Working group reports for the Workshop on parameter estimation and inverse modelling

Introduction

Under the umbrella of the EU-funded projects GEMS, MACC, and MACC-II, ECMWF has extended its data assimilation and forecasting system to include atmospheric composition. Analyses and forecasts of reactive gases, greenhouse gases, and aerosol are now routinely provided on a daily basis. At the same time the EU-funded GEOLAND and GEOLAND2 projects have extended the ECMWF land surface model to include a simple representation of the carbon cycle. As part of the European Copernicus framework it is envisaged that this atmospheric composition data assimilation and forecasting system will become fully operational to provide daily global forecasts complementing ECMWF's meteorological forecasts.

The modelling and data assimilation of atmospheric composition differs in several ways from the modelling and data assimilation for numerical weather prediction, such as the atmospheric chemistry and the available observations. However, another important aspect is the significance of the boundary conditions. Anthropogenic emissions, greenhouse gas surface fluxes, wild fire emissions, and volcanic eruptions form a significant part of the assimilation and forecasting problem, which can therefore not really be treated as an initial condition problem with a (almost) perfect model. Current practice is to prescribe these boundary conditions based on emission inventories, off-line carbon flux models, off-line fire detection systems, and ad-hoc definition of volcanic emissions. However, for a future operational forecasting system better solutions should be found in order to minimize the model error. This is especially important with the introduction of long-window 4D-Var.

ECMWF organised a workshop on parameter estimation and inverse modelling for atmospheric composition from 22 to 24 October 2013. The aim of the workshop was to explore the options to optimally define the boundary conditions in a near-real-time 4D-var data assimilation system. The first day and a half was spent with very interesting presentations from mostly external experts providing different angles on the challenge of constraining boundary conditions for atmospheric composition in an NWP environment. After that two working groups were formed to discuss in more detail the various aspects, focussing on surface modelling and parameter estimation on the one hand, and constraining emissions from volcanoes, fires, and anthropogenic sources on the other hand. Several recommendations were then presented in the final plenary. The working group reports as well as short abstracts for some of the presentations are combined in this workshop report.

ECMWF would like to thank all participants for their contributions to a successful and stimulating workshop.

Summary of Working Group 1: "Improving the 5-day CO2 forecast within the operational Copernicus service"

Wouter Peters (Wageningen University), Anna Agusti-Panareda (ECMWF) and Sébastien Massart (ECMWF)

Summary of recommendations

The aim of the working group 1 was to discuss how to improve the 5-day CO2 forecast in the global MACC system. This CO2 forecast is a new product offered by MACC-II which can be interesting to several users. These users range from the satellite retrieval community to the wider observational community in terms of continuous monitoring for the analysis and control of observations and planning of field experiments. Within ECMWF there is also an interest from the radiative transfer (both for forecasting and assimilation of satellite radiances) to the evaluation of boundary layer parameterization and land-surface feedbacks (e.g. vegetation and water cycle). Finally, this product is useful for the Carbon cycle research community as it already provides carbon fluxes, and in the near future it will also provide both their uncertainties and the covariance between the carbon flux errors and the surface concentration errors, as well as boundary conditions for regional modelling.

The working group reviewed two main themes that can contribute to the improvement of the forecast. Namely, (i) improving the availability of both ground-based and remote sensing observations of CO2 in near real time (NRT, note that here NRT means within 3 days of observation), and (ii) using new methodologies that will allow an adjustment of the CO2 fluxes, as well as the atmospheric CO2 mole fraction. The pros and cons of different methods and their feasibility within the Numerical Weather Prediction (NWP) framework of ECMWF were discussed.

The main recommendations of the working group to ECMWF and MACC are to:

1. Enhance efforts to create bi-lateral agreements with data providers. High level efforts are recommended in order to aim for a smooth data sharing capability similar to the WMO-GTS stream.
2. Exploit the coupling between the carbon and water cycles in order to evaluate the Carbon module and the land-surface model.
3. Continue with the plans to have a two-step approach where (i) the fluxes are bias corrected by applying re-scaling coefficients based on the climatological budget of optimized fluxes and (ii) the atmospheric CO2 mole fraction is further adjusted by assimilating CO2 satellite retrievals.
4. Explore the CO2 flux adjustment within the short window data assimilation at ECMWF in conjunction with the ensemble data assimilation (EDA) in order to be able to estimate the background error covariances that link the atmospheric concentrations to the fluxes.

