# Multivariate verification: Motivation, Complexity, Examples

A.Hense, A. Röpnack, J. Keune, R. Glowienka-Hense, S. Stolzenberger, H. Weinert

Berlin, May, 5th 2017

Meteorologisches Institut Universität Bonn
MiKlip
Hans-Ertel-Zentrum für Wetterforschung
universitätbonn RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT

- Motivations for MV verification
  - Data assimilation as a multivariate problem
  - Structures and physical processes
  - Detecting non-meteorological structures/patterns
- The problems with MV verification
  - univariate as subset of multivariate statistics
  - Dimensionality
  - Beyond multivariate Gaussian analysis?
- Some examples

**Definition:**

- univariate verification in weather prediction: single gridpoint, single lead time, single variable with ”many” observations
- multivariate verification: several gridpoints, several lead times, several variables in all possible combinations with respective observations
- all aspects of spatial verifications are covered by multivariate verification

**Question:**

Do observations and simulations coincide in structure ?

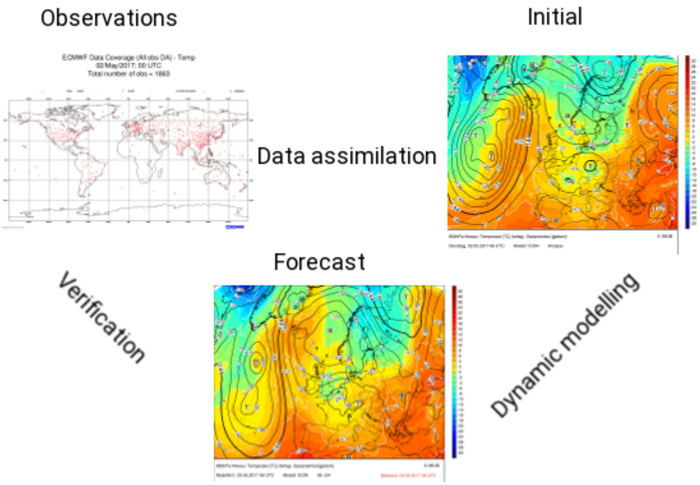
# The roots, 1

general approach to physics based weather forecasting was introduced by Vilhelm Bjerknes (1862-1951) in 1904

- observe the atmosphere
- generate a continous field of initial values (”data assimilation”)
- apply the laws of physics to advance in time
- issue as forecast
- (verification after the forecasts, not mentioned by V. Bjerknes)

https://en.wikipedia.org/wiki/Vilhelm_Bjerknes

#/media/File:Vilhelm_Bjerknes_Bust_01.jpg


Observations
ECMWF Data Coverage (All obs DA) - Temp
02/May/2017; 00 UTC
Total number of obs = 1663
Data assimilation
Initial
Dynamic modelling
Forecast
Verification

# The roots, 2

Let me remind you that ”everything in statistics” is explained by Bayes-Theorem (Thomas Bayes, $\sim$ 1701 - 1761)

$$[\vec{\theta}|\vec{o}] = [\vec{o}|\vec{\theta}]\frac{[\theta]}{[\vec{o}]}$$

- $\vec{o}$ the observations in space and time described by its pdf $[\vec{o}]$
- $\vec{\theta}$ the control variables in space, time and model parameters with pdf $[\vec{\theta}]$
- find the maximum of the conditional pdf $[\vec{\theta}|\vec{o}] \stackrel{!}{=}$ Max
- or estimates the most probable control variables given the observations

$$E(\theta|\vec{o}) = \int \theta[\vec{\theta}|\vec{o}]d\theta$$

- but the full conditional pdf $[\vec{\theta}|\vec{o}]$ contains much more information
- every pdf is necessarily a MV pdf

This can formally be solved by

$$
\begin{aligned}
[\vec{\theta}|\vec{o}] &= [\vec{o}|\vec{\theta}]\frac{[\theta]}{[\vec{o}]} \\
&= \int [\vec{o}, \vec{m}|\vec{\theta}] d\vec{m} \frac{[\theta]}{[\vec{o}]} \\
&= \int [\vec{o}, |\vec{m}\vec{\theta}][\vec{m}|\vec{\theta}] d\vec{m} \frac{[\theta]}{[\vec{o}]}
\end{aligned}
$$

in case of maximisation $[\vec{o}]$ is not necessary.

