meeting, October 15 – 17, 2021

This work was supported by the construction EFOP-3.6.3-VEKOP-16-2017-00002.
The project was supported by the European Union, co-financed by the European Social Fund.

Categorical forecasts

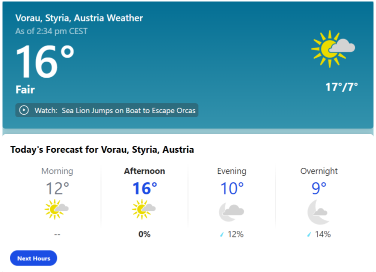
Point forecast: forecast obtained by a single run of a numerical weather prediction (NWP) model

A single forecast for each location, time and lead time

Numerical solution of a complex system of non-linear partial differential equations. Initial conditions from observations of the atmosphere

Problem: Lack of uncertainty information in predictions

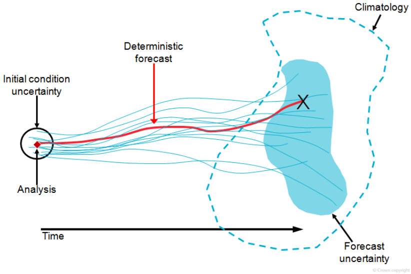
Solution: Ensemble prediction systems (EPS)[1]



www.weather.com

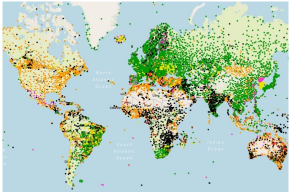
[1] Leith, C. E. (1974) Theoretical skill of Monte-Carlo forecasts. *Mon. Weather Rev.* **102**, 409—418

Ensemble forecasts



www.metoffice.gov.uk

- obtained by multiple runs of an NWP model
- better idea of what weather events may occur at a particular time
- forecast variation → uncertainty
- two factors of forecast skill decrease as forecast lead-time increases[1]
  1. uncertainties in the initial conditions
  2. approximations in the construction of a numerical model of the real atmospheric system



www.ecmwf.int

[1] www.ecmwf.int

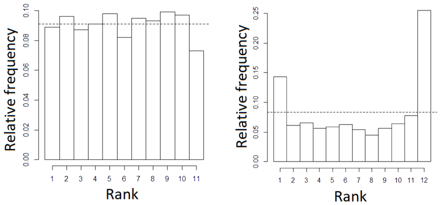
Ensemble forecasts

Nowadays, all major meteorological services operate ensemble prediction systems:

- European Centre of Medium-Range Weather Forecasts: ECMWF ensemble, 52 members
  - 1 control member obtained from the true initial conditions
  - 50 members from perturbed initial conditions. Statistically equivalent, hence exchangeable
  - 1 high resolution member
- Deutscher Wetterdienst: COSMO-DE ensemble, 30 members
- Hungarian Meteorological Service (HMS): AROME-EPS ensemble, 11 members

Problem: Ensemble predictions are often subject to forecast bias and dispersion errors, and are therefore uncalibrated[1]

Solution: Statistical post-processing[2]



Simulated reliable and the underdispersed temperature (K) ensemble predictions of the HMS

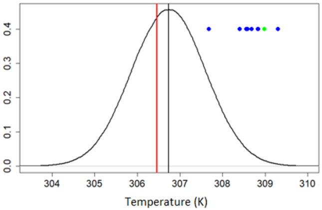
[1] Buizza, R. (2018) Ensemble forecasting and the need for calibration. In Vannitsem, S., Wilks, D. S., Messner, J. W. (eds.), *Statistical Postprocessing of Ensemble Forecasts*, Elsevier, Amsterdam, pp. 15–48.