All these recommendations should be addressed in parallel as they are all crucial for the improvement of the CO2 forecasting system. The availability of data not later than 3 days after observation will ensure the possibility of having a near-real time analysis which does not currently exist. This analysis of the carbon fluxes and concentrations will provide better initial conditions for the CO2 forecast. At the same time, the improvement of the land-surface carbon model will lead to a better background state for the data assimilation.

Observations

Currently there are no CO_2 observations provided in NRT, except for a few ICOS surface stations. This is the main limitation for not being able to provide a NRT flux estimation and atmospheric CO_2 analysis which is one of the key components for the improvement of the CO_2 forecast. Therefore, it is critical to work towards the establishment of links with the different observing networks in order to get the data as soon as possible after their measurement time.

There are two main streams of observations that could be used in the data assimilation of CO_2 in NRT: surface in situ observations and remote sensing retrievals from satellite and ground-based stations. Measurements of the surface CO_2 mole fractions have the highest accuracy, but do not have a global coverage. In order to achieve a coverage that would be sufficiently large to have an impact on the global CO_2 analysis, all the observations from different observing networks should be used. There are many data providers, e.g. ICOS, Environment Canada, NOAA, CMA, CSIRO, JMA, INPE, etc. High level discussions between ECMWF/MACC and these data providers are necessary in order to study the possibility of creating a system to share the data such as the WMO-GTS stream.

XCO_2 and XCH_4 products retrieved from GOSAT and IASI radiances are already used with a 6 month delay in the delayed-mode analysis of atmospheric CO_2 and CH_4 at ECMWF. There is also now an effort within MACC to try to make these products available with only a 2-day delay. These will offer a better coverage than the surface observations. However, the retrievals are more susceptible to biases and larger uncertainties than the surface observations. Therefore, the use of both data streams is not redundant, but complementary, by providing a good coverage and the possibility of correcting for the biases operationally. Retrievals from Fourier Transform Spectrometers of the TCCON network are widely used by some providers of CO_2/CH_4 retrievals in order to evaluate and bias correct their products. It would also be useful to have these data in NRT so to continuously monitor the quality and bias of the CO_2 analysis and forecast in NRT. The new upcoming satellite missions (e.g. OCO-2, TANSAT, MERLIN) are going to provide an increase in the data coverage. It is expected that the quality of the CO_2 analysis will also be improved, provided those satellite retrievals are made available in NRT.

The two streams of observations discussed above provide information on the surface and total column concentrations, thus leaving a data gap regarding the CO_2 vertical profile. There is a clear need to have more information about the vertical structure of CO_2. Aircraft observations would be very useful to evaluate the model and analysis. However, it is often difficult to access such data. The development of AirCore observations is promising. Finally, observations of parameters other than atmospheric CO_2, such as CO_2 fluxes, LAI, fPAR, biomass, soil moisture, albedo, etc. could also be useful to evaluate the land-surface carbon cycle.

Correcting the CO_2 surface fluxes

There was a consensus in the working group on the fact that flux adjustments of CO_2 are needed to get a better forecast. This is particularly important in order to reduce the background atmospheric CO_2 biases before any assimilation of surface observations, and similarly to avoid accumulating global biases in the long cyclic forecasts. Until recently, flux adjustments have been performed within flux inversion systems using mainly long windows (from a few weeks to decades). Long windows ensure mass conservation at large scales and they can provide larger-scale budgets which can be useful for carbon cycle process research. The use of long windows can compensate partially for the sparsity of observations. However, this approach is not feasible within the NWP setting.