# Data assimilation

Expressing the likelihood $[\vec{o}, |\vec{m}\vec{\theta}]$ and the prior $[\vec{m}|\vec{\theta}]$ as MV-Gaussians, making the assumption that the major contribution to the integral comes from the maximum of the exponent (Laplace method) we get

$$\mathcal{J} = \frac{1}{2}(\vec{o}-\vec{H}(\vec{m}))^T R^{-1}(\vec{o}-\vec{H}(\vec{m}))+\frac{1}{2}(\vec{m}-\vec{M}(\vec{\theta}))^T B^{-1}(\vec{m}-\vec{M}(\vec{\theta}))$$

$$\vec{\theta}_s = \min_{\vec{\theta}} \mathcal{J}$$

where $\vec{H}(\vec{m})$ is the socalled forward operator which maps the physical variables of the forecast $\vec{m}$ to the measurable quantities $\vec{o}$ and $\vec{M}(\vec{\Theta})$ is the forecast model which takes the parameters $\vec{\Theta}$ to produce the actual forecast $\vec{m}$ which is a very large dimensional vector containing all prognostic variables at all vertical levels and all horizontal gridpoints/grid volumes/wave amplitudes (typical size $\sim 10^7 - 10^9$)

# Dynamic modelling

The physics, e.g. continuity equation of a hydrostatic atmosphere in $\sigma = \frac{p}{p_s}$ coordinates

$$\frac{d}{dt} \ln p_s + \int_0^1 \vec{\nabla}_\sigma \cdot \vec{v}_h d\sigma = 0$$

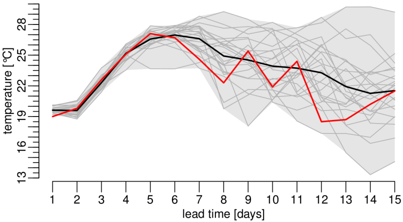
introduce dependencies

- in the horizontal through $\vec{\nabla}_\sigma \cdot \vec{v}_h$
- in the vertical through $\int_0^1 \vec{\nabla}_\sigma \cdot \vec{v}_h d\sigma$
- in time through $\frac{d}{dt} \ln p_s$
- and between the variables $p_s$ and $\vec{v}_h$
- similar for the remaining set of dynamic equations

# The Forecaster

- known from weather forecasting ”smoke plume”: mean ± Min,Max
- instead time also height
- instead 1 - 15 days also 1-15 years from medium range climate forecasts
- or global mean temperature of the 20th century from CMIP

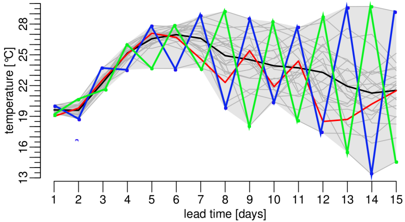
**$T_{2m}$ forecast Stuttgart summer 2010**



# The Forecaster

- known from weather forecasting ”smoke plume”: mean ± Min,Max
- instead time also height
- instead 1 - 15 days also 1-15 years
- or global mean temperature of the 20th century

**$T_{2m}$ forecast Stuttgart summer 2010**



Preliminary summary:
the Bjerknes weather forecasting chain has shown that

- data assimilation is a multivariate statistical process joining multiple observations in space, time and variable with their counterparts in a weather forecasting model
- weather forecasting with a dynamical model is based on physical connections between different variables in space and time
- use of forecasts from numerical processes implies the use of ”realistic” structures / features from the dynamical weather forecasting model

Preliminary summary cont.:

- it is only the verification step, which (mostly) ignores the dependency structure between different variables, in space and time using univariate verification
- but already the verification of a one gridpoint, one lead time, one variable forecast is a bivariate statistical problem because one evaluates the bivariate joint probability density function (e.g. estimated by contingency tables or scatter diagrams; Murphy and Winkler, 1987)

But what are the difficulties in multivariate verification/statistics?