[2] Gneiting, T. and Raftery, A. E. (2005) Weather forecasting with ensemble methods. *Science*, **310**, 248–249

Statistical post-processing

Ensemble Model Output Statistics (EMOS): yields in probabilistic forecasts that take the form of predictive probability density functions (PDFs) for continuous weather variables[1]

Parameter estimation: *CRPS* minimisation, by the help of training data



Example. Temperature (K) forecasts of the HMS for Debrecen, 8 July 2012, 12 UTC

EMOS models for different weather quantities:

Variable	Domain	Distribution
temperature	$y \in \mathbb{R}$	normal, skewed normal
wind speed	$y \in \mathbb{R}_+$	truncated normal, log-normal, $\mathcal{GEV}$, truncated $\mathcal{GEV}$
precipitation	$y \in \mathbb{R}_+$	censored $\mathcal{GEV}$, censored shifted gamma
global irradiance	$y \in \mathbb{R}_+$	censored logistic

[1] Gneiting, T., Raftery, A. E., Westveld, A. H. and Goldman, T. (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev.* **133**, 1098–1118.

EMOS model for temperature data

$f_{\overline{exch}}, f_{CTRL}, f_{HRES}$: mean of exchangeable members, control- and high resolution forecasts

y and m: corresponding observation and ensemble size

Regression model to capture the seasonal variation of temperature data:

$$y_{t_j} = c_0 + c_1 \sin\left(\frac{2\pi j}{365}\right) + c_2 \cos\left(\frac{2\pi j}{365}\right) + \epsilon_j, \qquad j = 1, \dots, n$$

$\hat{y}_t, \hat{f}_{HRES}, \hat{f}_{\overline{exch}}$ and $\hat{f}_{CTRL}$: climatological means (fitted values)

n : length of training period

Parameters of the normally distributed EMOS model[1]:

$$\mu = \hat{y} + \beta_0(f_{HRES} - \hat{f}_{HRES}) + \beta_1(f_{CTRL} - \hat{f}_{CTRL}) + \beta_2(f_{\overline{exch}} - \hat{f}_{\overline{exch}}) \qquad \text{and} \qquad \sigma^2 = c + ds^2$$

s^2: ensemble variance

$\beta_0, \beta_1, \beta_2, c$ and d: estimated parameters (constrained to be non-negative)

[1] Gneiting, T., Raftery, A. E., Westveld, A. H. and Goldman, T. (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev.* **133**, 1098–1118.

EMOS model for wind speed data

Left-truncated (at zero) normal distribution[1] on a square root-transformed space (model $\sqrt{y}$)

Predictive density function:

$$f(y;\mu,\sigma) = \frac{\frac{1}{\sigma}\phi((y-\mu)/\sigma)}{\Phi(\mu/\sigma)}, \qquad y \geq 0,$$

$\phi(\cdot)$ and $\Phi(\cdot)$: PDF and CDF of the standard normal distribution.

Parameters of the truncated normal distribution:

$$\mu = \beta_0 + \beta_1\sqrt{f_{HRES}} + \beta_2\sqrt{f_{CTRL}} + \beta_3\sqrt{f_{exch}} \quad \text{and} \quad \sigma^2 = c + d\,\boldsymbol{MD}_{\sqrt{f}}$$

$$\boldsymbol{MD}_{\sqrt{f}} = m^{-2}\sum_{i,i'=1}^{m}\left|\sqrt{f_i} - \sqrt{f_{i'}}\right|$$

$\beta_0, \beta_1, \beta_2, \beta_3, c$ and d: estimated parameters

[1] Thorarinsdottir, T. L., & Gneiting, T. (2010) Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *J R Stat Soc Ser A Stat Soc*, **173**(2), 371–388

EMOS model for precipitation data

Left-censored (at zero) generalized extreme value ($\mathcal{GEV}$) distribution[1]

Predictive CDF of the $\mathcal{GEV}$ distribution:

$$F(y;\mu,\sigma,\xi)=\begin{cases}\exp\left(-\left(1+\xi\frac{y-\mu}{\sigma}\right)\right), & \xi\neq 0,\\ \exp\left(-\exp\left(-\frac{y-\mu}{\sigma}\right)\right), & \xi=0,\end{cases} \qquad \text{where } 1+\xi\left(\frac{y-\mu}{\sigma}\right)\geq 0.$$

