Statistical Postprocessing of Ensemble Weather Forecasts

Tilmann Gneiting

Institute for Applied Mathematics
Heidelberg University

RMI Centenary, Brussels
September 26, 2013

joint work with colleagues and students at the University of Washington and Heidelberg University
supported by NSF, ONR, DWD, VWStiftung and ERC

Introduction: Probabilistic Forecasts

Statistical Postprocessing of NWP Ensembles:
EMOS/NR and BMA

Accounting for Dependencies:
Ensemble Copula Coupling (ECC)

Case Study

Discussion

Introduction: Probabilistic Forecasts

Statistical Postprocessing of NWP Ensembles:
EMOS/NR and BMA

Accounting for Dependencies:
Ensemble Copula Coupling (ECC)

Case Study

Discussion

Probabilistic forecasts

weather forecasting is considered the **ultimate problem** in meteorology (Bjerknes 1904)

in current practice, medium-range weather forecasting is based on **numerical weather prediction (NWP)** models that represent the physics and chemistry of the atmosphere

however, there are major sources of **uncertainty**, including uncertainty about **initial conditions** and **model parameters**

thus, attention has turned to **probabilistic forecasts**, taking the form of probability distributions over future weather states

preferred approach to probabilistic weather prediction is based on carefully designed **ensembles** of NWP model runs

global medium-range **ensemble prediction systems** have been operational at the **ECMWF** and **NCEP** since December 1992

Towards probabilistic weather forecasting

Tim Palmer (2000):

> Although forecasters have traditionally viewed weather prediction as deterministic, a *culture change towards probabilistic forecasting* is in progress.

Tim Palmer (2012):

> [. . .] in the coming decade, *NWP centres should* start to *focus exclusively on* developing *probabilistic forecast systems*, dropping their separate higher-resolution deterministic forecast systems, and, importantly *measuring progress*, and formulating strategic goals, principally *in terms of* improvements to *probabilistic scores*.

What is a good probabilistic forecast?

Gneiting, Balabdaoui and Raftery (2007) contend that the goal of probabilistic forecasting is to *maximize the sharpness of the predictive distributions subject to calibration*

calibration

refers to the **statistical compatibility** between the **predictive distributions** and the verifying **observations**

- joint property of the forecasts and the observations
- in a nutshell, the observations are supposed to behave like random numbers sampled from the predictive distributions
- can be assessed via **rank** or **probability integral transform (PIT)** histograms

sharpness

refers to the **spread** of the **predictive distributions**

- property of the probabilistic forecasts only

Proper scoring rules

proper scoring rules allow for the **joint** assessment of **calibration** and **sharpness**

a **scoring rule** is a function

$$\mathsf{s}(F, y)$$

that assigns a numerical score to each pair (F, y), where F is the **predictive distribution** and y is the realizing **observation**

we consider scores to be **negatively oriented** penalties that forecasters aim to **minimize**

a **proper scoring rule** s satisfies the expectation inequality

$$\mathbb{E}_G\, \mathsf{s}(G, Y) \leq \mathbb{E}_G\, \mathsf{s}(F, Y) \quad \text{for all} \quad F, G,$$

thereby encouraging **honest** and **careful** assessments (Gneiting and Raftery 2007)

Continuous ranked probability score

in meteorological applications, the most popular proper score is the **continuous ranked probability score (CRPS)**,

$$\begin{aligned} \mathrm{s}(F,y) &= \int_{-\infty}^{\infty} (F(x) - \mathbb{1}(x \geq y))^2 \, \mathrm{d}x \\ &= \mathbb{E}_F\,|X - y| - \frac{1}{2}\,\mathbb{E}_F\,|X - X'| \end{aligned}$$

where X and X' are independent random variables with cumulative distribution function F (Matheson and Winkler 1976; Hersbach 2000; Gneiting and Raftery 2007)

- the CRPS is reported in the **same unit as the observations**
- in the case of a single-valued forecast, the CRPS reduces to the **absolute error**
- thus, the CRPS provides a **direct way** of **comparing single-valued** forecasts and **probabilistic** forecasts
- the CRPS is a special case of the proper **energy score (ES)** for multivariate quantities

Introduction: Probabilistic Forecasts

Statistical Postprocessing of NWP Ensembles:
EMOS/NR and BMA

Accounting for Dependencies:
Ensemble Copula Coupling (ECC)

Case Study

Discussion

Statistical postprocessing of NWP ensembles

despite their undisputed success, **NWP ensembles** are subject to **model biases**, and typically they show a **lack of calibration**

- in typical experience, rank histograms are U-shaped, indicating underdispersion

thus, some form of **statistical postprocessing** is required to generate **calibrated** and **sharp** predictive distributions from NWP model output

- idea is to **exploit structure** in past forecast-observation pairs to **correct** for **systematic deficiencies** in the model output
- approach depends on the availability of a suitable **training set**, consisting of past forecast-observation pairs
- typically, a **rolling training period** of 20-40 days is used to estimate statistical parameters
- training sets can be usefully **augmented** by **reforecast data**
- simple bias correction doesn't suffice — e.g., in the case of precipitation, additive terms won't work

EMOS/NR and BMA

two general approaches to the **statistical postprocessing** of **NWP ensemble output** have emerged, namely

- **ensemble model output statistics (EMOS)** or **nonhomogeneous regression (NR)**, which fits a single, parametric predictive distribution using summary statistics from the ensemble (Gneiting et al. 2005)

$$y \,|\, x_1, \ldots, x_M \sim f(y \,|\, x_1, \ldots, x_M)$$

- **Bayesian model averaging (BMA)**, which fits a mixture density as predictive distribution, where each ensemble member is associated with a kernel function (Raftery et al. 2005)

$$y \,|\, x_1, \ldots, x_M \sim \sum_{m=1}^{M} w_m \, g(y \,|\, x_m)$$

in our experience, the two approaches yield similar predictive performance, with **BMA** being the more **flexible** and **EMOS/NR** the more **parsimonious** method