- MV statistics is only weakly covered during a typical meteorological education, despite one of the major text books
  *Anderson, T. W. (1984). Multivariate statistical analysis. Wiley and Sons, New York, NY.* with its first edition in 1958
- the dimensionality problem or the ”curse of dimension”
- standard multivariate Gaussian density is not applicable in all situations: cloud cover, precipitation (above threshold)

let's start with discrete forecasts

- in $K$ classes e.g. $K = 2$ for precip forecasts $\gtrless$ than a threshold at $q$ forecast positions
- to be verified at $r$ observational positions (in space and/or in lead time).

Then the joint probability mass distribution between the forecast vs observational outcomes

- has $K^{q+r} - 1$ independent entries
- $(-1)$ due of the normalization constraint that the sum over all joint probability entries is one.

- for contingency tables with $K = 2$ with $q = r = 1$ we get $2^2 - 1 = 3$ entries,
- for tables based on a tercile segmentation $K = 3$ we get $3^2 - 1 = 8$ a quadratic $q + r = 2$ increase
- increasing the number of points for the $K = 2$ case e.g. to $q = r = 2$ gives already $2^4 - 1 = 15$ necessary entries which leads to an exponential increase.

All entries have to be estimated from observations:

- you must have at least a sample size of $\mathcal{O}(K^{q+r} - 1)$ to fill in on average one observation into each joint probability bin.
- consider working with binary variables on a 3 by 3 grid in observations and forecasts,
- this would require the incredible sample size $> 2^{18} - 1 \sim 270,000$.

Problems can be remedied by turning to parametric probability mass distribution in case of discrete forecasts or parametric probability density functions

- Gibbs distributions $[\vec{x}] = \frac{1}{Z}\exp(-V(\vec{x}))$ with $Z$ as the normalizing constant (partition function) and $V$ a convex function (potential well)
- e.g. for a discrete binary field like precipitation below/aboe a threshold $x_i \in \{0, 1\}$

$$V = \sum_i m_i x_i + \frac{1}{2}\sum_i \sum_j J_{ij} x_i x_j$$

  with parameters $m_i$ und $J_{ij} = J_{ji}$, such that $(q+r) + \frac{1}{2}(q+r)(q+r+1) = \frac{(q+r)}{2}(q+r+3)$ unknowns have to be determined which grows quadratically
- unfortunately for multivariate parametric probability mass distribution $[\vec{x}]$ standard parameter estimation does not work. because $Z(m_i, J_{ij})$ is in general not known in closed form

Much easier for various (but not all) continous variables: using the multivariate Gauss density

$$[\vec{x}] = \frac{1}{Z}\exp(-V(\vec{x}))$$

with

$$Z = \sqrt{2\pi^{q+r}\det\Sigma}$$

$$V(\vec{x}) = \frac{1}{2}(\vec{x}-\mu)^T\Sigma^{-1}(\vec{x}-\vec{\mu}))$$

$$\vec{x} = (\vec{m},\vec{o}) \quad \mu = (\vec{\mu}_m,\vec{\mu}_o)$$

$$\Sigma = \begin{pmatrix}\Sigma_{mm} & \Sigma_{mo}\\ \Sigma_{mo}^T & \Sigma_{oo}\end{pmatrix}$$

with well known methods since decades (see the monograph by TW Anderson (1958, 2nd Ed. 1984)) e.g for estimating from samples of $\vec{f}, \vec{o}$ the location parameter $\mu$ and the covariance matrix $\Sigma$ using maximum likelihood techniques $\frac{(q+r)}{2}(q+r+3)$ parameters or a quadratic increase in complexity.
Unfortunately the estimated covariance matrix $\Sigma$ has to fulfill certain requirements

- positive definitness $\vec{x}^T \Sigma \vec{x} > 0$ if $\vec{x} \neq 0$
- non singular $\Sigma^{-1}$ has exist or $\Sigma$ has to be of full rank $\mathrm{rk}(\Sigma) = (q+r)$

