Data assimilation for atmospheric composition

Nicolas Bousserez
ECMWF
nicolas.bousserez@ecmwf.int

Contributions from: Antje Inness, Richard Engelen, Johannes Flemming, Sebastien Massart, Angela Benedetti, Alessio Bozzo
Why atmospheric composition at an operational weather prediction centre?

- Poor air quality is a major public health issue in many countries.
- Local authorities need accurate and timely information to implement effective air pollution mitigation measures.
- Accurate air quality forecasts require accurate transport models.
- Need to leverage sophisticated data acquisition infrastructures implemented at operational weather prediction centers.
- Atmospheric composition also impacts the weather.
Transforming satellite observations into user-driven services.

Copernicus Atmosphere Monitoring Service
Copernicus Global System

40km horizontal resolution at 60 model levels; two 5-day forecasts at 00z and 12z UTC each day
  • Aerosols (AOD and concentration): biomass burning, dust, sea-salt, sulphate
  • Reactive gases: CO, HCHO, NO₂, O₃, SO₂

9km horizontal resolution at 137 model levels; one 5-day forecast per day (CO₂, CH₄, linear CO)
Data Assimilation Methodology for Atmospheric Composition
Data Assimilation Methodology

Data assimilation for atmospheric composition is in principle no different from NWP data assimilation

\[ J(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i]) \]
Data Assimilation Methodology

Data assimilation for atmospheric composition is in principle no different from NWP data assimilation.

\[ J(x) = (x - \beta \gamma) \]

Control variables
- NWP:
  - vorticity
  - divergence
  - temperature
  - surface pressure (logarithm)
  - specific humidity
- Atmospheric Composition:
  - ozone
  - carbon monoxide
  - nitrogen dioxide
  - formaldehyde
  - sulphur dioxide
  - carbon dioxide
  - methane
  - aerosol mixing ratio

\[ x_i = M_{0\rightarrow i}(x) \]

Atmospheric Composition models
- Chemical module
- GHG module
- Aerosol module

\[ J(x) = \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i]) \]
Data Assimilation Methodology

\[ J(x) = (x - x_b) T_B^{-1} (x - x_b) + \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i]) \]

Atmospheric Composition models

\[ x_i = M_{0\rightarrow i}(x) \]

Control variables
NWP
- vorticity
- divergence
- temperature
- surface pressure (logarithm)
- specific humidity

Atmospheric Composition
- ozone
- carbon monoxide
- nitrogen dioxide
- formaldehyde
- sulphur dioxide
- carbon dioxide
- methane
- aerosol mixing ratio

Chemical Module
- TM5 (CB05)
- 54 species, 126 reactions
- Photolysis, dry and wet deposition

Data assimilation for atmospheric composition is in principle no different from NWP data assimilation

**Chemical Module**

**IFS**

**GHG module**

**Aerosol module**
Data Assimilation Methodology

Data assimilation for atmospheric composition is in principle no different from NWP data assimilation:

\[ J(x) = (x - \bar{x})^T \Sigma^{-1} (x - \bar{x}) \]

where:
- \( J(x) \) is the cost function
- \( x \) is the state vector
- \( \bar{x} \) is the background state
- \( \Sigma \) is the background error covariance matrix

Control variables:
- NWP
  - vorticity
  - divergence
  - temperature
  - surface pressure (logarithm)
  - specific humidity
- Atmospheric Composition
  - ozone
  - carbon monoxide
  - nitrogen dioxide
  - formaldehyde
  - sulphur dioxide
  - carbon dioxide
  - methane
  - aerosol mixing ratio

Atmospheric Composition models:
- Greenhouse Gas Module
  - IFS
  - CHTESSEL
- Chemical module
- GHG module
- Aerosol module

Control variables:
- GHG module
- Chemical module
- Aerosol module

Diagnoses the gross primary production of CO2 by plants and release of CO2 by soil

CH4 comes from prescribed emissions and climatological loss
Data Assimilation Methodology

\[ J(x) = (x) \]

Control variables

NWP
- vorticity
- divergence
- temperature
- surface pressure (logarithm)
- specific humidity