An intermediate approach has been explored within MACC at ECMWF, where the 10-day flux budget for different vegetation types is adjusted based on a climatology of optimized fluxes from a long-window flux inversion system also provided by MACC. This flux correction reduces the biases in the CO2 flux budget and lead also to a reduction in the atmospheric global CO2 bias of the cyclic forecasts. The second step is to then use this corrected forecast as a background state to assimilate the CO2 observations by adjusting only the atmospheric concentrations. This two-step approach is showing some promise, and there was a consensus on the need to continue pursuing this approach. This approach offers the potential to assimilate more efficiently the ICOS and other surface observations in NRT, after having reduced the background bias with the flux correction.

Recently, Eugenia Kalnay's group has shown that using short windows (few hours) to adjust both CO2 fluxes and concentrations is also possible and beneficial compared to longer windows. The main advantage is that the atmospheric CO2 signals coming from the fluxes do not mix. The main disadvantage is that mass conservation is not guaranteed if CO2 mole fractions are also adjusted. Also, for this approach to work, it is important to have data richness, such as satellite coverage. Because the short windows would be suitable for an operational/NWP context, ECMWF/MACC is strongly advised to try this approach. The developments in the EDA would be needed for estimating the background error covariance matrix. In this context, another possibility would be to use EnKF, if it is available.

Estimating forecast uncertainty

An estimate of the forecast uncertainty would be a valuable complementary product to the CO2 forecast. In order to provide the uncertainty of its weather forecast, ECMWF combines the deterministic forecast with an ensemble of perturbed forecasts (EPS). The spread of this ensemble is an estimation of the uncertainty associated with the forecast.

In parallel to the EPS, perturbed forecasts are also used individually as the background state for one particular member of an ensemble of analyses (EDA). This ensemble is used to estimate the uncertainty of the background state as well as the spatial correlation of the background error and the correlation between the prognostic variables of the system.

The EDA methodology has started to be used in the MACC system benefiting from the expertise at ECMWF. For the long-lived greenhouse gas analyses, the fluxes have also been perturbed. This allows estimating the covariance of the flux errors and the correlation between the flux errors and the atmospheric concentration errors at the surface. The information one could extract from these correlations is valuable for example for flux inversion systems and should be provided to the community. MACC is also encouraged to provide an ensemble of CO2 forecasts based on the ECMWF's EPS.

Summary of Working Group 2: "Emissions"

Andreas Stohl (NILU) and Johannes Flemming (ECMWF)

Summary of recommendations

The aim of the working group "Emission" was to discuss how the emissions used in the global MACC system can be improved, in particular by parameter estimation based on observations. The MACC system makes forecast of atmospheric composition and assimilates satellite retrievals to improve the realism of the tracer fields. The working group dealt mainly with anthropogenic emissions of air pollutants such as reactive gases and aerosols, and less with the long-lived green-house gases.

The working group reviewed methods to improve emissions by (i) further refining emission inventories and models and (ii) including emission corrections in the 4D-VAR data assimilation algorithm. Further, the working group summarised the current status and the future availability of atmospheric composition observations, which could be assimilated by the MACC system.

The main recommendations of the working group are:

1. MACC should enhance efforts to include emission corrections based on NRT observations in the data assimilation algorithm. The emissions are essential for the quality of atmospheric composition forecasts. They are in many cases more important than the initial conditions.
2. Modelling of the emissions in the global MACC system should be refined with respect to the temporal variability (diurnal cycle, weekly cycle) and dependence on meteorological situation (temperature). As the resolution of the global model is further increased, approaches developed for regional models become similarly important in the global model.
3. The MACC system and ECMWF's NWP forecasting system should be prepared to handle a volcanic eruption with a global or strong regional impact. An eruption of the magnitude of the Pinatubo eruption in 1991 is assumed to be "statistically" due.

The participants of the working group did not suggest prioritising recommendation 1) or 2). Improving bottom-up or process-oriented emission modelling and inverse modelling/data assimilation should be pursued together because (i) each approach alone may not lead to significant progress, (ii) they improve different aspects and (iii) a better a-priori model is beneficial for the inversion or the data assimilation. In the following section we give a more detailed overview of the discussion.