Standard maximum likelihood estimator for $\Sigma$ from a joint sample of forecasts and observations $\{\vec{d}_i = (\vec{m}_i, \vec{o}_i), i = 1, I\}$ reads

$$\Sigma \stackrel{est}{=} \Sigma_{mle} = \frac{1}{I-1} D'(D')^T$$

with $D'$ the $(q+r) \times m$ anomaly data matrix build from columns $\vec{d}_i' = \vec{d}_i - (\vec{m}_m, \vec{m}_o)$ and

$$(\vec{m}_m, \vec{m}_o) = \frac{1}{I} \sum_{i=1}^{I} \vec{d}_i$$

now lets calculate the rank of $\Sigma_{mle}$

$$\text{rk}(\Sigma_{mle}) = \text{rk}(\frac{1}{I-1} D'(D')^T) \leq \text{rk}(D') \leq \min(I-1, q+r)$$

meaning that $\Sigma_{mle}$ is only of full rank of the sample size $I$ is larger than the vector dimension $q+r$

It is even worse...
We do not need the actual, estimated covariance matrix $\Sigma_{mle}$ but its invers $\Sigma_{mle}^{-1}$, to model completely the multivariate probability density $[\vec{x}]$. It turns out that the estimated covariance matrix ist (almost) unbiased

$$E[\Sigma_{mle}] = \Sigma$$

but the invers of the estimated covariance is strongly biased

$$E[\Sigma_{mle}^{-1}] = \frac{l-1}{l-q-1}\Sigma^{-1}$$

depending on the ratio $\frac{l-1}{l-(q+r)-1}$, meaning that even non-singular estimated covariance matrices lead to massively distorted invers matrices as long as $l$ is not massively larger than $(q+r)$
This are the remains of the ”curse of dimensions” in case of a multivariate Gaussian density (also present in data assimilation)

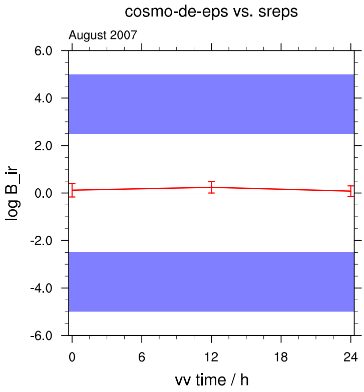
Ways out of the problem

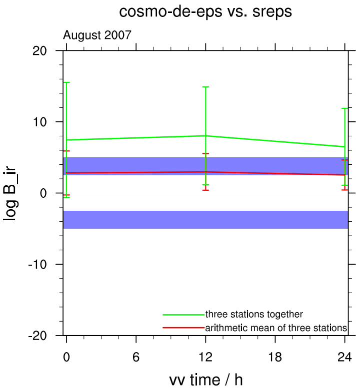
- data or dimension reduction: instead of $q + r$ grid points think and compute in $\tilde{q} + \tilde{r}$ ”structures”, ”modes”, ”patterns” defined by the problem/researcher e.g. from simple models with $\tilde{q}, \tilde{r} \ll q, r$
- not necessarily only principle component analysis (EOF) or comparable statistical techniques
- alternative methods to estimate non-singular invers covariance matrices: shrinkage methods and GLASSO methods
- combinations of both

**Added value of multivariate approach**, 21 day mean August 2007, 3 Radiosonde stations with 9 Levels each: Nancy, Idar-Oberstein, Stuttgart, Röpnack et al Mon.Weath.Rev. (2013) based on the log Bayes factor

classical univariate

two multivariate approaches



Multivariate extension of continous rank probability score CRPS for probabilistic forecasts: energy score

$$es\left(f_M(\vec{m}), \vec{o}\right) = E\{\|\vec{m} - \vec{o}\|\} - \frac{1}{2}E\{\|\vec{m} - \vec{m}'\|\}$$

parametrize predictive pdf as

- Gaussian-pdf $NV(\vec{\mu}_M, \Sigma_M^{-1})$
- Gaussian-mixture $\frac{1}{K}\sum_k NV(\vec{m}_k, \Sigma_e^{-1})$

both parameter sets estimated from ensemble realizations (post-processing).
Score calculated across all available observations

