Quantifying Uncertainty in Complex Simulation Models Using Ensemble Copula Coupling

Tilmann Gneiting

Heidelberg Institute for Theoretical Studies (HITS)
Karlsruhe Institute of Technology (KIT)

Innsbruck, May 21, 2014

Uncertainty Quantification
and Numerical Weather Prediction (NWP)

Statistical Postprocessing of NWP Ensembles

Ensemble Copula Coupling

Case Study

Discussion

Uncertainty quantification

decision-making frequently depends on the output of complex mathematical models or simulators, such as in weather and climate prediction, flood risk assessment, air quality, groundwater contamination, disease spread, . . .

increased recognition of the need for uncertainty quantification, as evidenced by the creation of SIAM and ASA interest groups, along with new journals, such as the joint SIAM/ASA Journal on Uncertainty Quantification

Nick Trefethen, SIAM News May 2012: "An answer that used to be a single number may now be a statistical distribution"

informed decision making depends on full joint probability distributions for insufficiently known quantities

typically, but not necessarily time forward, such as in numerical weather prediction

Numerical weather prediction (NWP)

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

numerical weather prediction (NWP) is based on complex mathematical models that represent the physics of the atmosphere

system of six partial differential equations (conservation of momentum, mass, energy and entropy, and equation of state) in six variables (two velocity components, density, pressure, temperature, humidity)

equations are discretized and run forward in time to obtain deterministic forecasts of future states of the atmosphere

data assimilation systems provide initial conditions that describe the current state of the atmosphere on a 3d grid: millions of inputs

Numerical weather prediction (NWP)

through the 1980s, weather forecasting had traditionally been viewed as a deterministic problem, despite two major sources of uncertainty, namely

- initial conditions: incomplete network of observations, measurement error, shortcomings in data assimilation, ...
- model formulation: incomplete knowledge of physical processes (e.g., inaccurate parameterizations of sub grid-scale processes), incomplete and inaccurate numerical schemes, ...

major shift of paradigms since the early 1990s, as expressed by Tim Palmer (2000):

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

probabilistic weather forecasts quantify uncertainty using predictive probability density functions (PDFs), as opposed to single-valued point forecasts

What is a good probabilistic forecast?
Maximizing sharpness subject to calibration

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

calibration

refers to the statistical compatibility between the predictive PDFs and the realizing observations

joint property of the forecasts and the observations that can be assessed using the probability integral transform (PIT)

sharpness

refers to the spread of the predictive PDFs

property of the forecasts only

proper scoring rules such as the logarithmic score or the continuous ranked probability score (Gneiting and Raftery 2007) allow for a joint assessment of calibration and sharpness

Proper scoring rules

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 CDF 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)

Proper scoring rules

the most popular example is the logarithmic score,

$$\mathrm{s}(f, y) = -\log f(y),$$

i.e., the negative of the predictive PDF, f, evaluated at the realizing observation, y

a local score of integer order k depends on the predictive distribution via $f(y), \ldots, f^{(k)}(y)$ only

the logarithmic score is local of order $k = 0$

local scores of general order have been studied by Parry, Dawid and Lauritzen (2012) and Ehm and Gneiting (2012), with the Hyvärinen score,

$$\mathrm{s}(f, y) = 2 \, \frac{f''(y)}{f(y)} - \left(\frac{f'(y)}{f(y)} \right)^2,$$

being an example of order $k = 2$

Proper scoring rules

in applications, our favorite score is the continuous ranked probability score,

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

where X and X' are independent random variables with cumulative distribution function F

the continuous ranked probability score is reported in the same unit as the observations and generalizes the absolute error, to which it reduces in the case of a point forecast

provides a direct way of comparing point forecasts and probabilistic forecasts

the kernel score representation allows for a direct multivariate analogue, the energy score