EMOS/NR and BMA for temperature

consider an **ensemble forecast**, $x_1, \ldots, x_M$, for **temperature**, y, at a given time and location

EMOS/NR employs a **single Gaussian** predictive density, in that

$$y \,|\, x_1, \ldots, x_M \;\sim\; \mathcal{N}(a_0 + a_1 x_1 + \cdots + a_M x_M,\, b_0 + b_1 s^2)$$

with location parameters b_0 and $b_1, \ldots, b_M$, and spread parameters c_0 and c_1, where s^2 is the ensemble variance

BMA employs Gaussian kernels with a linearly bias-corrected mean, i.e., the BMA predictive density is the **Gaussian mixture**

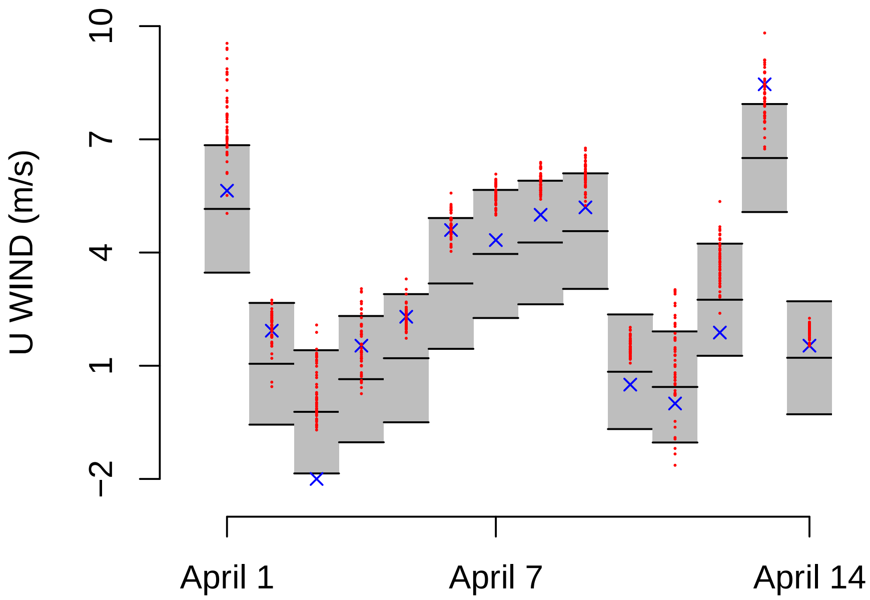
$$y \,|\, x_1, \ldots, x_M \;\sim\; \sum_{m\,=\,1}^{M} w_m \, \mathcal{N}(c_{0m} + c_{1m} x_m,\, \sigma_m^2)$$

with BMA weights $w_1, \ldots, w_M$, bias parameters $c_{01}, \ldots, c_{0M}$ and $c_{11}, \ldots, c_{1M}$, and spread parameters $\sigma_1^2, \ldots, \sigma_M^2$

for ensembles with **exchangeable** members, such as the ECMWF system, member specific statistical parameters are constrained to be **equal**; e.g., $a_1 = \cdots = a_M$ or $w_1 = \cdots = w_M = \frac{1}{M}$

Ensemble model output statistics (EMOS) or nonhomogeneous regression (NR)

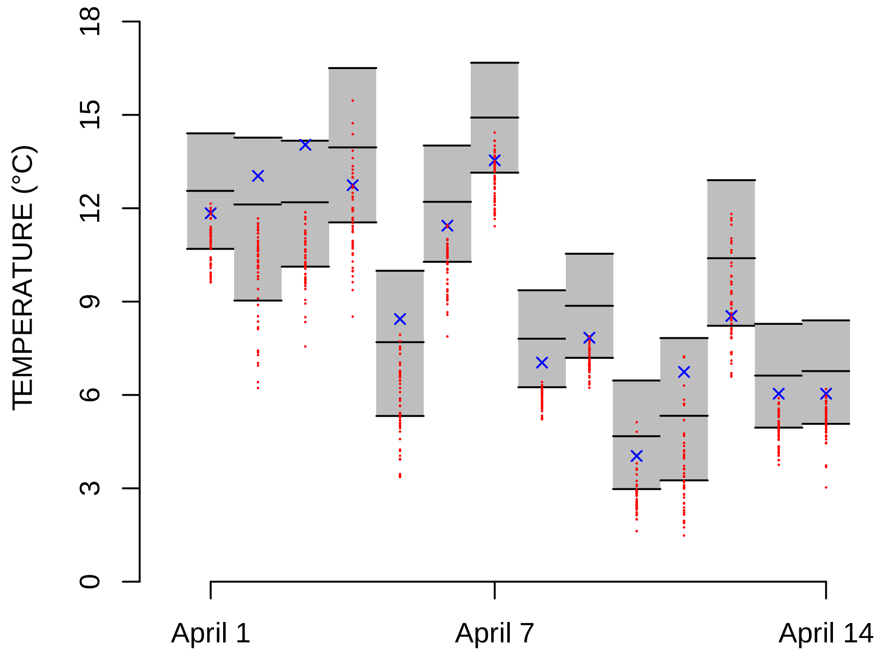
Weather Quantity	Range	Distribution (f)
Temperature	$y \in \mathbb{R}$	Normal
Pressure	$y \in \mathbb{R}$	Normal
Precipitation amount	$y^{1/2} \in \mathbb{R}^+$	Truncated logistic
	$y \in \mathbb{R}^+$	Generalized extreme value
Wind components	$y \in \mathbb{R}$	Normal
Wind speed	$y \in \mathbb{R}^+$	Truncated normal



