Downscaling in the Context of Data Assimilation

Juruš, Pavel
2006

Downscaling in the context of data assimilation

Post-Graduate Student:
MGR. PAVEL JURUŠ

Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2

182 07 Praha 8, CZ

Czech Technical University in Prague
Faculty of Transportation Sciences
Konviktská 20

110 00 Praha 1, CZ

jurus@cs.cas.cz

Supervisor:
DOC. ING. EMIL PELIKÁN, CSC.

Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
182 07 Praha 8

Czech Republic

emil@cs.cas.cz

Field of Study:
Engineering Informatics in Transportation and Telecommunications

This work was supported by the Grant Agency of the Academy of Sciences of the Czech Rep. (grant No. 1ET400300414, framework "Information Society") and by the Institutional research plan AV02 10300504 Computer Science for the Information Society "Models, Algorithms, Applications".

Abstract

Feasibility of common framework for the data assimilation and downscaling methods is investigated and discussed. Assimilation of measurements from ground level ozone stations with limited representativity to chemistry transport model is selected as an example and first results with simple decomposition to the local and background component are presented.

1. Introduction

Traditionally, data assimilation in environmental models is developed in the context of single scale. Resolution of a numerical model is given by the time discretization and grid spacing. Our typical objective is to find the 'analysis', which is as good estimate of reality as possible. In modern data assimilation techniques we usually combine information given by numerical model and by observations, as well as information about statistics of model and observational errors. Resulting analysis is defined on grid of numerical model and can be used, e.g., for further time integration or as initial and/or boundary condition for another coupled model.

Our efforts go beyond classical data assimilation – we don't want to have an estimate of model state as good as possible, but we also want to use information about model, observations and their respective errors for improved local predictions. This subject is often treated by downscaling methods or the model output statistics, but while some of these methods can utilize current observations, they don't use them to improve prediction of (coarse) numerical model and they don't use coarse model error statistics.

Local prediction is useful in number of cases. Here are some examples:

1. Wind direction and speed in location of wind farm or even specific wind turbine.
2. Prediction of road ice in given location.
3. Prediction of pollutant concentration in a specified place in an urban area.
4. Short-term prediction of wind field and of tracer dispersion in the vicinity of accidental release during dangerous goods transport.

All four cases are very useful applications of enviromental modelling that show several common features. We are interested in local properties of atmosphere and locality is of essential importance in these cases. While mesoscale NWP (Numerical Weather Prediction) models and/or CTM (Chemistry Transport Models) are the best sources of information available, they lack required space resolution. We need another model on top of mesoscale model(s) whose role is to downscale the prediction to the required location and resolution.

Models used for the downscaling vary depending on application and available data. The most widely investigated case of model output statistics is perhaps the first case above and models used in literature range from purely statistical regression and neural networks models to fine resolution CFD (Computer Fluid Dynamics) models.

Important aspect of aforesaid problems is that the finer the scale is, the more relevant is to predict the value *together* with its confidence interval or even to predict the whole probability distribution for non-gaussian quantities (e.g., road ice). Since the predicted quantity can be highly sensitive to the state of atmosphere (in atmospheric chemistry for example) and we lack the space averaging that helps in mesoscale models, the prediction errors might be very large in some cases. This means that sometimes the information about prediction error is even more important than the value itself.

Data assimilation framework gives us two basic benefits for downscaling. It gives us information about mesoscale state error statistics – this is especially straightforward for data assimilation based on ensemble methods. The simplest approach would be to perform downscaling for each ensemble member and regard the output as a sample from downscaled prediction. This approach is perhaps oversimplified because the resulting ensemble will lack information about error statistics of downscaling model, but there should be a remedy to this problem. The second benefit of data assimilation framework is the possibility to feed back the measurements into either mesoscale or downscaling model. This possibility is perhaps less pronounced for mesoscale NWP models where the analyses usually use so many various observations and quality is so high that assimilation of local observation might even deteriorate rather than improve the prediction. Usefulness of assimilation together with downscaling in mesoscale air quality model will be demonstrated in the next section and other possibilities of data assimilation also in downscaling model will be discussed in conclusion.