NWP ensembles

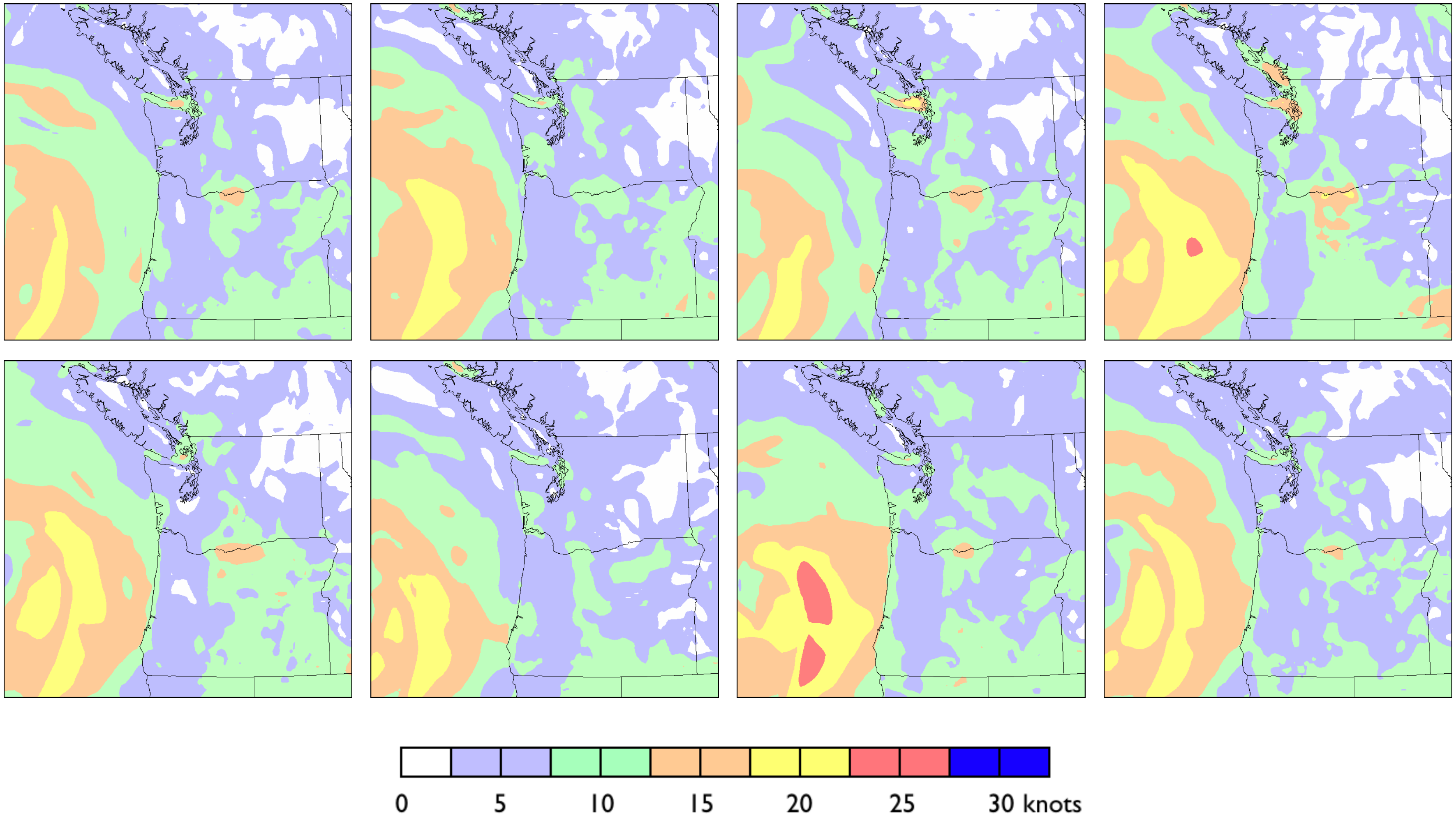
preferred approach to probabilistic weather prediction is based on ensembles of NWP forecasts:

- each ensemble member is a single-valued, deterministic forecast using an NWP model
- the forecasts differ from each other with respect to the two major sources of uncertainty: initial conditions and/or model formulation

global ensemble prediction systems have been operational at NCEP and ECMWF since December 1992

limited area systems such as the University of Washington Mesoscale Ensemble (UWME) or the COSMO-DE system run by the German Weather Service (DWD) operate at lead times up to three days

University of Washington Mesoscale Ensemble (UWME)



48-hour ahead UWME forecast of maximum wind speed valid August 7, 2003

Uncertainty Quantification
and Numerical Weather Prediction (NWP)

Statistical Postprocessing of NWP Ensembles

Ensemble Copula Coupling

Case Study

Discussion

Statistical postprocessing: EMOS/NR and BMA

NWP ensembles are subject to model biases and typically they show a lack of calibration

thus, some form of statistical postprocessing is required in order to properly quantify uncertainty, to generate calibrated and sharp predictive PDFs

we have developed two general approaches to the statistical postprocessing of NWP ensembles:

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

- Bayesian model averaging (BMA), which fits a mixture density as predictive PDF, where each ensemble member is associated with a kernel function, with a weight that reflects the member's relative skill (Raftery et al. 2005)

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

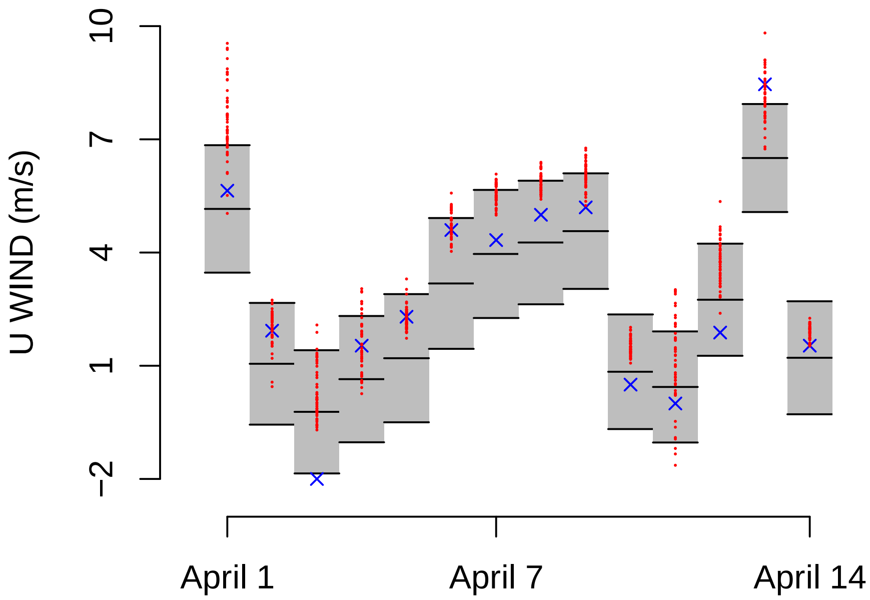
BMA employs Gaussian kernels with a linearly bias-corrected mean, i.e., the BMA predictive density is the Gaussian mixture

