Multivariate extremes in ensemble forecasting

Hans Wackernagel

MINES ParisTech, Fontainebleau (France) NERSC, Bergen (Norway)

10th International EnKF Workshop
Flåm, Norway, June 8-10, 2015

http://hans.wackernagel.free.fr

Anamorphosis and dependence structure

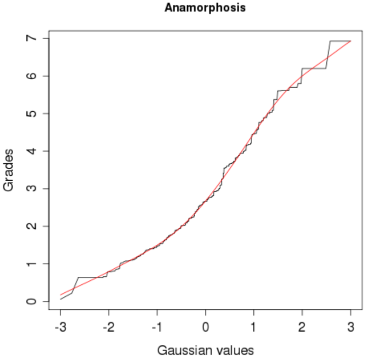
- Ensemble Kalman filter requires an assumption of Gaussian distribution at the analysis stage.
- Anamorphosis is a means of transforming data, so that the marginal distribution can be assumed Gaussian.
- Higher dimensional distributions in spatial and multivariate problems are however not made Gaussian this way.
- It is necessary to take care of the dependence structure in these problems.

Anamorphosis

Anamorphosis is widely used in geostatistics, in particular for simulation of Gaussian random functions.

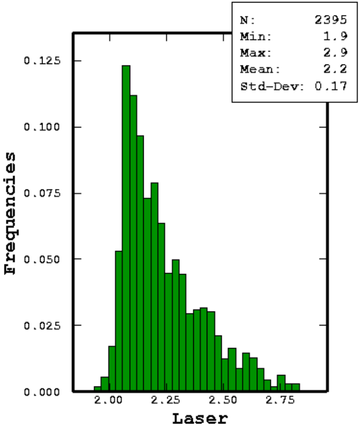
- Data for each variable is transformed into Gaussian equivalents;
- it is usually assumed that the multivariate distributions are Gaussian.

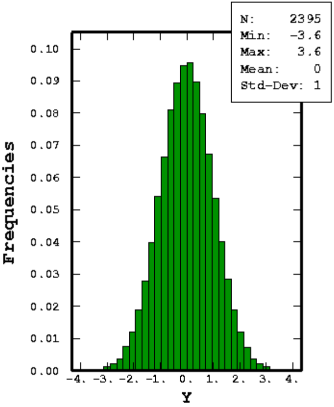
In the data ensemble assimilation litterature Gaussian anamorphosis appears in Bertino et al. 2003, Simon & Bertino 2009, ...



Example: laser data time series

Sea level heights measured during 20 minutes at Ekofisk platform (Jan 1st, 2002)

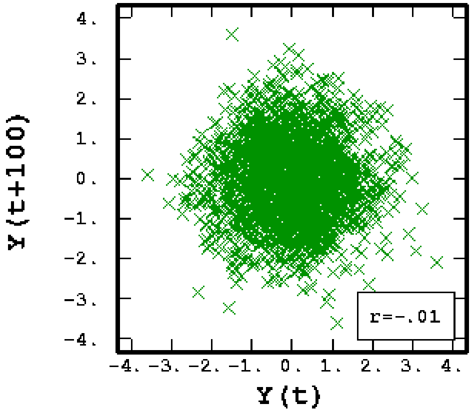




Histograms of the original laser data and the corresponding Gaussian values.

Lagged-scatterplot for $\Delta t = 100$ seconds

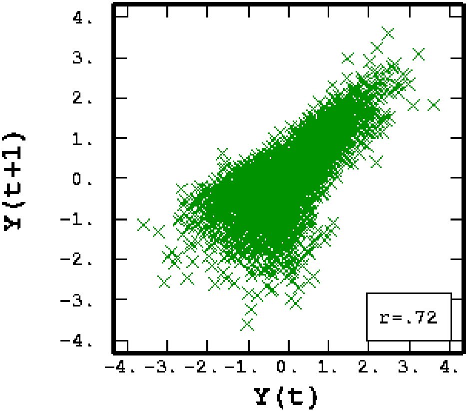
Comparison of Gaussian values 100 seconds apart



The shape is **circular**: it suggests a realization of a bivariate Gaussian distribution with zero correlation.

Lagged-scatterplot for $\Delta t = 1$ second

Comparison of Gaussian values 1 second apart



The shape is not **ellipsoidal**, ie the bivariate distribution is **not bi-Gaussian** - although the marginal distributions are Gaussian.

Discussion

- Anamorphosis secures that marginal distributions are Gaussian.
- However, bivariate and multivariate distributions are not necessarily Gaussian.