Atmospheric Composition
- ozone
- carbon monoxide
- nitrogen dioxide
- formaldehyde
- sulphur dioxide
- carbon dioxide
- methane
- aerosol mixing ratio

Data assimilation for atmospheric composition is in principle no different from NWP data assimilation

\[ + \sum_{i=0}^{n} (y_i - H_i[x_i])^{T} R_i^{-1} (y_i - H_i[x_i]) \]

Atmospheric Composition models

\[ x_i = M_{0 \rightarrow i}(x) \]

Aerosol bin scheme

12 aerosol-related prognostic variables:
- 3 bins sea-salt
- 3 bins dust
- Black carbon
- Organic matter
- Sulphate

Emissions, dry and wet deposition, sedimentation
Combining the atmospheric composition and NWP models

- Atmospheric composition models can be run coupled to NWP or fully integrated.

IFS

In the IFS the atmospheric composition and NWP models are fully integrated

Data Assimilation Methodology

Data assimilation for atmospheric composition is in principle no different from NWP data assimilation

\[ J(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i]) \]

Data assimilation for the atmospheric composition is ‘strongly coupled’ (see Coupled Data Assimilation lecture)
Data Assimilation Methodology

Data assimilation for atmospheric composition is in principle no different from NWP data assimilation

\[ J(x) = (x - x_b)^T B^{-1} (x - x_b) + \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i]) \]

Data assimilation for the atmospheric composition is "strongly coupled" (see Coupled Data Assimilation lecture)

IFS

Outer loops
- Full non-linear NWP model
- Full atmospheric composition models
- Full set of atmospheric composition variables

Inner loops
- Tangent linear NWP model linearised around outer-loop trajectory
- No atmospheric composition model
- Limited set of atmospheric composition variables treated as tracers

Tangent Linear and Adjoint of observation operators
Observations of Atmospheric Composition
Atmospheric composition observations traditionally come from UV/VIS measurements. This limits the coverage to day-time, cloud-free only. Infrared/microwave are now adding more and more to this spectrum of observations (MOPITT, AIRS, IASI, MLS, MIPAS …)
NRT data coverage for reactive gases

- **Global coverage in a few days (LEO) – fixed overpass time**

- **Atmospheric Composition also use retrievals rather than the direct radiances**

- **Fixed overpass times and daylight conditions only (UV-VIS) -> no daily maximum/cycle**

- **Often limited to cloud free conditions**
Challenges for Atmospheric Composition Data Assimilation
1. Initial vs Boundary Problem

- NWP 4D-Var is mostly defined as an initial value problem. Only initial conditions are changed and model error is relatively small.

- AC modelling depends on initial state and surface fluxes

- Large part of chemical system not sensitive to initial conditions because of chemical equilibrium, but dependent on model parameters (e.g. emissions, deposition, reaction rates, …)
Surface fluxes

An example of all the surface fluxes that need to be accounted for with CO₂
Emission Processes

- Combustion related (CO, NO\textsubscript{x}, SO\textsubscript{2}, VOC, CO\textsubscript{2}):
  - Fossil fuel combustion
  - Biofuel combustion
  - Vegetation fires (man-made and wild fires)

- Fluxes from biogeochemical processes (VOC, CH\textsubscript{4}, CO\textsubscript{2}, Pollen):
  - Biogenic emissions (plants, soils, oceans)
  - Agricultural emissions (incl. fertilisation)

- Fluxes from wind blown dust and sea salt (from spray)

- Volcanic emissions (ash, SO\textsubscript{2}, HBr …)
Example: volcanic eruptions

Both initial conditions and emissions are important to get it right

Flemming and Inness (2013)
Emission Estimates

- Emissions are one of the major uncertainties in modeling (can not be measured directly)
- The compilation of emissions inventories is a labour-intensive task based on a wide variety of socio-economic and land use data
- Some emissions can be “modeled” based on wind (sea salt aerosol) or temperature (biogenic emissions)
- Some emissions can be observed indirectly from satellites instruments (Fire radiative power, burnt area, volcanic plumes)
- “Inverse” methods can be used to correct emission estimates using observations and models – in particular for long lived gases such as CO2 (e.g. Chevallier et al. 2014) and Methane (Bergamaschi et al. 2009)
How to improve?