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

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

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

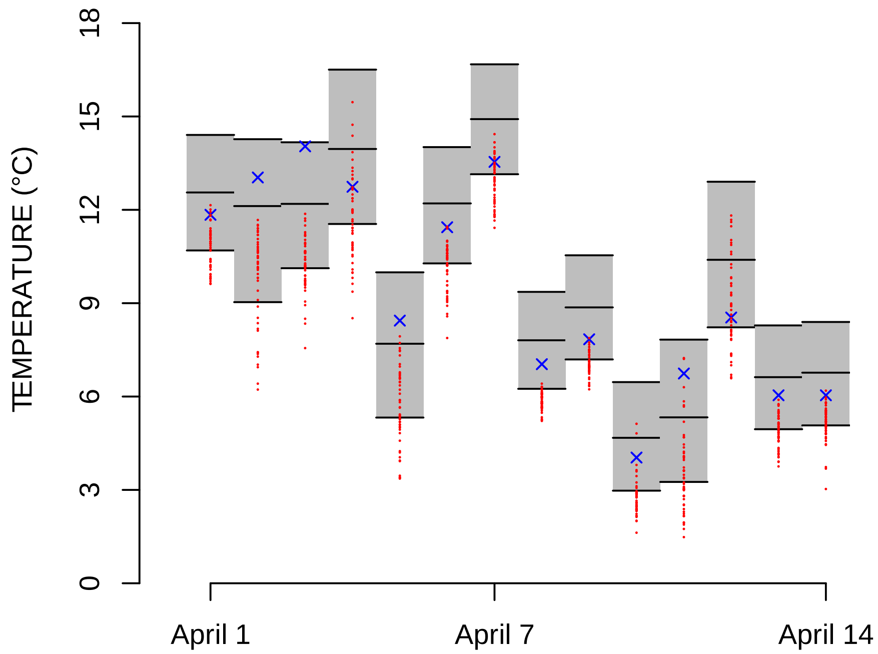
Weather Quantity	Range	Distribution (f)
Temperature	$y \in \mathbb{R}$	Normal
Pressure	$y \in \mathbb{R}$	Normal
Precipitation amount	$y^{1/2} \in \mathbb{R}^+$	Truncated logistic
	$y \in \mathbb{R}^+$	Generalized extreme value
Wind components	$y \in \mathbb{R}$	Normal
Wind speed	$y \in \mathbb{R}^+$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



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

Bayesian model averaging (BMA)

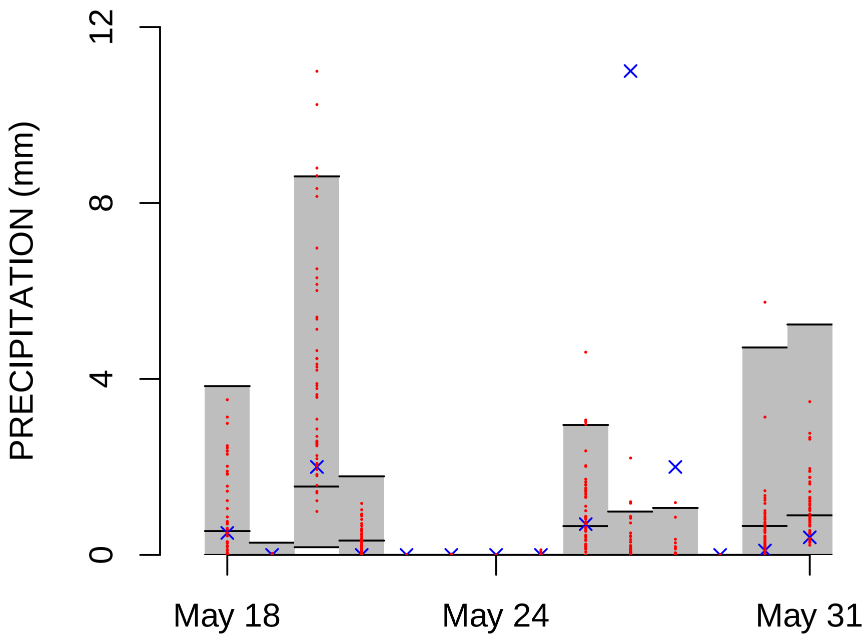
Variable	Range	Kernel (g)	Mean	Variance
Temperature	$y \in \mathbb{R}$	Normal	$c_{0m} + c_{1m} x_m$	σ_m^2
Pressure	$y \in \mathbb{R}$	Normal	$c_{0m} + c_{1m} x_m$	σ_m^2
Precipitation accumulation	$y^{1/3} \in \mathbb{R}^+$	Gamma	$c_{0m} + c_{1m} x_m$	$d_{0m} + d_{1m} x_m$
Wind components	$y \in \mathbb{R}$	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, rolling 30-day training period