It is interesting to study the dependence structure especially for inspection of the tails of the bivariate distributions.

Dependence structure

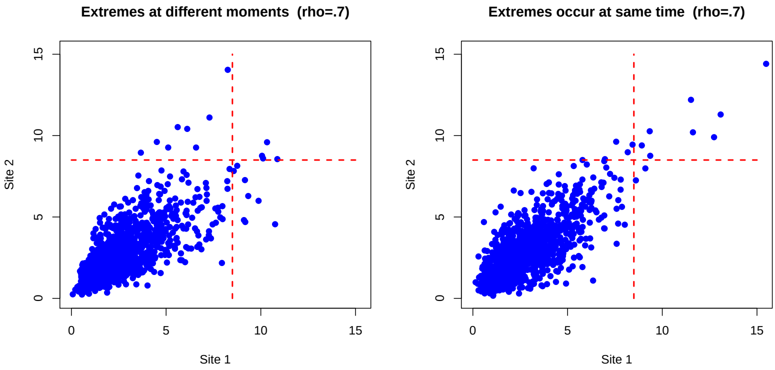
Multivariate extremes

In a multi-variate or a multi-location setting we may wonder:

- how likely is it that an extreme event occurs simultaneously for **two** (or more) **variables**?
- how likely is it that an extreme event occurs simultaneously at **two** (or more) **geographical locations**?

The bivariate distributions contain the answer.

Example: scatterplots between two sites



The correlation coefficient is the same for the two realizations:

$$\rho = 0.7$$

However, the behaviour of bivariate extremes different between left and right

Actually different dependence functions were used to construct these examples.

Joint distribution and copula

The idea in defining the copula C is to separate the **dependence structure** from the marginal distributions:

$$F(Z_1, Z_2, \ldots, Z_N) \quad = \quad C(\ F_1(Z_1),\ F_2(Z_2), \ldots, F_n(Z_n)\ )$$

The copula is itself a multivariate distribution, but with unit marginals.

Bivariate distributions with Gamma marginals

We consider two bivariate distributions $F_{\mathrm{Gau}}, F_{Gum}$ with **identical** Gamma(3,1) marginals F_1, F_2 and with **different** dependence structure:

$$
\begin{aligned}
F_{\mathrm{Gau}}(z_1, z_2) &= C_{\varrho}^{\mathrm{Gau}}(F(z_1), F(z_2)) \\
F_{\mathrm{Gum}}(x, y) &= C_{\beta}^{\mathrm{Gum}}(F(z_1), F(z_2))
\end{aligned}
$$

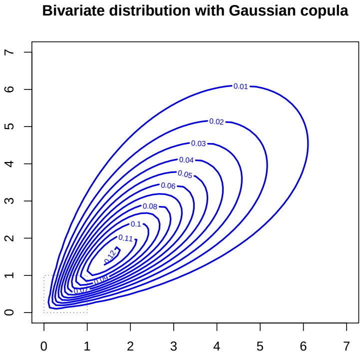
To obtain the **same overall linear correlation** $\rho = .7$ the parameters of the copula functions were set to:

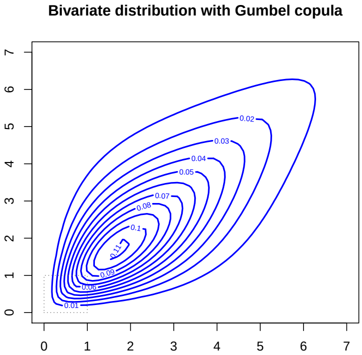
$$\varrho = .71 \quad \text{and} \quad \beta = .54$$

(for details see Embrechts et al. 2002, p22)

Plot of the two bivariate distributions

Marginals are Gamma(3,1)





Using **Gauss** copula

Using **Gumbel** copula

Discussion

Copulas are a convenient tool for analysing bivariate distributions and separating the dependence structure from the marginal distributions.

- The **Gaussian copula** belongs to the family of elliptic copulas, and implies asymptotically independent extremes.
  This may be unrealistic!!!

- The **Gumbel copula** does not belong to that family and is suitable for extreme value analysis/simulation.

Thus, paradoxically, from the point of view of extreme value theory a Gaussian dependence structure is generally not desirable!

Discussion (2)

For a detailed definition of asymptotic dependence/independence of extremes see e.g. Bacro & Toulemonde (2013) who list the following facts:

- jointly Gaussian variables which are not perfectly correlated are **asymptotically independent**;
- independence implies **asymptotic independence** but the converse is not true;
- detecting **asymptotic independence** is fundamental;
- fitting **asymptotically dependent** models to **asymptotically independent** data leads to over- or under-estimation of probabilities of extreme joint events.