Better modelling of fluxes and emissions

Modelling of sea salt and desert dust fluxes.
How to improve?

Use near-real-time off-line estimates of emissions

Copernicus Global Fire Assimilation System (GFAS)
How to improve?

Use near-real-time off-line estimates of emissions
How to improve?

Adjust emissions as well as concentrations

\[
J(x) = (x - b_T - 1) + n_y - H_i x_i T_i - 1 (y_i - H_i x_i)
\]

Control variables

- NWP:
  - vorticity
  - divergence
  - temperature
  - surface pressure (logarithm)
  - specific humidity

- Atmospheric Composition:
  - ozone
  - carbon monoxide
  - nitrogen dioxide
  - formaldehyde
  - sulphur dioxide
  - carbon dioxide
  - methane
  - aerosol mixing ratio

Future control variables

Emissions:

- nitrogen dioxide
- carbon monoxide
- carbon dioxide
- methane

All these species are dominated by emissions – changing initial conditions is not enough

Include as a scaling factor adjustment to the prior (inventory-based) emissions

Doubles the computation time
CO₂ Human Emission (CHE) project

- Build an anthropogenic CO₂ emission inversion capability to prepare the future Copernicus CO₂ service.
- IFS control vector has been extended to include emission (CO, NOₓ, CO₂) scaling factors in the 4D-Var minimization.

**Challenges:**
- Current IFS 4D-Var window is 12 hours → long lived tracers (e.g., CO₂, CH₄) requires longer assimilation windows (several weeks to months).
- Current CO₂ observations are not sufficient to constrain the CO₂ source attribution problem (e.g., biospheric vs anthropogenic sources) → need to use co-emitted short lived species information (NO₂, CO) → requires to account for chemical processes in the minimization (not currently available).

- Prior error = 100%
- Prior correlation length = 300km
- Observations: MOPITT CO, IASI CO
- IFS 4D-Var, 2 outer iterations

New IFS capability under development

2018/11/01

increased emissions over eastern Asia

reduced emissions over eastern US

CO emission scaling factor (analysis)
2. Mismatch between modelled and observed variables

\[ J(x) = (x - \frac{1}{b} T - 1 x - b) + \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i]) \]

Only a small subset of all chemical species are observed and therefore included in the control vector. This means the full chemical system is very under-constrained.
2. Mismatch between modelled and observed variables

\[ J(x) = (x - x_b) \] 

\[ + \sum_{i=0}^{n} (y_i - H_i[x_i])^T R_i^{-1} (y_i - H_i[x_i]) \]

Even for those species that are observed, it is often only total column data that is available.
Increment from a single total column ozone observation

Increment created by a single ozone observation of 375 DU, 10 DU higher than background.

Horizontal correlation from the B-matrix that spreads the information from the single observation in the horizontal.
Increment from a single total column ozone observation

Increment created by a single ozone observation of 375 DU, 10 DU higher than background

Standard deviation from the background matrix at the observation location

Vertical profile of the increment at the observation location

Background matrix has a significant impact on the distribution of information

Formulation of the B-matrix is very important for AC
Benefit of profile information – B matrix is not enough!

Ozone hole in CAMS reanalysis: Cross section along 8E over South Pole, 4 Oct 2003

Assimilation with profile data

Assimilation with total column data
3. Lifetime of species & chemical sensitivities

[Diagram showing temporal and spatial scales with different species' lifetimes and mixing times indicated.

After Seinfeld and Pandis [1998]
NO\textsubscript{2} data assimilation

12-hour 4D-Var window

Rapid chemical conversion within the 12-hour 4D-Var window means we cannot link an NO\textsubscript{2} observation at the end of the window correctly to the initial state without a full chemical adjoint.