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

Uncertainty Quantification
and Numerical Weather Prediction (NWP)

Statistical Postprocessing of NWP Ensembles

Ensemble Copula Coupling

Case Study

Discussion

Ensemble copula coupling (ECC)

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

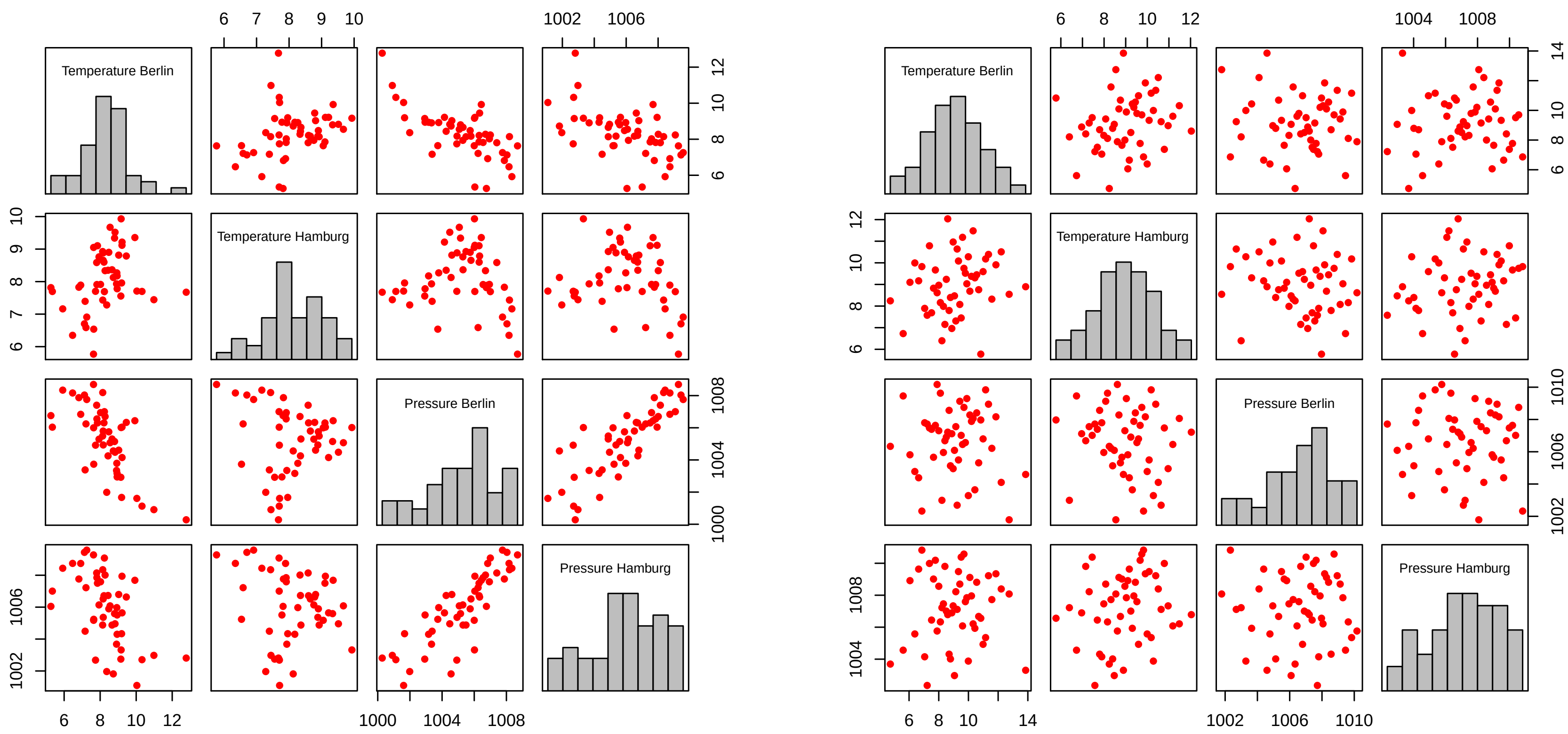
however, individually postprocessed distributions fail to account for multivariate dependence structures

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, hydrologic predictions and renewable energy management