This topic is clearly of interest in multivariate ensemble forecasting.

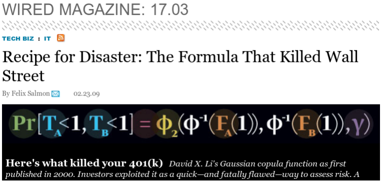
Anecdote: financial extremes

Gaussian copulas were the main ingredient of a formula proposed by Li for financial analysis in 2000.

It has been widely used by financial industry due to its simplicity.

Its inherent **underevaluation of joint risks** is deemed to be partly responsible for the unforeseen advent of the financial crisis of 2007-2009.

See the web paper by Salmon (2009):



WIRED MAGAZINE: 17.03

TECH BIZ : IT

Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon 02.23.09

$$\Pr[T_A<1, T_B<1] = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)), \gamma)$$

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A*

Ensemble forecasting

Multivariate empirical copulas

Schefzik (2013) defines multivariate discrete copulas and
applies them in an ensemble forecasting method (Ensemble Copula Coupling).

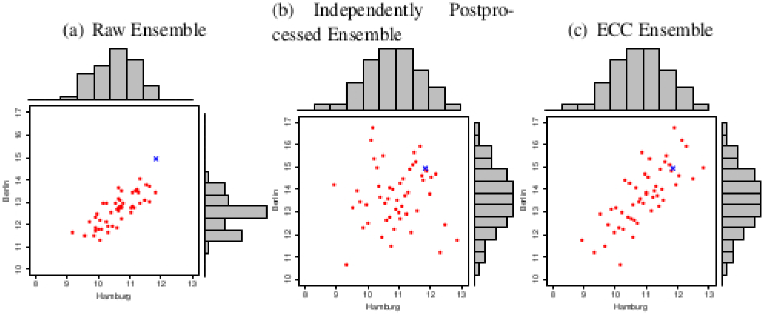
Three steps of ECC applied to a raw ensemble of numerical model runs:

1. Apply to each variable (or site) **individually** a statistical postprocessing technique (e.g. BMA) to get calibrated and sharp predictive distributions.
2. Draw a sample from each predictive distribution.
3. Rearrange these samples in the rank order structure of the initial raw ensemble to get the ECC postprocessed ensemble.

Step 3 represents comparably negligible numerical effort.

Example (two sites): ensemble copula coupling

24h-forecasts for temperature with 50-member ECMWF-ensemble (from Schefzik, 2013)



Raw (a), independently postprocessed (b) and ECC ensembles (c):

- the dependence structure of the raw ensemble is reproduced in the ECC ensemble (analog patterns).

Discussion

Empirical copula coupling: a simple and fast approach to assemble samples from individual predictive distributions into a joint multivariate distribution respecting the initial dependence structure.

- Wilks (2014) has compared ECC with the Schaake shuffle and recommends the latter type of empirical copula.

Conclusion

- It is essential in multivariate problems to generate predictive distributions with the appropriate dependence structure.
- Empirical copula methods provide a simple and fast method to assemble **individual** forecasts.
- It has not been examined yet how this compares to **joint** forecasting of predictive distributions.
- Modelling and forecasting extremal dependencies for multivariate processes are a challenging problem in ensemble forecasting and assimilation.

Acknowledgments

We acknowledge support from nordic Nordforsk EmblA project and from norwegian Petromaks program.

References

Bacro, J.-N., and Toulemonde, G.
Measuring and modelling multivariate and spatial dependence of extremes.
Journal de la Société Française de Statistique 154, 2 (2013), 139–155.

Davison, A. C., Padoan, S. A., and Ribatet, M.
Statistical modeling of spatial extremes.
Statistical Science 27 (2012), 161–186.

Embrechts, P., McNeil, A., and Straumann, D.
Correlation and dependence in risk management: properties and pitfalls.
In *Risk Management: Value at Risk and Beyond* (2002), M. A. H. Dempster, Ed., Cambridge University Press, pp. 176–223.

Schefzik, R.
Ensemble copula coupling as a multivariate discrete copula approach.
arXiv:1305.3445 (2013).

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

Wilks, D.
Multivariate ensemble model output statistics using empirical copulas.
Q. J. R. Meteorol. Soc. (2014), doi: 10.1002/qj.2414.