24-hour ahead u-wind at Hamburg, valid April 1–14, 2011 at 00 UTC, 24-hour lead time

Bayesian model averaging (BMA)

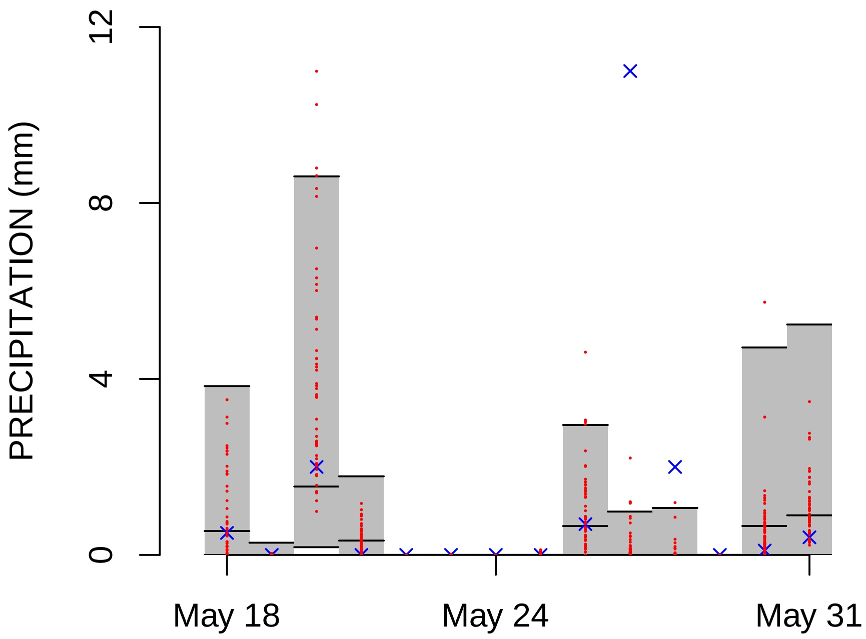
Variable	Range	Kernel (g)	Mean	Variance
Temperature	$y \in \mathbb{R}$	Normal	$c_{0m} + c_{1m} x_m$	σ_m^2
Pressure	$y \in \mathbb{R}$	Normal	$c_{0m} + c_{1m} x_m$	σ_m^2
Precipitation accumulation	$y^{1/3} \in \mathbb{R}^+$	Gamma	$c_{0m} + c_{1m} x_m$	$d_{0m} + d_{1m} x_m$
Wind components	$y \in \mathbb{R}$	Normal	$c_{0m} + c_{1m} x_m$	σ_m^2
Wind speed	$y \in \mathbb{R}^+$	Gamma	$c_{0m} + c_{1m} x_m$	$d_{0m} + d_{1m} x_m$
Visibility	$y \in [0, 1]$	Beta	$c_{0m} + c_{1m} x_m^{1/2}$	$d_{0m} + d_{1m} x_m^{1/2}$



temperature in Berlin valid April 1–14, 2011 at 00 UTC, 48-hour lead time

Bayesian model averaging (BMA)

Variable	Range	Kernel (f)	Mean	Variance
Precipitation accumulation	$y^{1/3} \in \mathbb{R}^+$	Gamma	$c_{0m} + c_{1m} x_m$	$d_{0m} + d_{1m} x_m$



precipitation accumulation in Frankfurt, valid May 18–31, 2011, 24-hour lead time, rolling 30-day training period

Introduction: Probabilistic Forecasts

Statistical Postprocessing of NWP Ensembles: EMOS/NR and BMA

Accounting for Dependencies: Ensemble Copula Coupling (ECC)

Case Study

Discussion

Accounting for dependencies

EMOS/NR and BMA apply to any **single weather variable** at any **single location** and any **single look-ahead time**

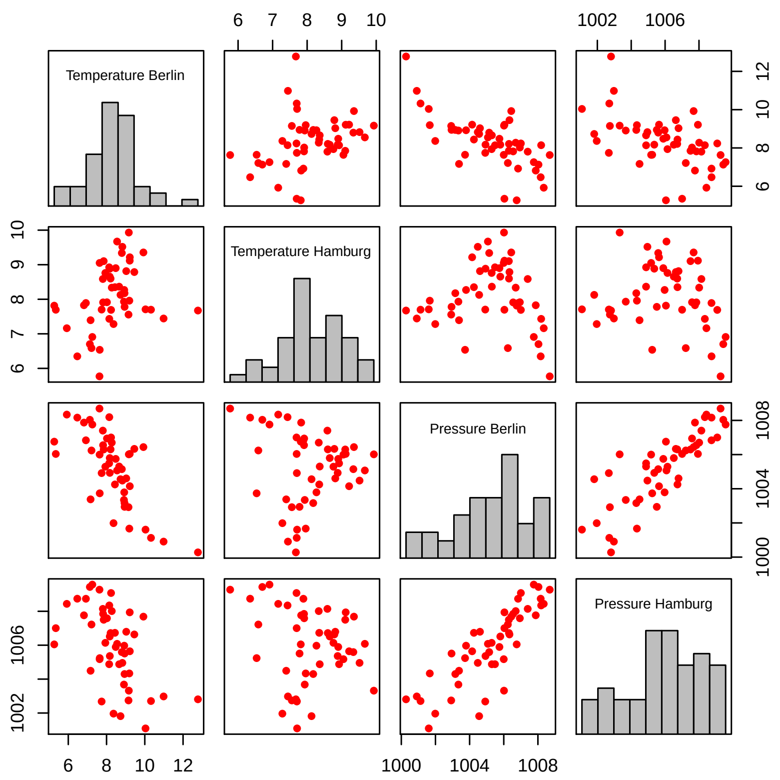
however, individually postprocessed distributions fail to account for **multivariate dependence** structures