Emissions optimisation as part of data assimilation with the MACC system

Including emissions parameters in chemical data assimilation in regional models has been successfully carried out with ensemble techniques (e.g. Kalman Filter in EUROS-LOTOS) and variational methods (e.g. 4D-VAR in EURAD). Any implementation will depend on the specific detail of the existing data assimilation method and no recommendation was given which of the methods is better suited for the problem.

It should be noted that including emission parameters in the data assimilation system should be distinguished from invers modelling of emissions to improve inventories of long-lived constituents. The latter uses assimilation windows of month to years and does not aim at the improvement of the initial conditions.

Including the emission in the control vector will lead to an improved representation of the emission for the period of the assimilation window, which can also be applied for the forecast period. An assimilation window of 24 h seems a good compromise between the amount of data available to constrain the model and the temporal validity range of the emission optimisation.

It was recommended not to include the emission fluxes as field of (independent) grid points in to the state vector. It seem more promising to introduce model parameters that describe the emission process for certain source groups (i.e. traffic) over a larger area.   This requires that the current approach of providing pre-calculated emission fluxes at grid point has to be altered to be able to simulate the emissions based on source groups within the IFS.

The use of adjoint model formulation could be helpful to identify the most sensitive parameters. The formulation of the background error for the emission parameters remains a challenge. The result from Ensemble / Kalman Filter approaches could be used to obtain basic estimates for the background error statistics to be used in 4D-VAR.

Improving temporal variability of emissions

Most of the working group members had a background in regional air pollution modelling. They were surprised by the lack of temporal variability of the anthropogenic fluxes in C-IFS and many other global Chemistry Transport Models. The anthropogenic C-IFS emissions for reactive gases and aerosol only vary with season. It was recommended to introduce diurnal and weekly emission factors on the levels of the source groups, i.e. SNAPs (Standardized Nomenclature for Air Pollutants).

The temperature dependence of fossil fuels combustion for heating (heating degree day concept) or evaporation of volatile organic compounds should also be taken into account. Diurnal emission factors could also be introduced for biomass burning and biogenic emissions.

The emission of emerging economies such as China and India are changing fast. It was therefore recommended that MACC should do more to update the global anthropogenic emissions for the near-real-time runs.

An important reason why global CTMs are less complex than regional CTM with respect to the temporal variability of the emissions is the fact that the temporal variability of the anthropogenic emissions is less well known in regions outside Europe and North-America. Further, changing anthropogenic emissions according to temperature may lead to values which are not in agreement with official emission inventories. However, as the spatial resolution of the global chemistry models is increased more attention should be given to the temporal variability of the emissions in the global simulation.

Observations

Satellite retrievals of atmospheric composition are assimilated in the global MACC system. Regional models predominately assimilate the in-situ observations of the air quality networks. Some of the satellite retrievals have a good availability in near-real-time whereas the NRT provision chain of the AQ observations from the different European networks is more fragmented. However significant improvements have been made in the past years. The working group participants stressed that it is important not only to sustain the AQ in-situ network but also the research type observations from networks such as ACTRIS and ICOS.

To improve emissions estimates with data assimilation in global models, satellite retrievals with high sensitivity to surface concentrations and high temporal resolution (GEO Sentinel) are required.

It was recommended that MACC scientists work on the assimilation of profile observations in the global and regional models. Lidar observations and IAGOS/MOZAIC are profile observations with the potential to complement total columns retrievals from satellite instruments. Observing system sensitivity experiments should be carried out to identify the impact of the assimilation of profile observations.

Of special importance for the global data assimilation system are retrievals with vertical information such as MLS and SAGE retrievals. For example, the assimilation of TES ozone retrievals and MLS HNO3 retrievals together with lightning observations to constrain NOx lightning emissions was suggested as an interesting project to better exploit synergies between different retrievals.

It is of great concern for the MACC work that MLS is close to the end of its nominal lifetime and no similar instrument is envisaged to fly on-board future satellite missions.

The working group recommended improving the NRT availability and quality of satellite products used to monitor volcanic eruptions. For example, volcanic ash retrievals from SEVIRI and AIRS should be produced in operational mode. Plume height information in SO2 retrievals could be inferred from a multispectral approach that combines UV with IR observations.