The first of listed cases is the example where a number of classical methods of downscaling is being used in practice. Assimilation of measurements (of wind or power production) back to the models is probably not important. Prediction of power production for a number of spatially distributed wind turbines can even help to cancel local wind turbulation and improve the forecast skill. Prediction is useful even without confidence intervals but probabilistic forecast would be certainly an enhancement.

Importance of probabilistic forecast is the main difference between the first and the second case. We are definitely interested in probability of road ice and a statistical model based on meteorological precursors could be appropriate solution. The design of a statistical model will depend on availibility of historical measurements and it could take into account the uncertainity in meteorological data.

The third case is an example when it makes sense to assimilate local measurements back into the mesoscale model. Exact method and the first results are selected as a demonstration of downscaling in data assimilation framework and will be main topic of the next section.

Final example is the most difficult among listed cases. There is usually not enough information to construct either deterministic or statistical downscaling model. There is no fine resolution model of terrain for CFD modelling and there is no history of model and observation values for statistical models – at least not at time of occurrence of an accident. Classical downscaling methods are therefore not applicable in this case. Design of model for local prediction in this case is a matter of further investigation.

2. Local observations in the air quality data assimilation

Data assimilating systems for tropospheric ozone modelling and prediction are currently being proposed and developed [1, 2]. The most common source of observations for air quality data assimilation are ground level stations that measure in situ concentrations of important pollutants. However, the spatial representativity of measurements differs greatly from station to station and can vary from order of 10^2m to 10^5m. Current practice is therefore only to assimilate the stations with the largest representativities – usually rural background stations, since only those stations have measurements aplicable on the scale of CTM. On the other hand most of stations are in urban areas where the air quality has the biggest impact and where mesoscale models are not suitable for modelling and forecasts of pollutant concentrations.

We have proposed and evaluated a data assimilating system for tropospheric ozone prediction with assimilation of rural background stations [3]. In this contribution we propose how to enhance system and how to allow it assimilate stations with smaller representativity and also how to improve local predictions of ground level ozone [4]. A test example of simple downscaling in data assimilation framework is also presented.

Our version of data assimilation is based on ensemble Kalman filter in square root formulation with localization and inhomogenous model error representation [5, 6]. Traditional discretized stochastic model of atmosphere evolution can be written as:

$$\mathbf{x}_k^t = \mathcal{M}(\mathbf{x}_{k-1}^t) + \nu_k \tag{1}$$

$$\mathbf{y}_k = \mathcal{H}(\mathbf{x}_k^t) + \varepsilon_k \tag{2}$$

where x_k^t is vector representing (inaccessible) discretized true state of atmosphere in timestep k on model grid, $\mathcal{M}$ is model of atmosphere used for time integration of state equation and ν_k is model error in timestep k. y_k is the vector of observation in timestep k, $\mathcal{H}$ is observation operator connecting observation and model state space and ε_k is its error in timestep k. ε_k is usually described as instrumental and representation error.

Observation operator in continuous space for a single observation $\mathbf{y}_i$ has usually the form [7]

$$y_i = \int_{\mathbb{R}^3} h(x) s(x, y_i) dx + \varepsilon(y_i) \tag{3}$$

where h is function connecting observation with physical space, and $s(x, y_i)$ aperture function depending on type of observation. For in situ observation is $s(x, y_i) = \delta(x - y_i)$ the Dirac delta function. For remote instruments (e.g. radars, lidars, satelites) implies $s(x, y_i)$ weighted area averaging over its support. For discretizated version it means that $\mathcal{H}$ is usually weighted average over values in some set of gridpoints – set of the nearest neighbours for in situ measurements or a larger set for remote observations.

We want to modify the equation (2) of stochastic model to employ a more general downscaling model instead of observation equation. The basic idea is to decompose ozone concentration to mesoscale component and local component. Decomposition is based on the assumption that part of ozone concentration have origin in long range transport of precursors and mesoscale weather conditions, the second part is added or substracted depending on local emmisions, deposition and other sub-mesoscale effects. The local component was estimated by very simple statistical model as 7-day moving average of the model update for the given hour of day – this reflects sometimes a strong diurnal cycle exhibited in ozone concentrations or a persistent bias. Modified stochastic model of atmosphere is now:

$$\mathbf{x}_k^t = \mathcal{M}(\mathbf{x}_{k-1}^t) + \nu_k \tag{4}$$

$$\mathbf{y}_k = \mathcal{H}(\mathbf{x}_k^t) + \mathbf{z}_k^t + \varepsilon_k \tag{5}$$

where the local component of ozone $\mathbf{z}_k^t$ is modelled as

$$\mathbf{z}_k^t = \frac{1}{7} \sum_{i=0}^{6} \mathbf{d}_{k-24i}^t + \eta_k \tag{6}$$

where $\mathbf{d}_k^t$ is model residual:

$$\mathbf{d}_k^t = \mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^t) \tag{7}$$