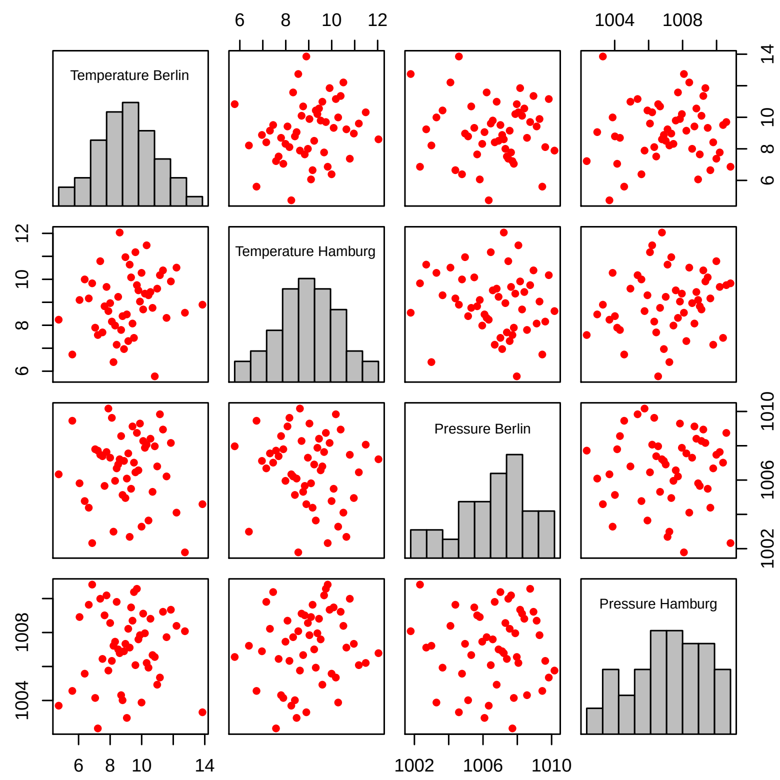
the most pressing need now is to develop postprocessing techniques that yield **physically realistic** probabilistic forecasts of **spatio-temporal weather trajectories** for **multiple weather variables** at **multiple locations** and **multiple look-ahead times**

key applications include **air traffic control**, **ship routeing** and **hydrologic predictions**

Example

illustration: 24-hour **ECMWF** ensemble forecast of surface temperature and pressure at Berlin and Hamburg valid May 27, 2010 **before** and **after** BMA **postprocessing**





Sklar's theorem

EMOS/NR and BMA apply to any **single weather variable** at any **single location** and any **single look-ahead time**

yielding a **univariate** or **marginal** predictive **cumulative distribution function** (**CDF**), F_l, for any given univariate weather quantity Y_l

with each multi-index $l = (i, j, k)$ referring to **weather variable** i, **location** j and **look-ahead time** k

we seek a **physically realistic** and consistent **multivariate** or **joint** predictive **CDF**, F, with margin F_l for each $l = 1, \ldots, L$

Sklar's theorem (1959): every multivariate CDF F with margins $F_1, \ldots, F_L$ can be written as

$$F(y_1, \ldots, y_L) = C(F_1(y_1), \ldots, F_L(y_L))$$

where $C : [0,1]^L \to [0,1]$ is a **copula**, i.e., a multivariate CDF with standard **uniform margins**

Copula approaches

in order to issue **physically realistic** and **consistent** probabilistic forecasts of **spatio-temporal weather trajectories**

it remains to specify and fit a suitable **copula** $C : [0,1]^L \to [0,1]$

if L is small, or if specific structure can be exploited, **parametric** families of copulas work well

- Gel et al. (2004), Berrocal et al. (2007), Pinson et al. (2009), Schuhen et al. (2012) and Möller et al. (2013) use **Gaussian copulas**

- parametric or semi-parametric alternatives include **elliptical**, **Archimedean**, hierarchical Archimedean and **pair** copulas

if L is huge and no specific structure can be exploited, we need to resort to **non-parametric** approaches, based on **empirical copulas**, with the **Schaake shuffle** (Clark et al. 2004) and **ensemble copula coupling (ECC)** being particularly attractive options

Ensemble copula coupling (ECC)

given an **NWP ensemble** of size M for the weather variables Y_l, where $l = 1, \ldots, L$, **ensemble copula coupling (ECC)** proceeds in three steps

univariate postprocessing: for each $l = 1, \ldots, L$, apply **EMOS/NR** or **BMA** to obtain a postprocessed **predictive CDF**, F_l

quantization: for each $l = 1, \ldots, L$, obtain a discrete **sample** of size M from F_l, namely

$$\tilde{x}_m = F_l^{-1}\left(\frac{m}{M+1}\right), \qquad m = 1, \ldots, M$$

ensemble reordering: take the function $C : [0,1]^L \to [0,1]$ in Sklar's theorem to be the **empirical copula** of the raw ensemble, i.e., arrange the postprocessed values in the same **rank order** as the raw ensemble values

Ensemble copula coupling (ECC)

the method is **implicit** or **explicit** in scattered **recent work**, including that of Bremnes (2007), Krzysztofowicz and Toth (2008), Flowerdew (2012), Pinson (2011), Roulin and Vannitsem (2012) and Schuhen, Thorarinsdottir and Gneiting (2012)

