COSMIC 1D-Var Update and Validation

Tae-Kwon Wee

COSMIC Program Office
UCAR

IROWG-8 Virtual Meeting, 8 April 2021
Background

• **Purpose**
  Optimal estimation of the thermodynamic states (T,p,q; unobserved but needed by many data users) along the trajectory of ray tangent points.
  Based on the same principle/method with var data assimilation but differs in purpose and practice.

• **Requirements**
  Accurate (meeting mission requirements and other standards), reliable (stable), and fast enough for real-time data processing.

• **Information utilized**
  Observation and a priori (background) and their error statistics, and physical constraints (hydrostatic, sub-adiabatic, non-negative q) and relationships (e.g., forward models).
  Uses the background to tackle underdeterminancy, while attempting to minimize its adverse influence on the retrieval.

• **Status**
  Comprehensive validation completed for C2 and other missions including COSMIC.
  In use for C2 data processing and will be used for re- and post-processing of other missions.
<table>
<thead>
<tr>
<th></th>
<th>OLD (v 1.0)</th>
<th>NEW (v 2.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation</td>
<td>Incremental (Courtier et al., 1994), Control-Variable Transform (Parrish-Derber, 1992)</td>
<td></td>
</tr>
<tr>
<td>Control Variables</td>
<td>$T$, $p_L$, $R\text{H}^*$ (other options available)</td>
<td></td>
</tr>
<tr>
<td>Observation error</td>
<td>Extreme quality</td>
<td>Statistical (Hollingsworth-Lönnberg, 1986)</td>
</tr>
<tr>
<td>Background error (B)</td>
<td>Empirical</td>
<td>Statistical (HL86 + NMC-method)</td>
</tr>
<tr>
<td>Correlation</td>
<td>Univariate</td>
<td>Multivariate (NMC-method)</td>
</tr>
<tr>
<td>Resolution</td>
<td>3 latitude zones, seasonal</td>
<td>$5^\circ \times 5^\circ$ lat-lon grid, monthly</td>
</tr>
<tr>
<td>Construction</td>
<td>Run-time (expensive)</td>
<td>Precomputed $C^{1/2}$ on fixed levels (“no cost”)</td>
</tr>
<tr>
<td>Observation operator</td>
<td>Refractivity</td>
<td>Variational Abel Transform (Wee, 2018)</td>
</tr>
<tr>
<td>Terrain</td>
<td>None</td>
<td>Digital Elevation Model</td>
</tr>
<tr>
<td>Ducting</td>
<td>Disregarded</td>
<td>Truncation of model grids</td>
</tr>
<tr>
<td>I/O</td>
<td>ascii</td>
<td>netcdf</td>
</tr>
</tbody>
</table>
Observation error specification

- Locally accurate (statistically precise) OE, varies with height, latitude, and month
- Provides corresponding background errors for both bending angle refractivity

**annual-mean structure**

**temporal variation**
Background error spec.

- NMC method scaled towards the Hollingsworth-Lönnberg (1986) with radiosonde (RS)
- Varies with height, latitude, longitude, and month

**zonal- and annual-mean structure**

- **a)** Temperature [K]
- **b)** Specific Humidity [g/kg]
- **c)** Relative Humidity [%]

**horizontal structure**

- **a)** Temperature (5 km)
- **b)** Surface Pressure
- **c)** Specific Humidity (2 km)
- **d)** Relative Humidity (2 km)
Background error correlation

- 1D-Var can utilize **full error covariance matrices** (w/o approx. & simplification)
- NMC method with ECMWF forecasts over a 9-y period
- Varies with latitude and season
1D-Var response to single-level ob pert.

(Useful to understand how information is spread vertically)

- 1D-Var response is non-local and broad, because $B$ and $R$, along with Abel transform and hydrostatic balance op, spread information to wide extents
- $R$ and $B$ (cross error correl. in particular) are important to tackle the underdetermined retrieval problem
- If the observation holds a good number of independent pieces of information, 1D-Var might be able to reconstruct the whole atmosphere
Validation using synthetic data

- A sanity check whether 1D-Var behaves as designed and intended
- A controlled experiment where the truth and $B$ and $R$ are well known, opposed to real-world validation
- For a given RO event, a smoothed GRUAN sounding serves as the truth:
  - 200 observed RO soundings (made available by perturbing the true phase Doppler)
  - 200 first-guess (FG) soundings (by perturbing FG using EOFs of $B$)
  - 40,000 realizations of RO-FG pairs and 1D-Var retrievals
  - Refractivity is taken as the observation (for simplicity’s sake)
Synthetic: RMS error wrt the known truth

- 1D-Var is significantly smaller than FG in the error. The aspect, however, depends on atmospheric condition.
- RO observation is very sensitive to moisture; moisture error reduction dominates in moisture-abundant atmosphere.
COSMIC compared to (1-s) RS92

