Experiences from implementing GPS Radio Occultations in Data Assimilation for ICON

Harald Anlauf

Data Assimilation Section, Business Area Research and Development
Deutscher Wetterdienst, Offenbach, Germany

OPAC-6 / IROWG-5 Workshop, Seggau Castle, Austria
8–14 September 2016
Outline

1 Introduction
   - Global NWP at DWD
   - History of GPS Radio Occultations at DWD

2 Utility of GPSRO in Data Assimilation and NWP Model Development

3 Do We Get The Most out of Radio Occultation Data?
   - Tuning of Observation and Background Error Covariances
   - Observation Operator Issues
   - Tangent Point Drift?

4 Why are NWP People concerned about Observation Error Correlations?

5 BUFR File Contents
Global NWP at DWD: ICON

- Joint project by DWD and MPI-M (Hamburg)
  - Unified framework for NWP and climate modeling
  - ICOsahedral-triangular (Arakawa C) grid, Non-hydrostatic core
  - unstructured mesh, local refinement options
    global/regional mode, self-nesting, 1- and 2-way nesting
  (G. Zängl et al., doi:10.1002/qj.2378)

- ICON Global NWP version operational on 2015-01-20
  - 13 km avg. mesh size, 90 levels (top: 75 km)
  - replacing former GME

- Europe nest (2015-07-21)
  - 6.5 km mesh size, 60 levs.
  - 2-way horizontal/vertical nesting with feedback to global domain
Global NWP at DWD: Data Assimilation

- 2008 – 2015: 3D-Var, 3 hourly cycle
- Jan. 2016: Hybrid Variational / Ensemble Data Assimilation ("VarEnKF")
  - High-resolution deterministic model: 13 km global / 6.5 km Europe
  - Ensemble: 40 members, 40 km global / 20 km Europe
  - High-resolution analysis: Ensemble-variational system ("EnVar"), with background error covariance
    \[ B = \alpha B_{EnKF} + (1 - \alpha) B_{clim} \]
    \( \alpha = 0.7 \) since Feb. 2016
    - \( B_{clim} \): climatological \( B \) matrix of 3D-Var
    - \( B_{EnKF} \) provides flow-dependent error covariances by ensemble
  - Ensemble analysis: Local Ensemble Transform Kalman Filter (LETKF)
History of GPS Radio Occultations at DWD

- **Bending angle** operator, based on code by Michael Gorbunov
  - 2000: 3d ray tracer for ECHAM
  - 2002: adapted 3d ray tracer to GME grid, non-operational 3D-Var
  - 2006: 1d operator (Abel integral) based on ray tracer code
    - fixed/effective tangent point for whole profile, or
    - individual tangent point for each ray

- Evaluation (monitoring) in collaboration with GFZ using CHAMP and GRACE data
  - Ray tracer needs (drifting) satellite positions and velocities
    - Processing using CT2 was done at DWD (Pingel and Rhodin, 2009).
  - Ray tracer best in terms of std.dev.(O-B), numerically expensive!
History of GPS Radio Occultations at DWD

- **2008**: DWD’s 3D-Var-PSAS for GME becomes operational
  - replacing former OI

- **2010**: operational implementation of GPSRO
  - 1d-operator with effective location of occultation
    (H. Anlauf et al., doi:10.5194/amt-4-1105-2011, some later refinements, see my IROWG-2 talk)
  - only affordable option on former NEC SX-9 computer

- **2015+**: evaluate operator variants accounting for tangent point drift
  - More important for ICON than for GME due to:
    - increased model top, horizontal resolution
    - major overhaul of model physics
  - We got another computer (Cray XC40) and could afford it now

- **2015**: revised refractivity calculation (incl. compressibility), tuning of observation and background error
Utility of GPSRO in NWP Development

- **Promises** of GPSRO:
  - High vertical resolution (in UT/LS better than most NWP models)
  - Globally distributed, with almost uniform coverage
  - Essentially bias-free ⇒ assimilate without bias correction
  - Well understood error characteristics (really?) of disseminated data

- By-product of data assimilation: “feedback files” with comparisons of observational data to first guess (O-B) and to analysis (O-A).
  - Assessment of model bias (relative to observations)
  - Diagnostics of DA performance and of problems
  - Statistical inference of (parts of) background and observation error covariances ($B$, $R$)
    - May require additional assumptions, as e.g. in Desrozières’ method
    - Optimality for non-linear systems, non-gaussian errors?

- Understanding/control of biases and proper choice of $B$ and $R$ are prerequisite to optimal utilization of data
Diagnosing a Problem in ICON’s Vertical Nest Interface

- Comparison of bending angles against model over Europe
  - Reduction of standard deviation of observation minus FG in troposphere, but increase around vertical interface levels (∼ 22 km)
  - Sharp structure in bias at vertical interface levels due to nest feedback: Inconsistencies in radiation fluxes – solved in a subsequent ICON release

129 = nest exp.  
130 = control (no nest)  
(normalization: \( \sigma_o \))

GPSRO in Data Assimilation for ICON
Tuning of Observation and Background Error Covariances

- **First-guess, analysis departures in observation space:**

  
  \[
  d^o_b := y^o - H(x^b) \, , \quad d^o_a := y^o - H(x^a) \, , \quad d^a_b := d^o_b - d^o_a
  \]

  Assuming no biases:

  \[
  \mathbb{E}\{d^o_b(d^o_b)^t\} = \mathbf{R} + \mathbf{HBH}^t
  \]

  and for an optimal linear DA system (c.f. Desroziers et al., 2005):

  \[
  \mathbb{E}\{d^a_b(d^o_b)^t\} = \mathbf{HBH}^t
  \]

  \[
  \mathbb{E}\{d^o_a(d^o_b)^t\} = \mathbf{R}
  \]