Example

illustration: 24-hour 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

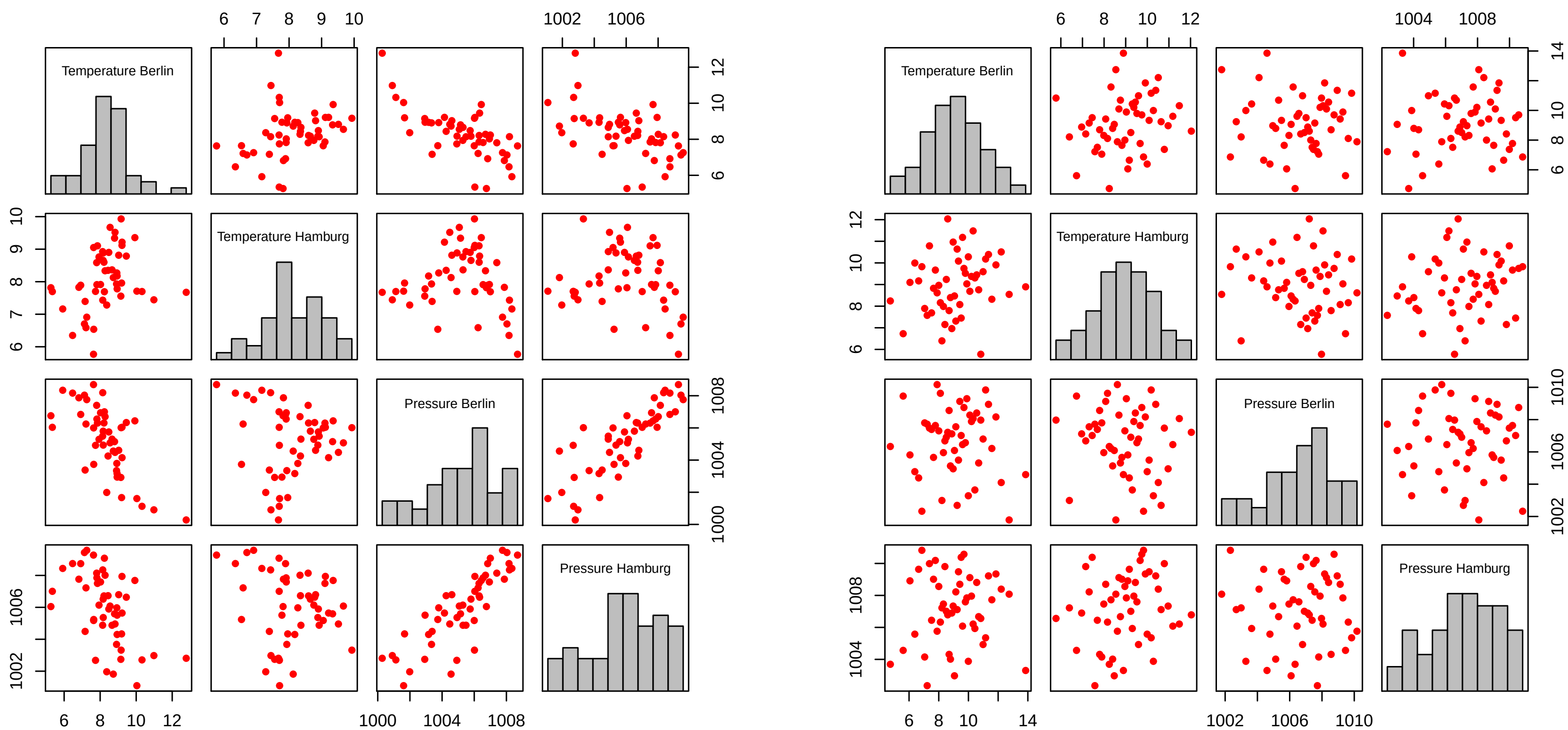
$$\tilde{x}_m = F_l^{-1}\left(\frac{m}{M+1}\right), \qquad m = 1, \ldots, M$$

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

implicit in scattered recent work in the meteorological and climatological literatures

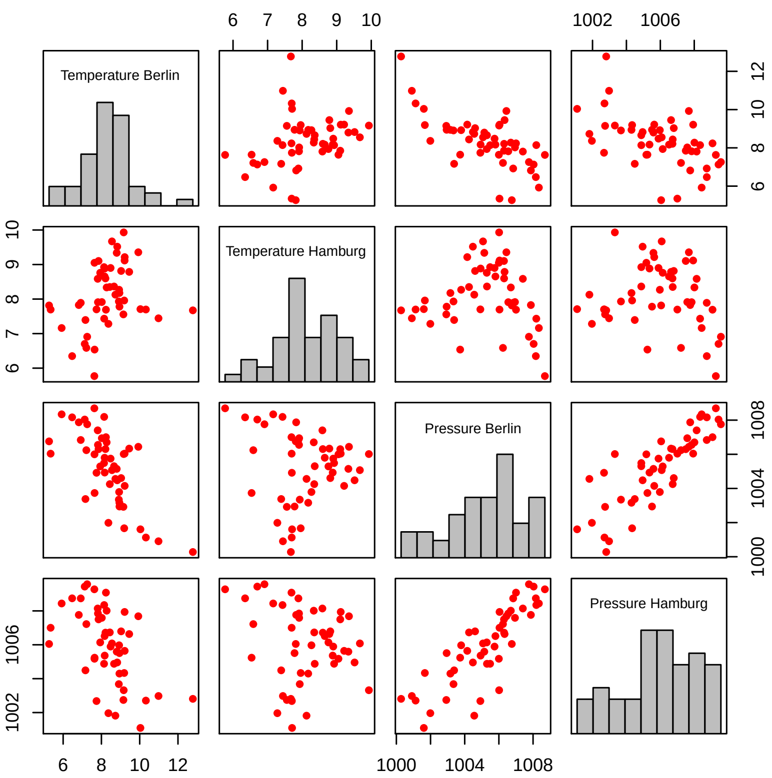
Ensemble copula coupling (ECC)