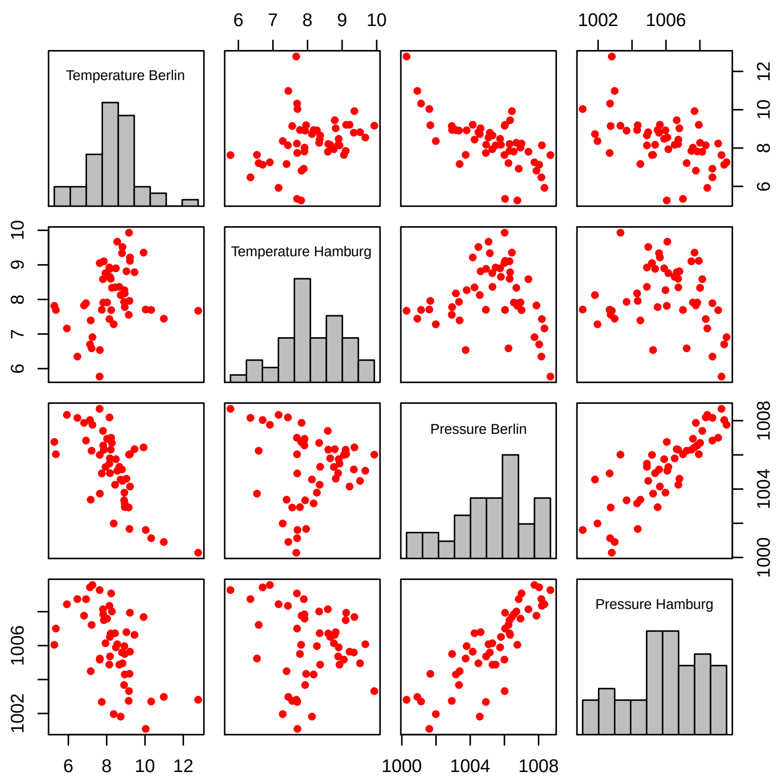
Flowerdew (2012, p. 15) explains the **idea** colorfully:

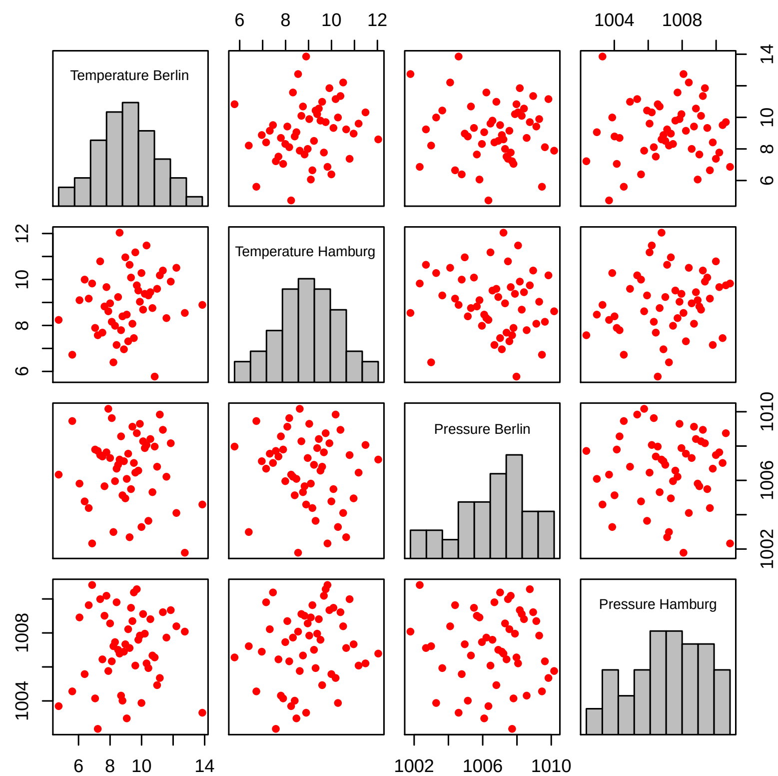
> The key to preserving spatial, temporal and inter-variable structure is how this set of values is distributed between ensemble members. One can always construct ensemble members by sampling from the calibrated PDF, but this alone would produce spatially noisy fields lacking the correct correlations. Instead, *the values are assigned to ensemble members in the same order as the values from the raw ensemble: the member with the locally highest rainfall remains locally highest, but with a calibrated rainfall magnitude.*

an up-to-date **review** of ECC type techniques is provided by Schefzik, Thorarinsdottir and Gneiting (2013)

Ensemble copula coupling (ECC)

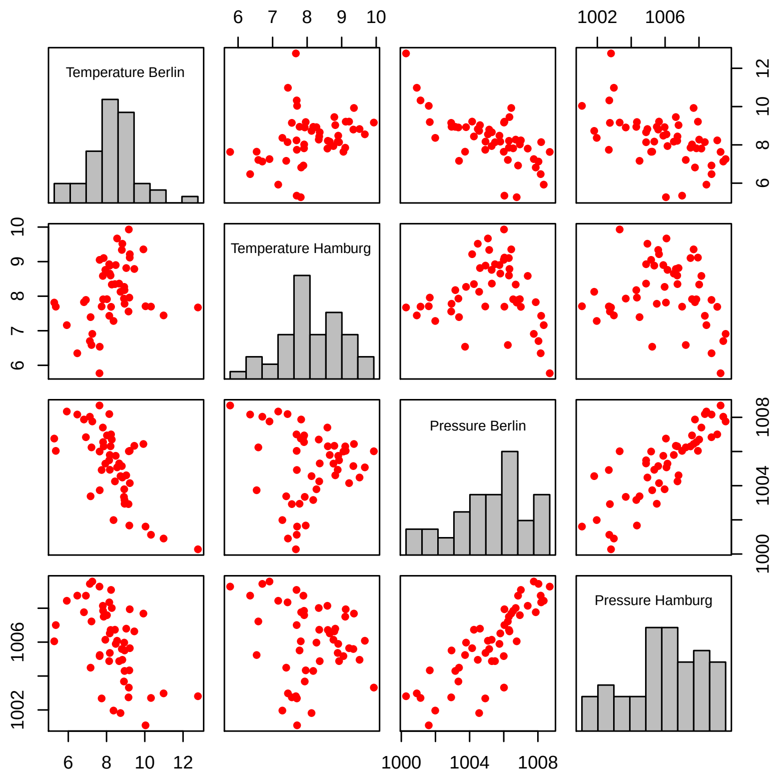
illustration: 24-hour **ECMWF** ensemble forecast of surface temperature and pressure at Berlin and Hamburg valid May 27, 2010 before and after postprocessing with **BMA**

