The priority program SPP1167 "Quantitative Precipitation Forecast" PQP and the stochastic view of weather forecasting

Andreas Hense

15. Juni 2007
Overview

• The priority program SPP1167: mission and structure

• The stochastic view of weather forecasting
  – Basic theory
  – Verification, Assimilation and Calibration
  – Applications
History:

- first round table discussions 2000 and 2001
- submission preproposal November 2001
- withdrawal February 2002
- Elbe flooding Summer 2002
- finalization of proposal end of 2002
- submission in February 2003
• approval by the senate of DFG in May 2003 for 6 years

• submission of individual proposals Oct. 2003, approval in January 2004

• kick-off meeting in April 2004

• second phase since April 2006, now halftime
The goals from the initial proposal:

- Identification of physical and chemical processes responsible for the deficiencies in quantitative precipitation forecast

- Determination and use of the potentials of existing and new data and process descriptions to improve quantitative precipitation forecast

- Determination of predictability of weather forecast models by statistico-dynamic analyses with respect to quantitative precipitation forecast
These goals are related to the following tasks

- Understanding of atmospheric processes relevant to precipitation formation has to be improved considerably. These processes have to be modeled in a more realistic manner.

- Initial distribution of the atmospheric water content in the three phases of vapor, liquid, and ice has to be improved by data that have not been used so far and new data. Their potential for quantitative precipitation forecast has to be quantified.
• Methods of assimilating measurement data into atmospheric simulation models and application of these methods to data of any type for a statistico-dynamically consistent retrieval of the initial water distribution and, hence, optimum use of all data have to be improved.

• The stochastic character of precipitation has to be taken into account when using observational data, interpreting deterministic simulations, and developing alternative forecast strategies.
Quantitative Precipitation Forecasts (QPF)

Research of atmospheric processes and models to enhance the predictability of quantitative precipitation.

A) processes and model physics

B) data base preparation and allocation

C) assimilation and statistical-dynamical methods

E) GOP und field experiment

D) operational test environment (NUMEX/PF-tools)

DFG-Schwerpunktprogramm

Universities and Research institutions

Operational weather forecast system

DWD
Why did we include the stochastic aspect?

- for practical reasons
- to treat the physics correctly
The atmosphere is a stochastic system due to

- its high-dimensionality

- its non-linear dynamics
• total volume of the atmosphere: spherical shell thickness $h \sim 100\text{km}$

• smallest unit volume $V_0 \sim 1\text{mm}^3 = 10^{-9}\text{m}^3$

• within each unit volume about $n_i \sim \mathcal{O}(100)$ variables like $\Theta, p, \rho_i, u, v$, etc

\[
\begin{align*}
V & \sim 4\pi h R_{\text{Earth}}^2 \\
& \sim 4\pi 10^5 (6.37^2 10^{12}) \\
& \sim 5 \times 10^{19} \text{m}^3 \\
\frac{V}{V_0} & \sim 5 \times 10^{28} \\
\frac{V n_i}{V_0} & \sim 10^{30}
\end{align*}
\]
We need about $10^{30}$ numbers to characterize the actual state of the atmosphere $\vec{T}$.
The observed state $\vec{o}$ is much smaller in dimension through a projection by measurements $H$: $\vec{o} = H(\vec{T}) + \vec{e}_o$.
- detailed computations are impossible

- statistical physics: probability density description $p(\vec{T})$ with

$$\int p(\vec{T})d\vec{T} = 1$$
• high resolution global models: discrete spherical shell

• with about $2000 \times 1000 \times 100$ grid points

• unit volumes $20\text{km}^2 \times 1\text{km}$

• within each unit volume about $n_i \sim \mathcal{O}(10)$ variables like $\Theta, p, \rho_i, u, v,$ etc
We can treat numerically about $10^9$ degrees-of-freedom to characterize the actual state $\tilde{f}$ of high resolution atmospheric model.

- the remaining d-o-f’s in the state description introduce observational or aleatoric uncertainty $\epsilon_o$

- the influence of the remaining d-o-f’s ($\sim 10^{21}$) on the dynamics has to be parametrized introducing structural or epistemic uncertainty

- both have to be treated by probabilities
Additionally: atmosphere and its quasi-realistic model counterparts are nonlinear dynamical systems with embedded instabilities like

- convective instability
- inertial instability
- barotropic and baroclinic instability

Nonlinearities and instabilities amplify small uncertainties (e.g. from the aleatoric part of initial conditions):

The famous Lorenz butterfly (using the Lorenz (1984) model)
Zeit [d]
Zonaler Grundstrom
Wahrscheinlichkeitsdichte
Consequences (Murphy and Winkler, 1987):

- any comparison of simulation $\vec{f}$ and observation $\vec{o}$ e.g. for verification has be done using the joint pdf $p(\vec{o}, \vec{f})$.

- can be done with the Bayes-theorem

$$
p(\vec{o}, \vec{f}) = p(\vec{o}|\vec{f})p(\vec{f})
= p(\vec{f}|\vec{o})p(\vec{o})
$$

$$
p(\vec{f}|\vec{o}) = \frac{p(\vec{o}|\vec{f})p(\vec{f})}{p(\vec{o})}
$$

- note that $\vec{o} = (\vec{o}_t, \vec{o}_{t+1}, \ldots, \vec{o}_{t+\tau})$ and similar for $\vec{f}$
Verification at lead time $\tau$ and stationarity assumption

$$p(f_{\tau}|\tilde{o}_{\tau}) = \frac{p(\tilde{o}_{\tau}|f_{\tau})p(f_{\tau})}{p(\tilde{o}_{\tau})}$$