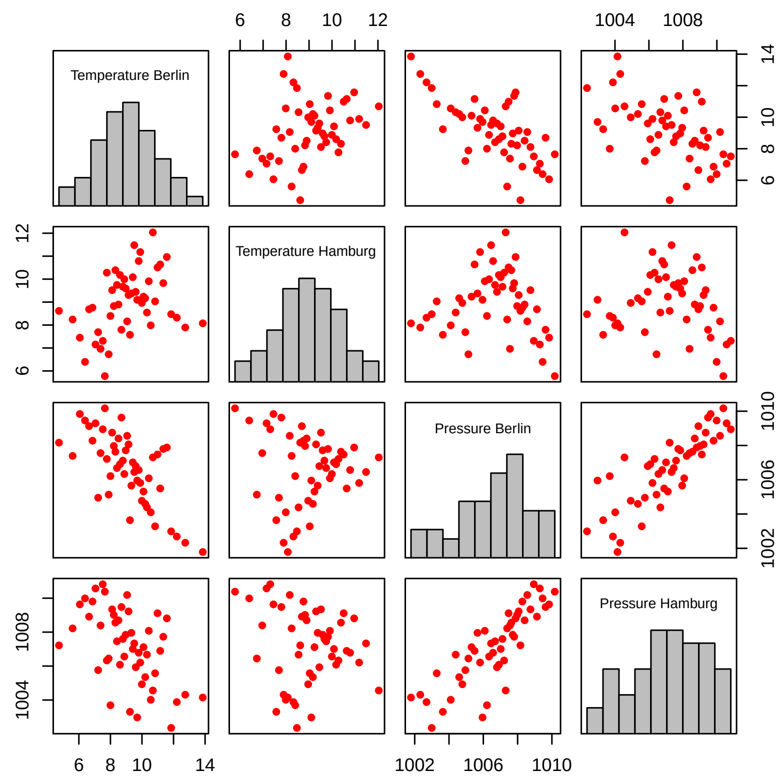
illustration: 24-hour ensemble forecast of surface temperature and pressure at Berlin and Hamburg valid May 27, 2010 before and after postprocessing with BMA



Ensemble copula coupling (ECC)

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





Uncertainty Quantification
and Numerical Weather Prediction (NWP)

Statistical Postprocessing of NWP Ensembles

Ensemble Copula Coupling

Case Study

Discussion

ECMWF ensemble

statistical postprocessing for the European Centre for Medium-Range Weather Forecasts (ECMWF)'s 50-member ensemble

BMA, EMOS/NR and ECC applied to 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