Partial solution through simple approximation of main chemical reaction

\[
\frac{[NO\textsubscript{2}]}{NO_x} \approx \frac{k[O_{3eff}]}{JNO\textsubscript{2} + k[O_{3eff}]} \]

\[
NO\textsubscript{2} + \text{sunlight} \rightarrow NO + O \\
NO + CH\textsubscript{3}O\textsubscript{2} \rightarrow NO\textsubscript{2} + CH\textsubscript{3}O
\]
Short lived memory of NO2 assimilation

- Large positive increments from OMI NO2 assim
- Large differences between analyses of ASSIM and CTRL
- Impact is lost during subsequent 12h forecast
- It might be more beneficial to adjust emissions (instead of IC)
Minimization with chemical processes

➢ Chemical processes currently only included in the non-linear forward model integration (outer loop).
➢ Solutions:
  o Tangent-linear and adjoint of full-chemistry scheme → too expensive for operational implementation.
  o Simplified chemistry in TL/AD models (e.g., NO/NO2 photochemical equilibrium): fast but may not contain enough information.
  o Use of ensemble information: small ensembles of non-linear forward model trajectories with full-chemistry could be used to explicitly propagate the background error covariance within the 4D-Var window.
  o Hybrid approach: combines simplified chemistry in TL/AD models with ensemble information to propagate the background error covariance (chosen method for the CAMS/CHE emission inversion prototype).
Potential Benefit for NWP
Potential benefit for NWP

- Interactive aerosols: Feedback on dynamics via radiation scheme: First Tegen AER climatology used in radiation scheme, CAMS interim climatology from CY43R3 onwards

- Use of O3 (& other fields) in the radiation scheme: MACC climatologies used

- RTTOV observation operator: Use of O3, CO2 analysis fields to improve the use of radiances sensitive to O3, CO2: model O3 is used, but climatologies used for other tracers (e.g. fixed CO2 value)

- Dynamical coupling with wind/T through TL and AD: turned off

- Multivariate JB: Correlations between tracers and dynamical variables, e.g. O3 and vorticity; correlations between chemical species: univariate
Benefit of AC for NWP: Updated climatology in the radiation scheme

Climatological AOD 550nm distribution CAMS vs Tegen et al 1997

- CAMS interim reanalysis (2003-2018): sources of biomass burning from GFAS, sulphate aerosol precursor from EDGAR 4.1, prognostic for sea salt and dust, revised dust model
- Optical properties recomputed for RRTM spectral bands and for each aerosol type/size bin. Mass mixing ratio as input to radiation
- Vertical distribution following an exponential decay with scale height derived from the CAMS model for each aerosol type. Monthly varying for dust.

Credits: Alessio Bozzo
Improvements to NWP forecast errors

- Change in mass distribution and optical properties -> reduction in SW absorption -> reduction in temperature (positive)
- This is of the order of 0.1K for a bias of the order of 0.3K – it explains at least ~30% of the temperature error.
- Similar for winds at upper levels
Improvements to NWP forecast errors

Too strong monsoon circulation in Indian Ocean in the model leads to too high precipitation over west India.

Revised aerosols affect the circulation and reduce the bias both in the wind circulation and in the precipitation amounts in Summer.
Benefit of AC for NWP: Variable CO2 in radiance assimilation

Reduced AIRS and IASI Bias Correction

- Using modelled CO$_2$ in AIRS/IASI radiance assimilation leads to significant reduction in needed bias correction.
- Small positive effect on T analysis and neutral scores/ small positive impact at 200 hPa T in Tropics
- Stratospheric T in variable CO2 exp more consistent with AMSU-A
- It would be beneficial to replace the fixed value by more realistic values

Engelen and Bauer, QJRMS, 2011
Benefit of AC for NWP: Wind information from tracers

• Prospect to extract wind information from long lived tracers in stratosphere and upper troposphere, e.g. O₃, H₂O, N₂O.
• Similar to cloud track winds but data coverage worse.
• Potential to extract wind info indirectly through TL and AD of tracer advection
• Potential was demonstrated in early studies for H₂O (Thepaut 1992) and O₃ (Daley 1995; Riishojgaard 1996; Holm 1999; Peuch et al. 2000)
• Could compliment existing wind observations and help in areas where there is a lack of adequate global wind profile data
Benefit of AC for NWP: Example of link between wind and tracer