- **Note:**
  - Above estimators for \(\mathbf{R}\) and \(\mathbf{HBH}^t\) are not positive definite!
  - Off-diagonal elements less reliable than the diagonals (variances)
  - Cross-validation with non-assimilated data may be crucial
  - Long-term aim: situation-dependent errors (model and data!)
Tuning of GPSRO Observation Error

- Observation error initially chosen more conservatively for ICON
- Tuning of observation error needed, and possible:
  - Improvement of model biases since first operational implementation
  - Improved understanding of model ⇔ DA interaction, feedbacks
  - Desroziers diagnostics suggested:
    - Reduction of observation error (most regions/impact heights)
    - Latitudinal dependence
    - Representativity error (esp. near cold-point tropopause)

Harald Anlauf (DWD)
Further Improvements: Background Error, Refractivity

- Tuning just observation error did not lead to better forecast quality
  - Positive feedback between DA and the model’s convection scheme, required reducing of background humidity error
- Systematic bias between GPSRO and model in the stratosphere
  - Previously used refractivity expression (Rüeger) did not work well
  - Revised refractivity calculation, incl. non-ideal gas treatment (Aparicio and Laroche, 2011)

⇒ Consistent improvements in data assimilation cycle and forecasts!
Observation Modeling: Tangent Point Drift

- Initial tests with “optimized” 1d operator for tangent point drift
  - assigned each ray to nearest model column, batch model-column-wise
  - apparently small but systematic improvement in obs-fg statistics
    😞 Not confirmed in experiments: degradation in lower troposphere!?
    😞 Degradation of long forecasts with full set of operational observations
- Reasons for lack of positive impact not yet understood
  - Tangent-point drift not yet implemented operationally

---

**ICON GPSRO Bending Angle Statistics, Tropics**

- |lat| < 20

**ICON GPSRO Bending Angle Statistics, Polar regions**

- |lat| > 60

---

Harald Anlauf (DWD)  
GPSRO in Data Assimilation for ICON  
8 September 2016  
12 / 17
Why the Concerns about Observation Error Correlations?

- Data Assimilation is sensitive to appropriate knowledge and modeling of error statistics, even off-diagonal terms
  - Many (most?) DA codes optimized for essentially diagonal $R$ matrices. Wrongly neglecting off-diagonal terms may lead to degraded analyses.

- Toy example: model with 2 sites, $x = (x_1, x_2)^T$, with diagonal $B$, and 2 correlated observations $y = (y_1, y_2)^T$:

  $\begin{align*}
  H &= I \\
  B &= \sigma_b^2 I \\
  R &= \sigma_o^2 \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} \\
  r &= \frac{\sigma_o^2}{\sigma_b^2}
  \end{align*}$

Analysis in observation space: $y_a = Hx_a$

$$y_a = S y + (I - S) H x \quad S = R^{-1} H (B^{-1} + H R^{-1} H^T)^{-1} H^T$$

“Influence matrix” of observations on analysis:

$$S = \begin{pmatrix}
\frac{1+r}{(1+r)^2-\beta^2 r^2} & -\frac{\beta r}{(1+r)^2-\beta^2 r^2} \\
-\frac{\beta r}{(1+r)^2-\beta^2 r^2} & \frac{1+r}{(1+r)^2-\beta^2 r^2}
\end{pmatrix}$$
Why the Concerns about Observation Error Correlations?

- Remark: \( r, \beta \) are **assumptions** by the DA system, not the truth!

\[
S_{11} = \frac{1}{1+r} + \frac{r^2}{(1+r)^3} \beta^2 + O(\beta^4) \quad S_{12} = -\frac{r}{(1+r)^2} \beta + \ldots
\]

- **Assuming** observation error correlations increases observation impact, and leads to correlated contributions of other observations

- Let \( \tilde{r}, \tilde{\beta} \) be the **true** values of these parameters, and calculate true analysis error covariance \( A \) depending on **chosen** \( r, \beta \):

\[
A_{11}(r, \beta) = A_{11}(\tilde{r}, \tilde{\beta}) + c_{11}(r - \tilde{r})^2 + c_{22}(\beta - \tilde{\beta})^2 + 2c_{12}(r - \tilde{r})(\beta - \tilde{\beta}) + \ldots
\]

with a positive definite Hessian \( (c_{ij}) \), as to be expected.

- There is no compensation for wrong choice of \( r \) or \( \beta \!\
- It’s even hard to get one of them right!
- Traditional treatment in NWP: thin correlated data, increase obs. error
Data Smoothing by Processing Center or Super-Obbing?

- EUMETSAT test data, different smoothing settings: $\gamma = 1, 2, 3, 4$
  1. Vertical thinning to DWD’s operational setup for ICON, or
  2. Super-Obbing to same target vertical resolution
     - User-defined smoothing with user-controlled correlations
     - Residual $\gamma$-dependence due to information loss in thinning to BUFR

---

EUMETSAT's smoothing + data thinning

EUMETSAT's smoothing + 'superobbing'
Rick Anthes (IROWG-4, with references):

“Radio Occultation: World’s most accurate, precise, and stable thermometer from space”

- (Not only) for NWP: Good data should come with good metadata!
- WMO standard for dissemination of observational data: BUFR
  - Description of format available on ROM SAF website
  - Not all metadata fields filled in properly by some processing centers
  - Interpretation of WMO language, specifications needs getting used to
    - Additional clarifications: BUFR action group initiated after IROWG-4. Discussion paper and teleconference, summarized in:
      - Awaiting feedback ...
- Please help to get the most out of these valuable data!
Thank you for listening!