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

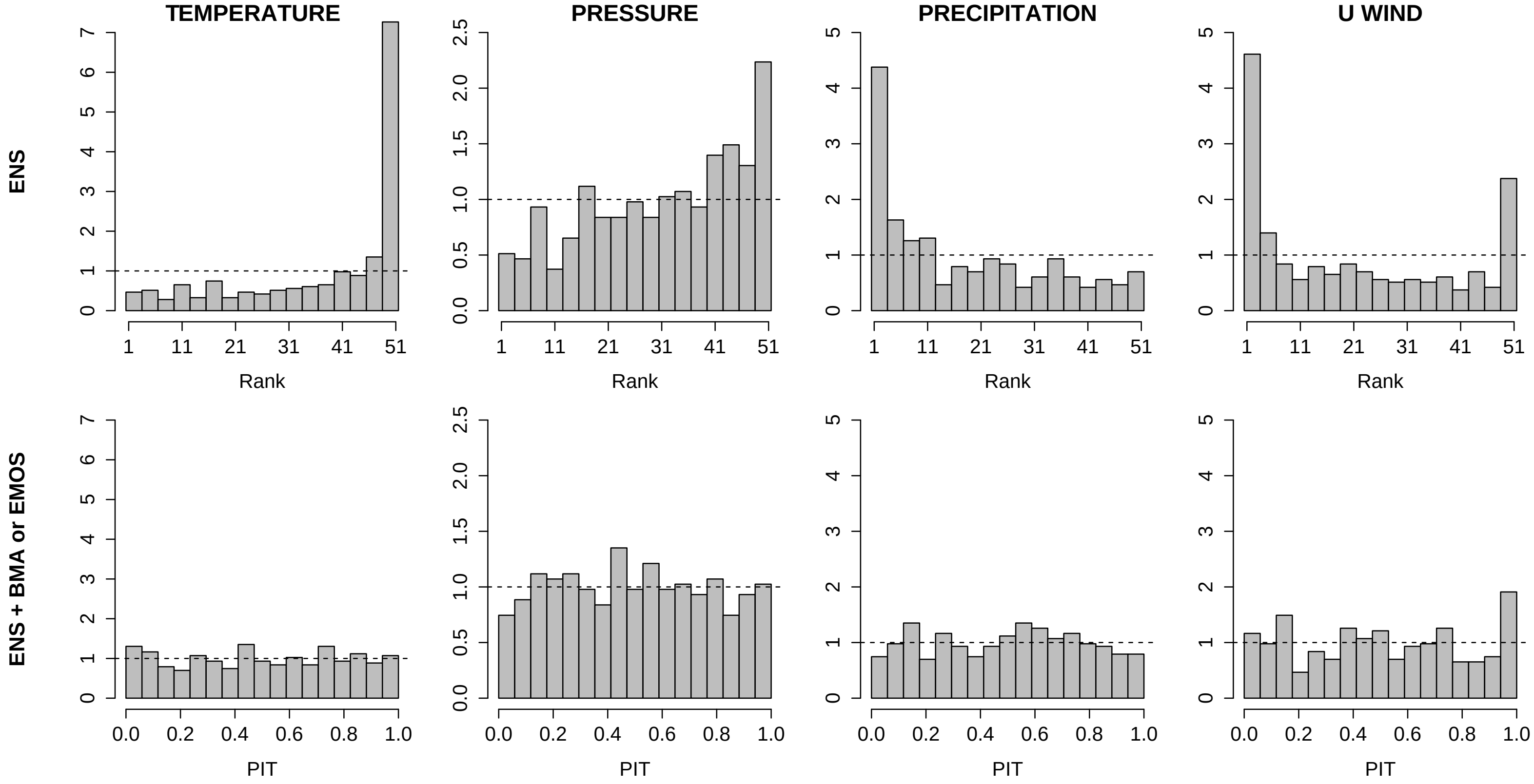
the test period ranges from May 1, 2010 through April 30, 2011