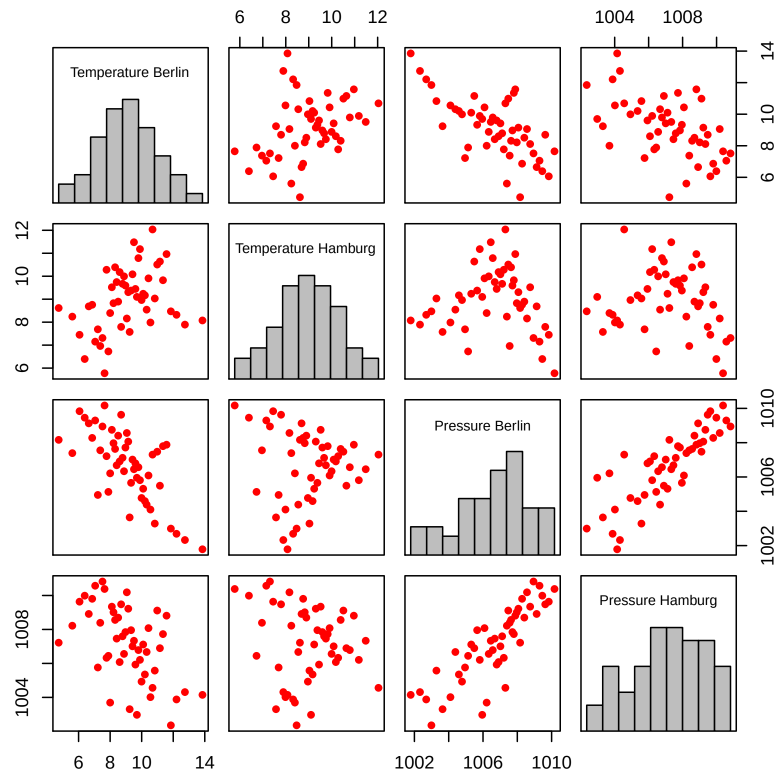




Ensemble copula coupling (ECC)

illustration: 24-hour **ECMWF** ensemble forecast of surface temperature and pressure at Berlin and Hamburg valid May 27, 2010 before and after postprocessing with **BMA + ECC**





Introduction: Probabilistic Forecasts

Statistical Postprocessing of NWP Ensembles:
EMOS/NR and BMA

Accounting for Dependencies:
Ensemble Copula Coupling (ECC)

Case Study

Discussion

Case study: ECMWF ensemble

we consider **statistical postprocessing** for the **ECMWF**'s operational **50-member** system

using the **BMA**, **EMOS/NR** and **ECC** techniques for surface **temperature**, **pressure**, **precipitation**, and the u **wind** component

at the airports in **Berlin**-Tegel, **Frankfurt** and **Hamburg**, Germany

at lead times of 24 and 48 hours

the statistical parameters for BMA and EMOS/NR are **estimated** on a rolling 30-day **training period**, with the member specific parameters constrained to be equal

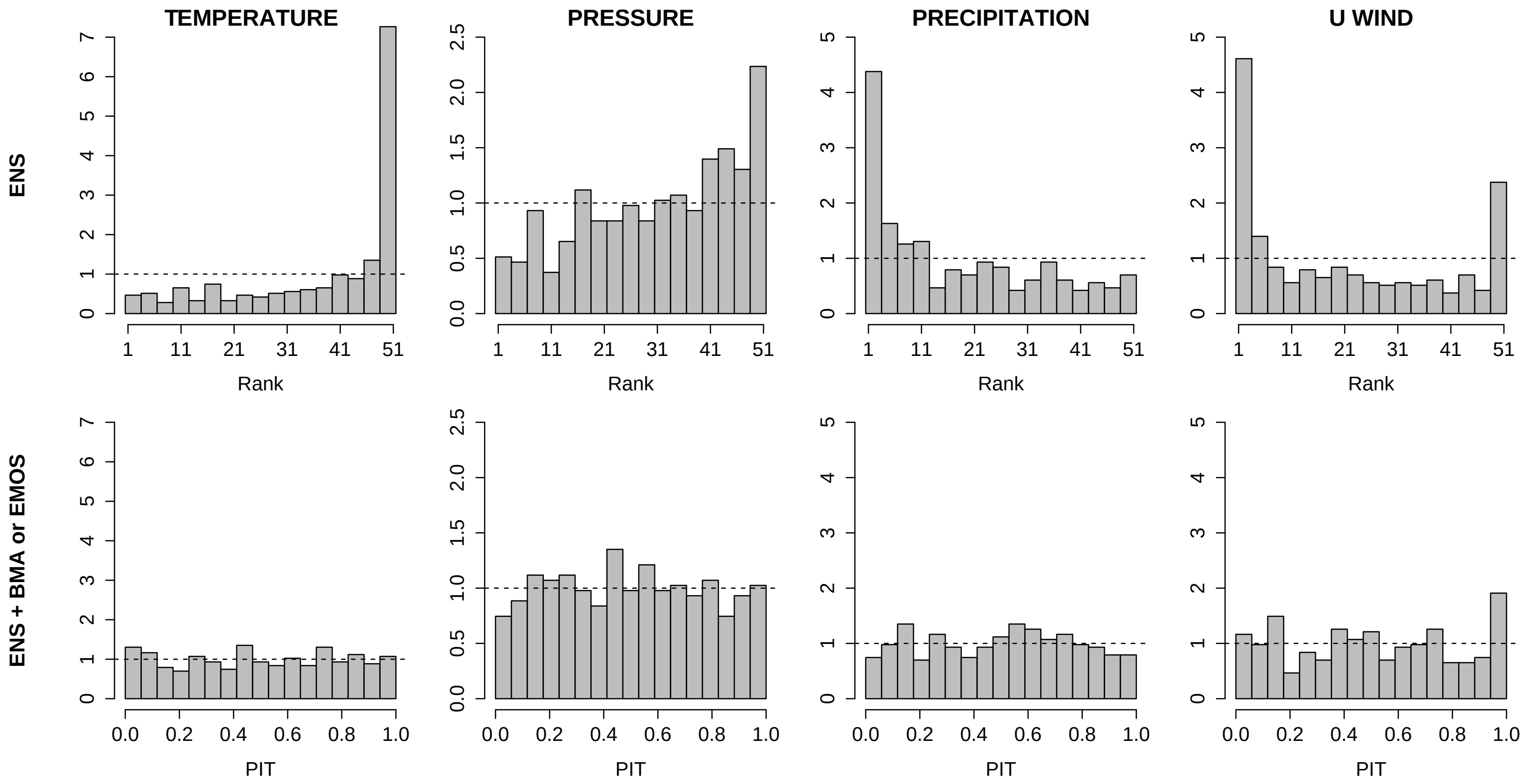
the **test period** ranges from May 1, 2010 through April 30, 2011