COSMIC-RS92, GLOBAL, 150 KM / 90 MIN, 2006/05-2015/08

- OLD is close to DRY in the heights above 12 km (as intended)
- NEW agrees significantly better with RS, compared to OLD and DRY
- As a minimum-error-variance estimator, 1D-Var can be smaller than FG in the error (shown later)
Observation types (operators)

- Refractivity (N) and bending angle (BA) are currently admissible (others under testing)
- N versus BA boils down to inverse versus forward Abel transforms
- Use of BA suffers from the fact that it belongs to a space different from N
- **COSMIC 1D-Var**
  - Uses BA through a variational Abel transform
  - Carries out two minimizations, one in the observation space (impact parameter) and the other in the model space (height coord.)
  - Rationale & Support are given by Wee (AMT, 2018)
Comparison of observation operators

- Stats for 7-y COSMIC (FG: NCEP fcst, Verification: ECMWF anal)
- Uses consistent OEs between BA and N (same method applied to a common data set)

OLD (N) Prior 1D-Var version which takes refractivity (N) as the observation
NEW (N) Uses N as the ob
NEW (BA) Uses BA
NEW (VR) Uses BA through Var Abel transform (default)

- NEW(N) v. OLD(N) shows the gross improvement brought by the recent upgrade (other than the ob type)
- NEW(N) v. NEW(BA) relates to the relative effectiveness btw inverse v. forward Abel transforms. Disputable, but N serves better in lower troposphere, because of large fwd modeling error and high volatility of BA there
- NEW(BA) is the most effective
Can 1D-Var add any value to 1\textsuperscript{st} guess?

\textbf{C2 comparison to collocated (operational; GTS) radiosondes}

- 1D-Var 1\textsuperscript{st} guess (FG) used here is NCEP GFS short-term forecasts
- FG and 1D-Var (C2) are compared to nearby radiosondes (RS) on ML (Mandatory p. levs) and ST (Significant T. levs)

1D-Var (C2, heavy solid lines) agrees better with RS than FG (dashed) does
Can 1D-Var add any value to 1\textsuperscript{st} guess? (II)

FG (GFS\textsubscript{FCST}) v. 1D-Var retrieval (C2) wrt ECMWF anal (2019.197 -- 2020.032)

- 1D-Var shows a significantly better agreement than its FG
- This exemplifies the benefit that C2 can bring to weather forecasting/analysis
C2 dependence on FG

- While 1D-Var needs FG to deal with the underdeterminancy and to better use the observation, its dependence on FG is unavoidable and not necessarily bad. It is, though, important to quantify/understand the dependence.

- The dependence is measured by the diff. btw two 1D-Var retrieval sets that make use of different FGs (NCEP and ECMWF), relative to the diff. btw the FGs:

  \[ \text{dep.} = \frac{||C2_{\text{NCEP}} - C2_{\text{EC}}||}{||\text{NCEP} - \text{EC}||} \]

- 1D-Var retrieval sets are closer to each other than the distance between FGs
- The weak FG dependence indicates that 1D-Var retrieval is constrained well by RO
- If RO observation is accurate, the uncertainty of 1D-Var retrieval is lower than FGs
Weak FG dependence enables long-term stability
(COSMIC data record 2008-2014)

GFS departure from ERA-Interim

Past GFS major upgrades (marked A and B) result in discontinuities

- 1D-Var data record is stable enough to show the GFS breaks
- 1D-Var shows no obvious dependence on FG (lower-right p.)
Outlying C2 soundings (not flagged bad)

- Existence of few obvious C2 outliers were informed (S.-Y. Chen, NCU)
- Currently, no QC is applied within 1D-Var on purpose. With QC enabled, a small number of unphysical outliers (< 0.5% for Oct 2019) were detected (considered not bad by lower-level data processing)
- Despite the tiny fraction, these outliers inflate RMSD significantly by up to 60%
- “Robust” statistics, rather than straightforward evaluation, is advised

---

**Standard deviation from ECMWF analysis (w/ NCEP as FG)**

**wetPf2**

**wetPf2 + QC (< 0.5%)**

![Diagram showing standard deviation from ECMWF analysis](image)
Conclusions

• COSMIC 1D-Var (v2) behaves as designed
• Recent changes made to the 1D-Var brought in significant overall improvement
• 1D-Var retrievals show good agreements with radiosonde and global forecasts/analyses
• Dependence on background is weak, enabling long-term stability of 1D-Var data records
• Some issues (new with C2 or unnoticed earlier) remain such as few outliers, dry bias (due to biased observation) in the lowest 2 km, and larger than expected C2 temperature error in the middle troposphere