$p(\tilde{o}_{\tau}|f_{\tau}) = p(\tilde{o}_{\tau}|f_{\tau}, \tilde{H}, \lambda)$ models the measurement process, $\lambda$ unknown parameters describe the difference between the actual measurement $H$ and its discrete analog $\tilde{H}$

$\lambda$ are fixed by maximizing $p(\tilde{o}_{\tau}|f_{\tau})$

$$p(f_{\tau}|\tilde{o}_{\tau}) = \arg\max_{\lambda}(\frac{p(\tilde{o}_{\tau}|f_{\tau})p(f_{\tau})}{p(\tilde{o}_{\tau})})$$

$$= \frac{1}{p(\tilde{o}_{\tau})}\arg\max_{\lambda}(p(\tilde{o}_{\tau}|f_{\tau})p(f_{\tau}))$$
• calibration or model output statistics MOS (Glahn and Lowry, 1972)
• prior $p(\vec{f}_τ)$ can be obtained from marginalisation

$$p(\vec{f}_τ) = \int \ldots \int p(\vec{f}_0, \ldots, \vec{f}_τ) df_0 \ldots df_{τ-1}$$

• and dynamics

$$p(\vec{f}_0, \ldots, \vec{f}_τ) = p(\vec{f}_τ|\vec{f}_0, \ldots, \vec{f}_{τ-1})p(\vec{f}_0, \ldots, \vec{f}_{τ-1})$$

$$p(\vec{f}_τ|\vec{f}_0, \ldots, \vec{f}_{τ-1}) = \frac{δ(\vec{f}_τ - M(\vec{f}_0, \ldots, \vec{f}_{τ-1}))}{\text{in case of a deterministic model}}$$

• iterating leads to (Markov-chain)

$$p(\vec{f}_0, \ldots, \vec{f}_τ) = \underbrace{p(\vec{f}_0)}_{\text{Initial conditions}} p(\vec{f}_τ|\vec{f}_0, \ldots, \vec{f}_{τ-1})p(\vec{f}_{τ-1}|\vec{f}_0, \ldots, \vec{f}_{τ-2}) \cdots$$
- if $p(f_0)$ should depend on observations $\tilde{o}_0$: data assimilation

- with $p(f_0)$ replaced by $p(f_0|\tilde{o}_0)$
Theoretical consideration about the structure of the problem show that

- probabilistic treatment interrelates the verification, the calibration (MOS) and the assimilation

- but is unusable: multidimensional integrals and optimization

- but allows one to ”boil” down the problem

- and register the assumption necessary to arrive at the final problem
Bayesian Forecasting for Complex Systems Using Computer Simulators

Peter S. CRAIG, Michael GOLDSTEIN, Jonathan C. ROUGIER, and Allan H. SEHEULT

J. American Statistical Association (2001)

1. The simulator is not an exact representation of the system.

2. Values of many of the simulator input parameters are unknown.

3. The simulator can be regarded as a "black box"

4. The simulator is slow and expensive to run

5. Collections of both past observations and future outcomes may be very large and have complicated spatial/temporal structure.
• linear Bayes-Statistics

• Normal distribution with mean and covariance matrices

• e.g. Assimilation as 3d/4d-Var: optimization of a quadratic function

\[
\mathcal{J}(f_0) = (\tilde{\sigma}_0 - \tilde{H} \tilde{f}_0)^* \Sigma_o^{-1} (\tilde{\sigma}_0 - \tilde{H} \tilde{f}_0) + (\tilde{f}_0 - M(\tilde{f}_{T=-1}))^* \Sigma_b^{-1} (\tilde{f}_0 - M(\tilde{f}_{T=-1}))
\]

• appropriate for any situation?

• e.g. calibration of univariate observed variables vs. multivariate forecasts: standard linear multiple regression

• e.g. verification of univariate variables: mean square error based statistics and skill scores
Estimation

- to be estimated probabilities, probability density functions or parameters describing these (e.g. mean and covariance matrices)

- under ergodicity (transitivity) assumption: counting events in time "frequencies"

- under homogeneity and isotropy assumption: counting in space "intensities"

- model probabilities from (small-size) ensemble simulations

- point estimates need confidence interval estimates
A calibration example (T. Winkelinkemper, diploma thesis)
multiple linear regression with Kalman updating of regression coefficients

$T_{\text{max}}$ Stuttgart August 2003
Bootstrap estimation: a computer based statistics method for point and confidence intervals
Summer 2005 LM forecasts radiosonde Essen

- **MSE [K^2]**
  - LM+00h
  - LM+24h
  - LM+48h

- **Brier Skill Score [1]**
  - LM+00h
  - LM+24h
  - LM+48h

- **sq. correlation [1]**
  - LM+00h
  - LM+24h
  - LM+48h

- **Temperature [K]**
  - Obs time mean
  - LM+00h
  - LM+12h
  - LM+24h
  - LM+48h

- **Bias**
  - LM+00h
  - LM+24h
  - LM+48h

- **MSE [K^2]**
  - LM+00h
  - LM+24h
  - LM+48h

- **Brier Skill Score [1]**
  - LM+00h
  - LM+24h
  - LM+48h

- **sq. correlation [1]**
  - LM+00h
  - LM+24h
  - LM+48h

- **Temperature [K]**
  - Obs time mean
  - LM+00h
  - LM+12h
  - LM+24h
  - LM+48h

- **Bias**
  - LM+00h
  - LM+24h
  - LM+48h
Conclusions

- Stochastic weather forecasting and its relatives are part of the SPP1167 mission

- Stochastic weather forecasting has its origin in the physics of atmospheric dynamics

- Stochastic weather forecasting can be structured through Bayes theorem

- Assimilation, Calibration and Verification are interrelated
• A theoretical overview is necessary to assess the assumption for specific problems

• Avoid "ad-hoc" statistics

• Practical advice: use bootstrapping