for details and further results see Schefzik, Thorarinsdottir and Gneiting (2013)

Results: Univariate weather quantities

			CRPS			AE		
			Berlin	Frankfurt	Hambg	Berlin	Frankfurt	Hambg
Temp. (°C)	24	ENS	1.21	1.23	1.01	1.50	1.53	1.26
		ENS+BMA	0.90	0.88	0.79	1.27	1.23	1.10
	48	ENS	1.25	1.26	1.06	1.62	1.63	1.39
		ENS+BMA	0.99	0.97	0.92	1.41	1.33	1.31
Pressure (hPa)	24	ENS	0.54	0.55	0.51	0.75	0.75	0.71
		ENS+BMA	0.43	0.43	0.39	0.62	0.61	0.54
	48	ENS	0.80	0.78	0.77	1.12	1.08	1.09
		ENS+BMA	0.77	0.74	0.73	1.08	1.03	1.03
Precip. (mm)	24	ENS	0.25	0.41	0.31	0.32	0.51	0.39
		ENS+BMA	0.23	0.40	0.37	0.30	0.49	0.44
	48	ENS	0.26	0.41	0.36	0.34	0.50	0.45
		ENS+BMA	0.26	0.43	0.39	0.32	0.52	0.48
u-Wind (ms^{-1})	24	ENS	0.83	0.96	0.89	1.06	1.19	1.11
		ENS+EMOS	0.70	0.60	0.68	0.98	0.81	0.96
	48	ENS	0.82	0.89	0.88	1.09	1.15	1.18
		ENS+EMOS	0.75	0.62	0.75	1.05	0.83	1.04