We focus on testing the feasibility of the approach and do not try to find optimal solution for $\mathbf{x}$ and $\mathbf{z}$ simultaneously. We simplify solution by first estimating the local component $\mathbf{z}$ at the location of observations and then substituing our estimate into equation (5) and solving classical filtering problem. The solution gives us an estimate of mesoscale state. We want to estimate a vector of target variables $\mathbf{q}_k^a$ (as e.g. local ozone concentrations, probability of icing in specified point). For our testcase concerning ozone concetrations it is simply a sum of mesoscale analysis and local component $\mathbf{q}_k^a = \mathcal{H}(\mathbf{x}_k^a) + \mathbf{z}_k^a$, where superscript a denotes analysis. For linear observation operator and update equation for Kalman filter we can write:

$$\mathbf{q}_k^a = \mathcal{H}(\mathbf{x}_k^a) + \mathbf{z}_k^a = \mathcal{H}(\mathbf{x}_k^f) + \mathcal{H}(\mathbf{K}_k(\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^f) - \mathbf{z}_k^a)) + \mathbf{z}_k^a \tag{8}$$

Superscript f denotes forecast, K_k is Kalman gain matrix and $\mathbf{K}_k(\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^f))$ is analysis increment.

Forecast of mesoscale component is done by the NWP and CTM couple, and forecast of local component is done by persistent model, i.e. it repeats 24-hours old value. Index p denotes prediction horizon $(1 \leq p \leq 24)$, and $\mathcal{M}^p$ means p successive applications of atmosphere model $\mathcal{M}$.

$$\mathbf{q}_{k+p}^f = \mathcal{H}(\mathbf{x}_{k+p}^f) + \mathbf{z}_{k+p}^f = \mathcal{H}(\mathcal{M}^p(\mathbf{x}_k^a)) + \mathbf{z}_{k+p-24}^a \tag{9}$$

Inspired by the equation (8) one could also alternatively try to predict analysis increment – in this case also by a simple persistent model:

$$\mathbf{q}_{k+p}^f = \mathcal{H}(\mathcal{M}^p(\mathbf{x}_k^a)) + \mathcal{H}(\mathbf{K}_{k+p-24}(\mathbf{y}_{k+p-24} - \mathcal{H}(\mathbf{x}_{k+p-24}^f) - \mathbf{z}_{k+p-24}^a)) + \mathbf{z}_{k+p-24}^a \tag{10}$$

3. Testcase and results

Testcase configuration was chosen to be as close as possible to operational system MEDARD [8] in the Institute of Computer Science in Prague, because one of the goals is to make data assimilation usable in operational mode. Selected NWP-CTM model couple was MM5 (PSU, NCAR) and CAMx (Environ) with NCEP FNL analyses as initial and boundary conditions. Model grid is identical to the outer domain of MEDARD system and has 80×65 cells with 27 km horizontal resolution, covering western and central Europe. CAMx model has 12 vertical levels and uses SAPRC-99 [9] chemistry mechanism.

Ozone epizode modelled took place in 24–29 June 2001. Two previous weeks were used for model calibration and spinup. We assimilated observations of $\mathrm{O_3}$, NO, and $\mathrm{NO_2}$ from the Airbase database [10]. 640 stations was selected for the assimilation – among them also mountain stations and urban background stations that are usually not assimilated due to representativity and/or bias problems.

We wanted to stay close to prospective operational mode and 24-hour tropospheric ozone prediction was made for the analysis at 5 p.m. Typical behaviour of local forecast compared to observation on station is illustrated in figures 1–5.