Censoring of the predictive CDF:

$$F_c(y;\mu,\sigma,\xi)=\begin{cases}F(y;\mu,\sigma,\xi), & y\geq 0,\\ 0, & y<0.\end{cases}$$

Parameters of the censored $\mathcal{GEV}$ distribution:

$$\mu=\beta_0+\beta_1 f_{HRES}+\beta_2 f_{CTRL}+\beta_3 \overline{f_{exch}}+\beta_4\pi_0, \qquad \sigma^2=c+d\boldsymbol{MD}_f \qquad \text{and} \qquad \xi=0.2.$$

π_0 and $\boldsymbol{MD}_f$: proportion of zero forecasts in the ensemble and ensemble mean difference

$\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,c$ and d: estimated parameters

[1] Scheuerer, M. (2014) Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Q. J. R. Meteorol. Soc*, **140**(680), 1086–1096.

Model evaluation

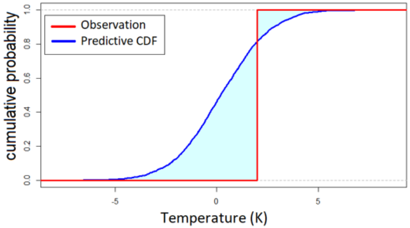
$\hat{x}_t^\ell$: verifying observation of the weather quantity at location ℓ and time t.

$P_{\ell,t}(x)$: estimated predictive CDF at location ℓ and time t with a given lead time.

Mean continuous ranked probability score (CRPS)[1]:

$$\overline{CRPS} := \frac{1}{N}\sum_{\ell,t} CRPS(P_{\ell,t}(x), \hat{x}_t^\ell)$$

$$CRPS(P,x) := \int_{-\infty}^{\infty} (P(y) - \mathbb{1}_{\{y \geq x\}})^2 dy$$



CRPS based on a sample from the forecast distribution:

$$CRPS(\hat{F}_m, x) = \frac{1}{m}\sum_{i=1}^{m} |f_i - x| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} |f_i - f_j|$$

Negatively oriented score: the smaller, the better

[1] Gneiting, T., Raftery, A. E., Westveld, A. H. and Goldman, T. (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev.* **133**, 1098–1118.

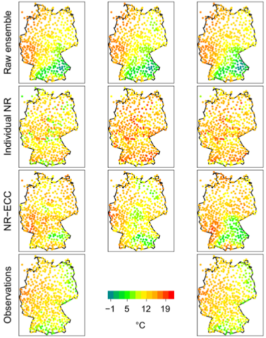
Multivariate post-processing

The downside of post-processing: Spatial- and temporal dependencies across marginals may be lost after applying univariate post-processing

Aim: Reinstall the possibly lost dependencies

Techniques:

- Fitting multivariate distribution[1]
- Application of parametric copulas[2]
- Empirical copula-based post-processing



Example. Effect of MVPP[3]

[1]Baran, S. & Möller, A. (2017) Bivariate ensemble model output statistics approach for joint forecasting of wind speed and temperature. *Meteorol. Atmos. Phys.*, **129**, 99—112

[2]Möller, A., Lenkoski, A., & Thorarinsdottir, T. L. (2013) Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Q. J. R. Meteorol Soc*, *139*, 982–991.

[3]Pinson, P. & Messner, J. W. (2018) Ensemble Postprocessing Methods Incorporating Dependance Structures. In Vannitsem, S.,Wilks, D. S. & Messner, J. W. (eds.), Statistical Post-processing of Ensemble Forecasts, Elsevier, Amsterdam, pp. 93.