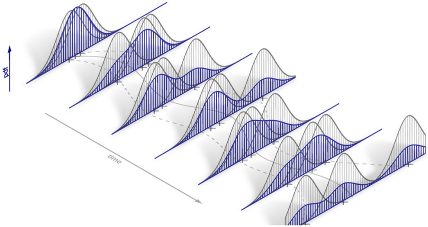
$$ES_M = \frac{1}{T}\sum_{t=1}^{T} es(f_M(\vec{m}, t), \vec{o}_t)$$

with the skill score relative to climate

$$ESS = 1 - \frac{ES_M}{ES_{clim}}$$

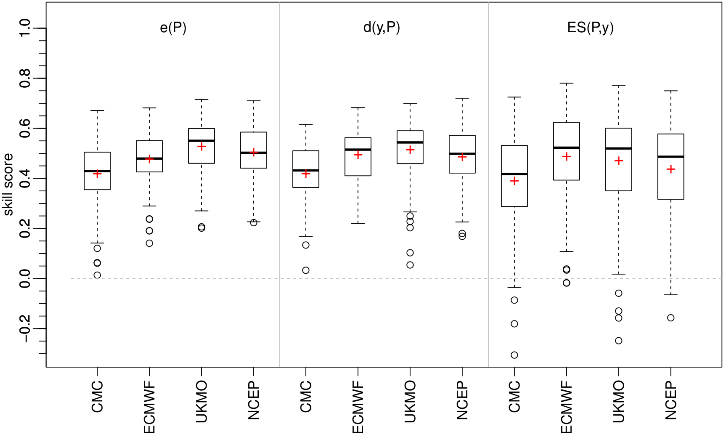
Non-Gaussian probability density functions: Gaussian mixtures combine Gaussian versatility with modelling non-Gaussian pdf's

$$[\vec{x}|K,\vec{x}_k,\Sigma_e^{-1}] = \sum_{k=1}^{K} NV(\vec{x}|\vec{x}_k,\Sigma_e^{-1})$$



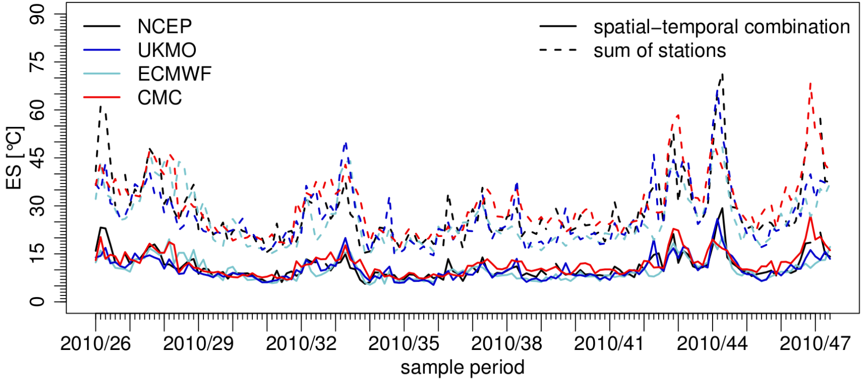
# Comparison of 4 EP systems TIGGE data base, Stuttgart, $T_{2m}$, July-Nov. 2010, energy score based, ten-day forecasts

Keune et al. Mon. Weath. Rev. (2014)



Combine ten day forecast sequences at eight stations: 80-dim vector

**With vs without spatial correlations between eight German station $T_{2m}$**



- The whole Bjerknes chain for an integrated forecasting system is based on multivariate statistics, relevant structures, dynamical connections in space, time and between variables
- except the verification: current verification measure largely ignore these connections dictated by physics
- taking into account the structural information or ”correlations”: better scores compared to the univariate case in two examples
- MV verification comes with extra expenses related to the ”curse of dimension”
- which can be treated by methods from MV statistics coming from image processing, mode expansion etc.