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

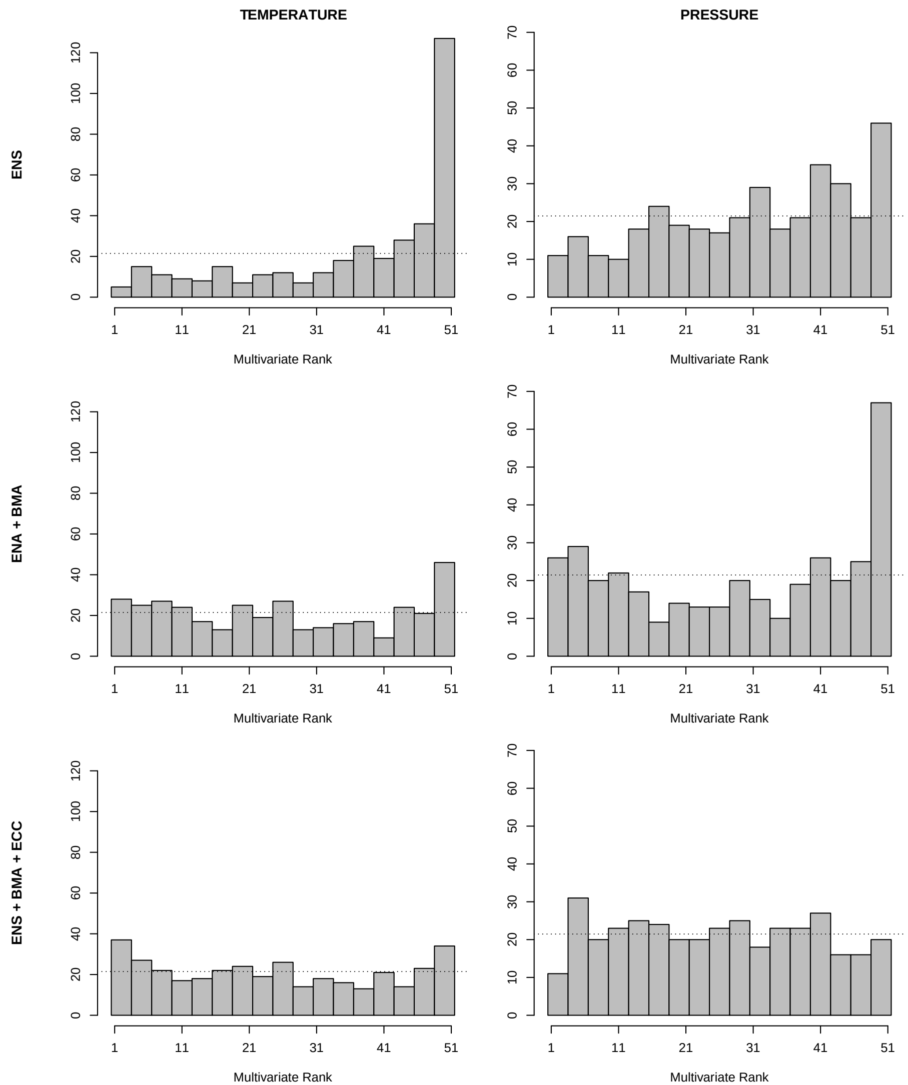


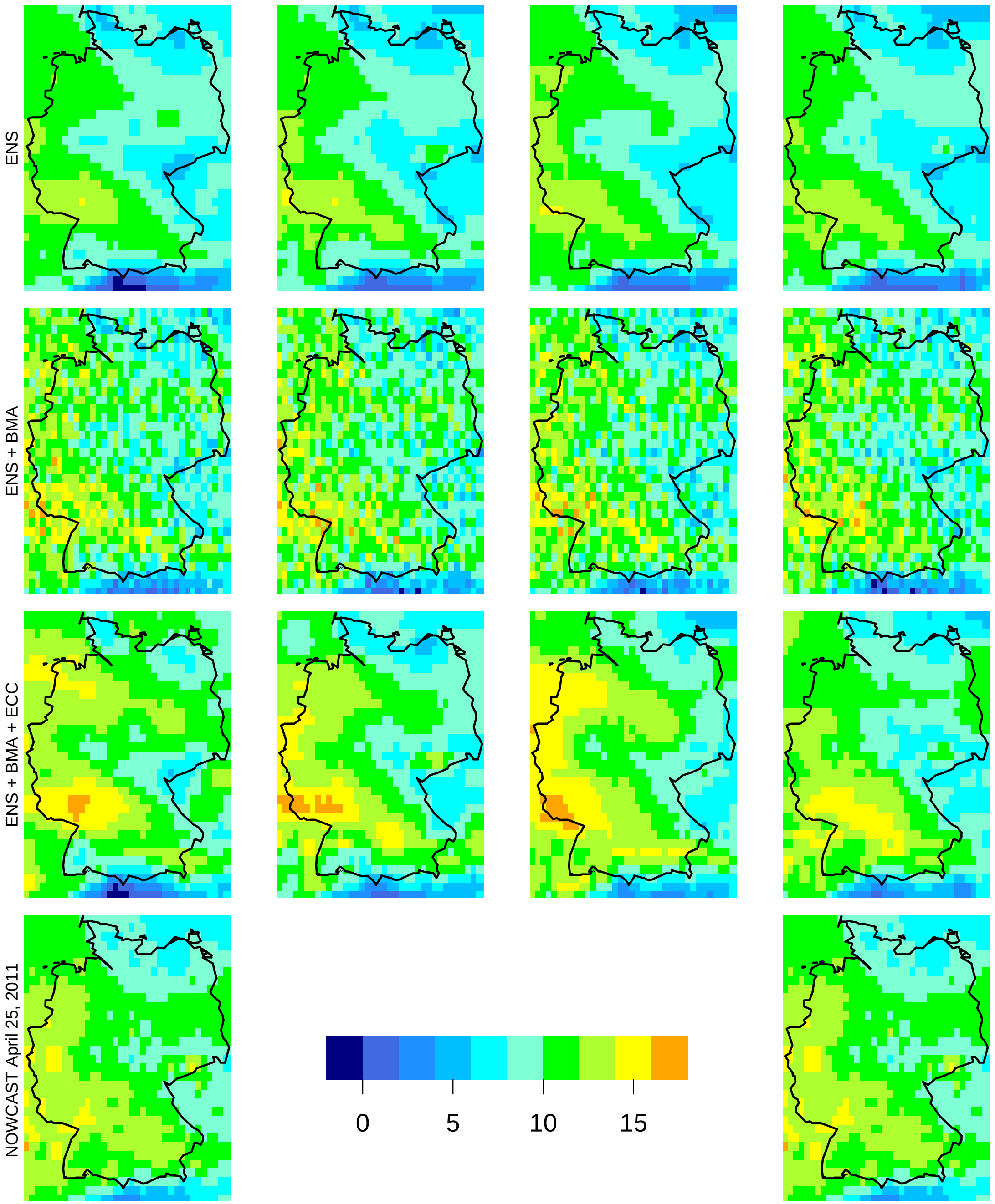
Frankfurt, 48-hour lead time

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

Uncertainty Quantification
and Numerical Weather Prediction (NWP)

Statistical Postprocessing of NWP Ensembles

Ensemble Copula Coupling

Case Study

Discussion

The future of 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.

Uncertainty quantification

strong recognition of the need for uncertainty quantification in complex simulators

the term ensemble copula coupling (ECC) refers to a general four-stage approach to uncertainty quantification

1. generate a forecast ensemble, using multiple runs of the simulator with perturbed initial conditions or model parameters

2. apply statistical postprocessing techniques to obtain calibrated and sharp univariate predictive CDFs

3. sample from the postprocessed predictive CDFs

4. merge the discrete univariate margins using the empirical copula of the forecast ensemble

ECC approaches combine analytic, numerical and statistical modeling and are likely to be very broadly applicable

Selected references

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

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102, 359–378.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106, 746–762.

Schefzik, R., Thorarinsdottir, T. L. and Gneiting, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28, 616–640.