Applied methods

Ensemble Copula Coupling (ECC):[1]

1. A sample $\hat{x}_1^\ell, \dots, \hat{x}_m^\ell$ is drawn from each of the post-processed CDFs F_θ^ℓ of station/lead time ℓ where m the number of ensemble members
2. The sampled values are rearranged in the rank order structure of the raw ensemble

Variants:

- ECC-Q: equidistant quantiles
- ECC-S:[2]
    1. Generate $u_1, \dots, u_m$ random numbers, where $\mathcal{U}(\frac{i-1}{m}, \frac{i}{m})$
    2. $\hat{x}_i^\ell$ is set to the quantile of F_θ^ℓ at level u_i
- ECC-R: Random sampling
- d-ECC: Combination of forecast error-correlation and raw ensemble rank order structure[3]

[1] Schefzik, R., Thorarinsdottir T. L., & Gneiting, T. (2013) Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, **28**, 616–640

[2] Hu, Y., Schmeits, M. J., van Andel, J. S., Verkade, J. S., Xu, M., Solomatine, D. P., & Liang, Z. (2016) A stratified sampling approach forimproved sampling from a calibrated ensemble forecast distribution. *J. Hydrometeorol*, **17**, 2405–2417.

[3] Ben Bouallègue, Z., Heppelmann, T., Theis, S. E., & Pinson, P. (2016) Generation of scenarios from calibrated ensemble forecasts with a dual-ensemble copula-coupling approach. *Mon. Weather Rev.*, **144**, 4737–4750.

Applied methods

Schaake Shuffle: Uses observations from a historical dataset as dependence template

Variants:

- Random Schaake Shuffle[1]
- Similarity-based Schaake Shuffle[2]
- Minimum Divergence Schaake Shuffle[3]

Gaussian copula approach (GCA): Linking calibrated marginals with a Gaussian copula and draw multivariate samples from the resulted distribution[4]

Reference models: Univariate samples from the calibrated CDFs

[1] Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., & Wilby, R. (2004) The Schaake Shuffle: a method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol*, **5**, 243-–262.

[2] Schefzik, R. (2016) A similarity-based implementation of the Schaake shuffle. *Mon. Weather Rev.*, **144**, 1909–1921.

[3] Scheuerer, M., Hamill, T. M., Whitin, B., He, M., & Henkel, A. (2017) A method for preferential selection of dates in the Schaake shuffle approach to constructing spatiotemporal forecast fields of temperature and precipitation. *Water Resour. Res.*, **53**, 3029–3046.

[4] Möller, A., Lenkoski, A., & Thorarinsdottir, T. L. (2013) Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Q. J. R. Meteorol Soc*, **139**, 982–991.

Multivariate verification scores

$\boldsymbol{x} = \left(x^{(1)}, \dots, x^{(d)}\right) \in \mathbb{R}^d$

F forecast distribution on $\mathbb{R}^d$ given through m discrete samples $\boldsymbol{f}_i, \dots \boldsymbol{f}_m$ from F with $\boldsymbol{f}_i = (f_i^1, \dots, f_i^d)$

Energy score (ES)[1]: multivariate generalization of CRPS

$$ES(F, x) = \frac{1}{m} \sum_{i=1}^{m} ||\boldsymbol{f}_i - \boldsymbol{x}|| - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} ||\boldsymbol{f}_i - \boldsymbol{f}_j||$$

where $||\cdot||$ denotes the Euclidean norm on $\mathbb{R}^d$.

Variogram score of order p (VS-p)[2]:

$$VS(F, x) = \sum_{i=1}^{d} \sum_{j=1}^{d} \omega_{ij} \left(|x^{(i)} - x^{(j)}|^p - \frac{1}{m} \sum_{k=1}^{m} |f_k^{(i)} - f_k^{(j)}|^p \right)^2$$

where ω_{ij} is a non-negative weight. Typical choices of p include 0.5 and 1. Negatively oriented scores

Skill scores: Improvement in ES or VS with respect to a F_{ref} reference forecast. Energy skill score (ESS) and Variogram skill score (VSS) respectively (positively oriented):

$$ESS := 1 - \overline{ES}/\overline{ES}_{ref} \qquad VSS := 1 - \overline{VS}/\overline{VS}_{ref}$$

[1] Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., & Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST*, **17**, 211–235.