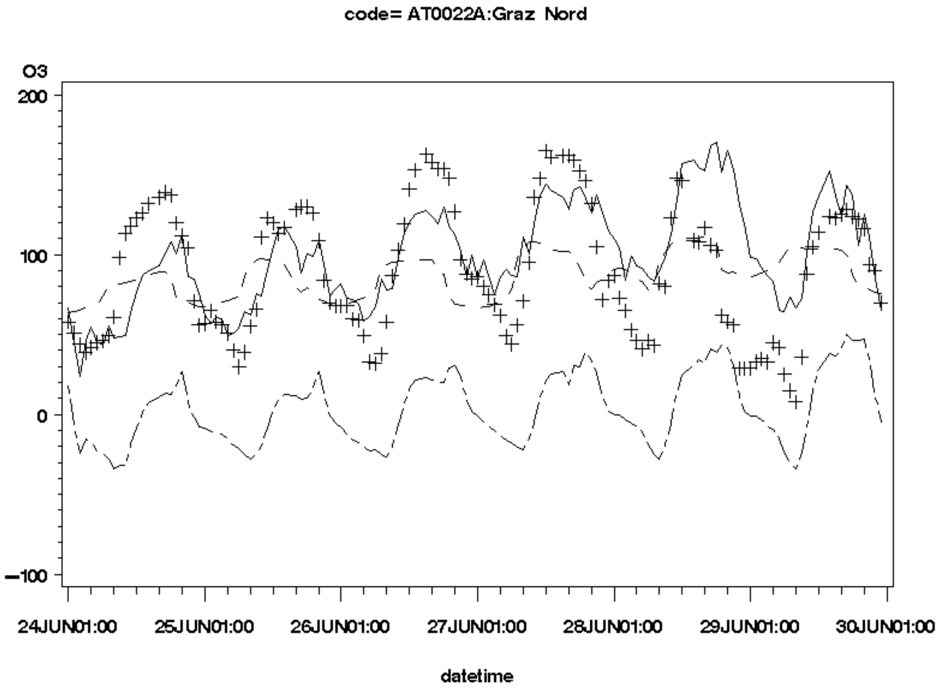

Figure 1: Observations and forecasts for selected station (Graz, Austria). Crosses: observations, dash-dot curve: estimate of local component, dash-curve: mesoscale model ($\mathcal{H}(\mathbf{x}^f)$), solid curve: target variable $\mathbf{q}_k^f$ containing term for analysis increment prediction (equation (10)).

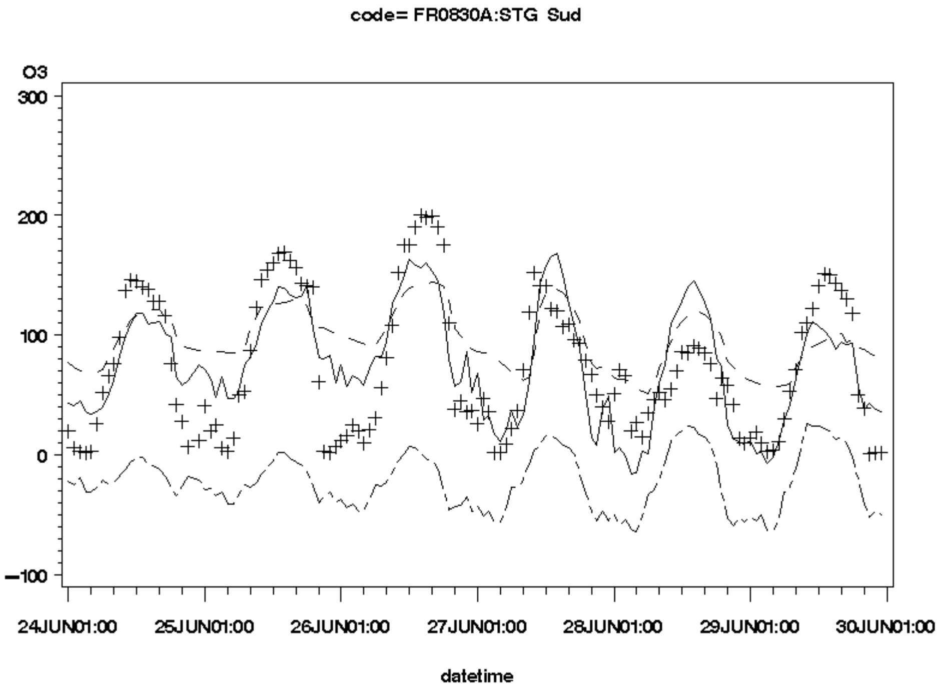

Figure 2: Observations and forecasts for selected station (Strasbourg, France). Crosses: observations, dash-dot curve: estimate of local component, dash-curve: mesoscale model ($\mathcal{H}(\mathbf{x}^f)$), solid curve: target variable $\mathbf{q}_k^f$ containing term for analysis increment prediction (equation (10)).