Results: Univariate weather quantities

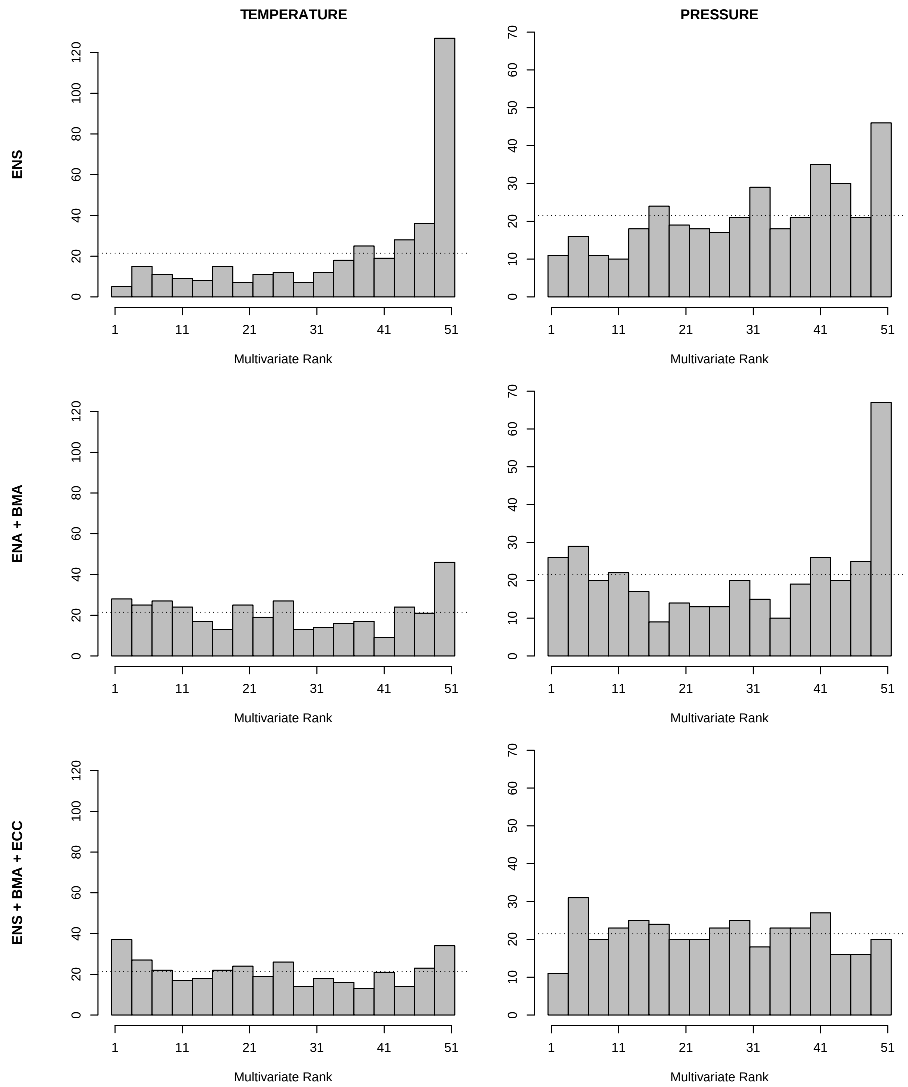


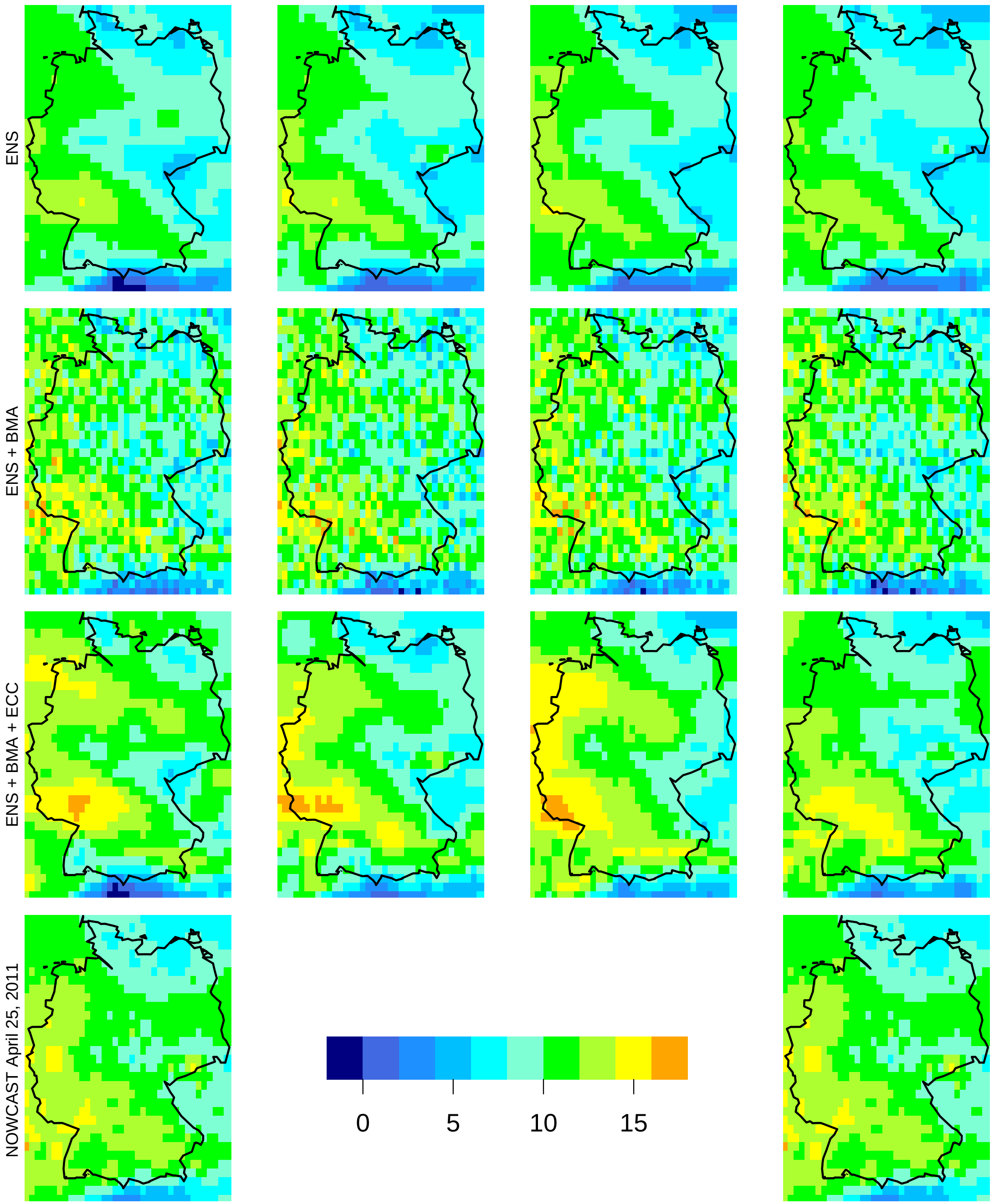
Frankfurt, 48-hour lead time

Results: Multivariate weather quantities

ensemble forecasts of **temperature** or **pressure** at all **three sites simultaneously**, at 48-hour lead time

Energy score	Temp (°C)	Pressure (hPa)
ENS	2.34	1.48
ENS+BMA	1.93	1.48
ENS+BMA+ECC	1.92	1.43




ENS
ENS + BMA
ENS + BMA + ECC
NOWCAST April 25, 2011
0
5
10
15

Introduction: Probabilistic Forecasts

Statistical Postprocessing of NWP Ensembles:
EMOS/NR and BMA

Accounting for Dependencies:
Ensemble Copula Coupling (ECC)

Case Study

Discussion

Discussion

statistical postprocessing techniques such as **ensemble model output statistics (EMOS/NR)** and **Bayesian model averaging (BMA)** can yield **gains** in lead time **of several days** for **surface weather** variables

statistical parameters need to be **estimated** from **training data**, which can be usefully augmented by using **reforecasts** (Hamill, Hagedorn and Whitaker 2008)

future research and development is expected to focus on ensemble postprocessing techniques for **multiple weather variables** at **multiple locations** and **multiple look-ahead times** simultaneously

with the goal of generating **calibrated** and **sharp** ensemble forecasts of **spatio-temporal weather scenarios**

in this setting, the **ensemble copula coupling (ECC)** method can serve as a **benchmark**

Selected references

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

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133, 1155–1174.

Schefzik, R., Thorarinsdottir, T. L. and Gneiting, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, in press.