[2] Scheuerer, M., & Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Mon. Weather Rev.*, **143**, 1321–1334.

Ongoing work

Methods for comparing different multidimensional techniques:

- Simulation-based comparison[1]
- Comparison on real data

Data

ECMWF ensemble forecasts and corresponding observations

- 2 meter temperature (T2M): 4588 stations, January 2004 to March 2014
- 10 meter wind speed (V10): 4509 stations, January 2004 to March 2014
- 24 hour precipitation accumulation (PPT24): 2956 stations, January 2007 to March 2014

Lead time: 1 – 10 days

Data period: approx. 12 years

[1]Lerch, S., Baran, S., Möller, A., Groß, J., Schefzik, R., Hemri, S., & Graeter, M. (2020) Simulation-based comparison of multivariate ensemble post-processing methods. *Nonlinear Process. Geophys.*, **27**, 349–371

[2]ECMWF Directorate (2012). Describing ECMWF's forecasts and forecasting system. *ECMWF Newsletter*, **133**, 11–13.

Workflow

1. Univariate post-processing of ensembles[1]
   - T2M: Variant of standard normal EMOS model, which captures the seasonal component of temperature data (length of training period: 720 days)
   - V10: Standard truncated normal EMOS model on square root transformed space (length of training period: 720 days)
   - PPT24: Censored generalized extreme value distribution (length of training period: 1826 days)
2. Multivariate post-processing

Spatial	Temporal
Step 1: Select the 10 closest stations	Step 1: Select 1 station and consider all lead times

Step 2: Apply MV techniques

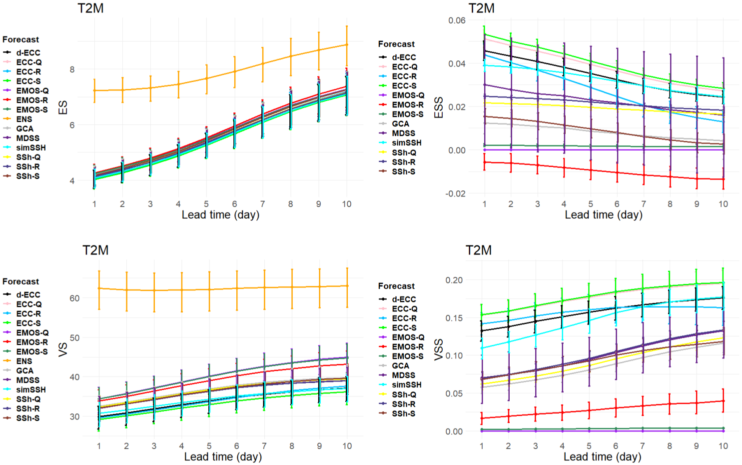
Step 3: Evaluation

} Repeat 100 times

[1] Hemri, S., Scheuerer, M., Pappenberger, F., Bogner, K., & Haiden, T. (2014). Trends in the predictive performance of raw ensemble weather forecasts. *Geophys. Res. Lett.*, **41**, 9197–9205.

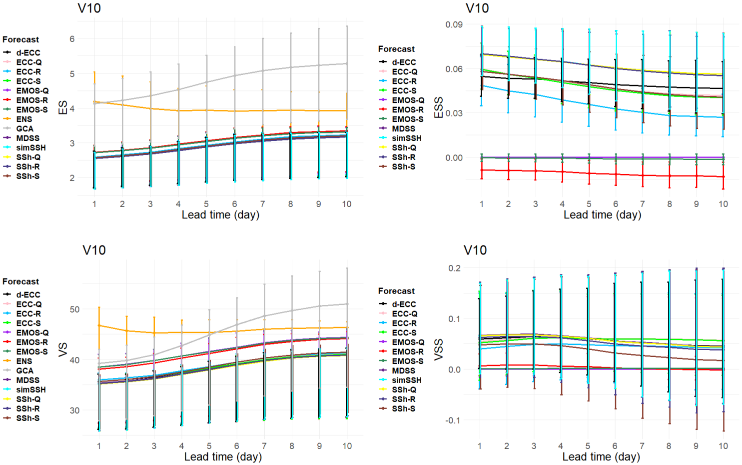
Spatial dependencies

Mean energy-, energy skill-, variogram- and **variogram skill scores** and their 95% CIs for 2 meter temperature forecasts (T2M) respectively.