Model equations
\[
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2} \\
\frac{\partial q}{\partial t} + u \frac{\partial q}{\partial x} = 0
\]

Tangent linear equations:
\[
\frac{\partial \delta u}{\partial t} + u \frac{\partial \delta u}{\partial x} + \delta u \frac{\partial u}{\partial x} = \nu \frac{\partial^2 \delta u}{\partial x^2} \\
\frac{\partial \delta q}{\partial t} + u \frac{\partial \delta q}{\partial x} + \delta u \frac{\partial q}{\partial x} = 0
\]

Adjoint equations:
\[
-\frac{\partial \delta' u}{\partial t} - u \frac{\partial \delta' u}{\partial x} + \frac{\partial \delta u}{\partial x} - \nu \frac{\partial^2 \delta' u}{\partial x^2} + \delta' q \frac{\partial q}{\partial x} = 0 \\
-\frac{\partial \delta' q}{\partial t} - \frac{\partial (u \delta' q)}{\partial x} = 0
\]

\(u = u(x,t)\) = wind over periodic domain \([0,L]\) \\
\(q = q(x,t)\) = passive tracer \\
\(\nu\) = diffusion coef. \\
\(\delta u, \delta q\) = perturbations \\
\(\delta' u, \delta' q\) = adjoint variables
Single observation experiments - Ozone and wind increments

3D-Var

4D-Var 9z

4D-Var 12z

4D-Var 15z

Level 20, ≈ 30 hPa

Observation at T0: 4D-Var = 3D-Var

Observation at T3: wind increments

Observation at T6: wind increments

6h assimilation window
Benefit of AC for NWP: Requirements to extract wind info from tracers

- Complete data coverage (3D), frequent observations.
- Accurate observations
- High quality background field
- No bias between observations and background
- Depends on accuracy of TL model compared to full model (better for passive tracers/ long chemical lifetime) => E.g. extracting wind information from O$_3$ is more difficult in the tropics and summer hemisphere where photochemical lifetime is shorter
- Studies have looked at this in idealized experiments (e.g. Daley 1995; Riishojgaard 1996; Peuch et al. 2000; Allen et al. 2013, 2014) focussing on long lived tracers O$_3$, H$_2$O, N$_2$O and found positive impact for perfect observations.
- Few studies used real data (e.g. MLS O3, Semane et al. 2009) and positive results are less clear
Benefit of AC for NWP: Example from ERA-Interim

The stratosphere is not well constrained by observations:

- Ozone profile data generate large temperature increments
- 4D-Var adjusts the flow where it is least constrained, to improve the fit to observations

=> IFS O3 analysis is completely uncoupled now

GOME 15-layer profiles (~15,000 per day)
SBUV 6-layer profiles (~1,000 per day)

Ozone increments at 10S

Large systematic O3 increments

Associated Temp increments

D. Dee
Summary
What we have seen today…

• Basic Data Assimilation theory is the same
• Particular challenges related to DA for atmospheric composition
  – Boundary conditions (emissions) as well as initial conditions
  – Mismatches between modelled and observed variables
  – Fast reactions and short life time of some species
  – Cost of using chemical processes in the minimization (i.e., in the TL/AD models)
• Atmospheric composition has the potential to improve various aspects of NWP
• CAMS system produces useful Atmospheric Composition forecasts and analyses, freely available from atmosphere.copernicus.eu
More information about the environmental monitoring activities at ECMWF and how to access the data can be found on:

atmosphere.copernicus.eu
References: Reactive gases


References: Reactive gases


References: Reactive gases


References: Aerosols


References: Aerosols


References: Greenhouse gases


A. Agusti-Panareda; S. Massart; F. Chevallier; G. Balsamo; S. Boussetta; E. Dutra; A. Beljaars
A biogenic CO2 flux adjustment scheme for the mitigation of large-scale biases in global atmospheric CO2 analyses and forecasts. ECMWF Technical Memorandum, no 773, 2015
http://www.ecmwf.int/en/elibrary/technical-memoranda


References: Greenhouse gases


Massart et al. (2016)  


References: Fires


References: General