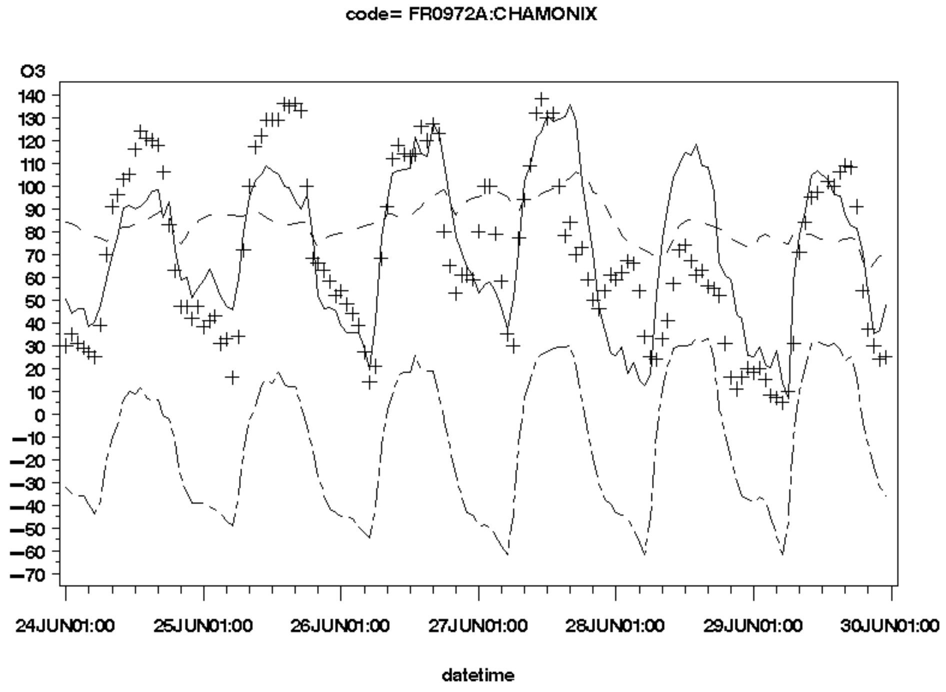

Figure 3: Observations and forecasts for selected station (Chamonix, France). Crosses: observations, dash-dot curve: estimate of local component, dash-curve: mesoscale model ($\mathcal{H}(\mathbf{x}^f)$), solid curve: target variable $\mathbf{q}_k^f$ containing term for analysis increment prediction (equation (10)).