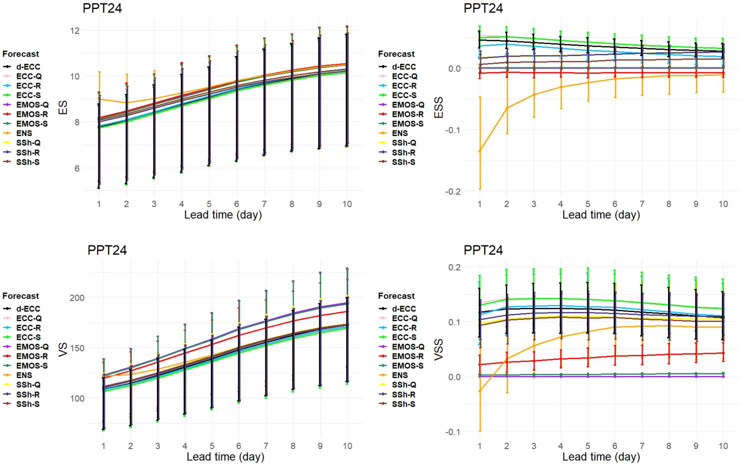
Spatial dependencies

Mean energy-, energy skill-, variogram- and **variogram skill scores** and their 95% CIs for 10 meter wind speed forecasts (V10) respectively.



Spatial dependencies

Mean energy-, energy skill-, variogram- and **variogram skill scores** and their 95% CIs for 24-hour precipitation forecasts (PPT24) respectively.



Temporal dependencies

Mean energy- and **variogram scores** respectively for the ECMWF ensembles

Verification score	ES			VS		
Weather quantity	T2M	V10	PPT24	T2M	V10	PPT24
ENS	6.8484	5.2119	9.9841	43.0150	52.6785	152.8449
EMOS-Q	5.9250	4.1480	12.2164	57.8222	67.1869	226.9377
EMOS-R	5.9611	4.1683	12.3025	55.1884	60.8643	227.0591
EMOS-S	5.9175	4.1426	12.1932	57.5562	67.1451	222.5307
ECC-Q	5.5480	3.8415	11.8853	42.2730	47.9586	205.2279
ECC-R	5.5983	3.8732	12.1150	42.9228	48.7029	210.6007
ECC-S	5.5421	3.8373	11.8998	42.3440	48.3893	205.9031
SSh-Q	5.7169	3.8505	11.7345	48.6426	47.2339	205.0069
SSh-R	5.7053	3.8422	11.8559	48.0636	47.3477	206.9210
SSh-S	5.7596	3.8796	11.6954	47.7979	47.6965	203.2355
MDSS	5.7862	3.8044	11.6832	48.8265	46.4642	203.7148
simSSH	5.7171	3.8503	11.7340	48.6501	47.2360	205.0071
d-ECC	5.5518	3.8522	10.0266	42.4865	48.7400	154.4293
GCA	5.7631	5.6649	-	48.6404	49.9520	-

Conclusions

- The multivariate approaches are generally successful in the modelling of possibly lost dependencies (esp. for T2M and V10)
- However, no single best performing method can be identified
- It may depend on:
  - weather variable
  - lead time
  - specification of correlation structure
  - length of historical datasets

Future plans

- Testing approaches with extremely long execution time (for PPT24)
- Evaluate multivariate performance with additional verification scores and graphical diagnostic tools.

Thank you for your attention!