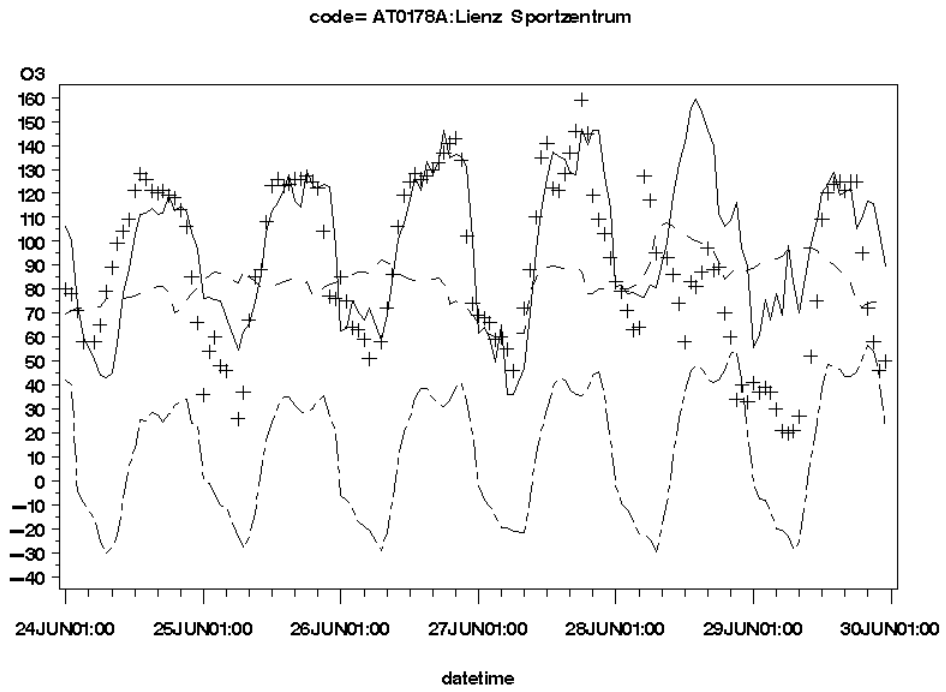

Figure 4: Observations and forecasts for selected station (Lienz, Austria). Crosses: observations, dash-dot curve: estimate of local component, dash-curve: mesoscale model ($\mathcal{H}(\mathbf{x}^f)$), solid curve: target variable $\mathbf{q}_k^f$ containing term for analysis increment prediction (equation (10)).

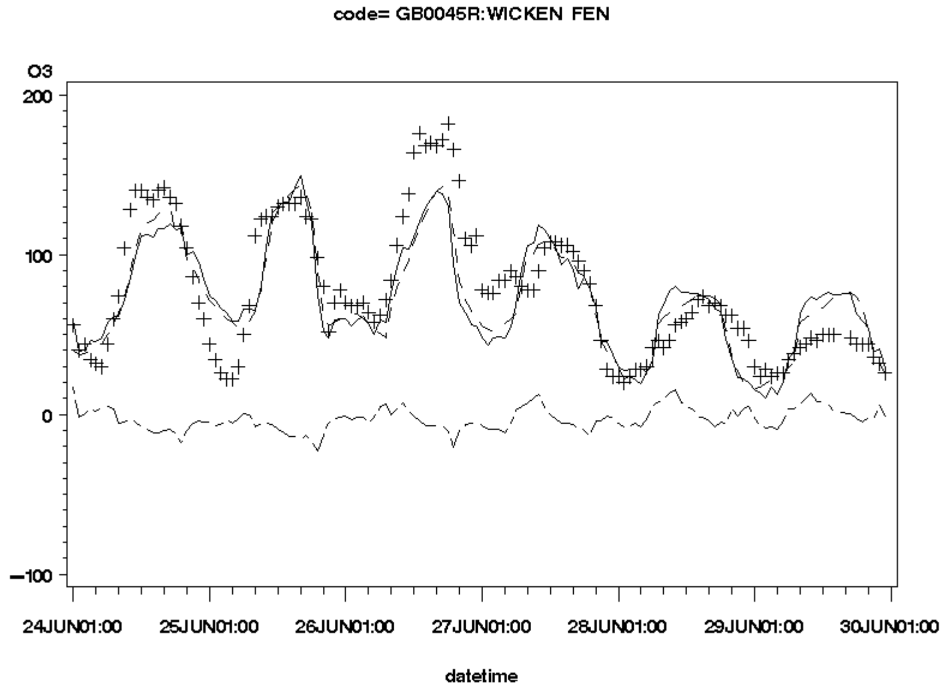

Figure 5: Observations and forecasts for selected station (Wicken Fen, UK). Crosses: observations, dash-dot curve: estimate of local component, dash-curve: mesoscale model ($\mathcal{H}(\mathbf{x}^f)$), solid curve: target variable $\mathbf{q}_k^f$ containing term for analysis increment prediction (equation (10)).

Figures 1 and 2 belong to urban stations with strong diurnal cycle. This cycle is not sufficiently described by mesoscale model. Model of local component rectifies this situation and local component increases with the beginning of the ozone episode. Figures 3 and 4 represent mountain stations with high orographic error and probably also inadequate traffic emmisions. Mesoscale model completely lacks the strong diurnal cycle that must be modelled by downscaling model. High persistence of simple moving average model causes worse performance in detection of the end of episode in the case of Lienz station. Figure 5 belongs to background rural station that is already accurately modelled by mesoscale model. Local component is negligible in this case.

First plot in figure 6 shows that forecast combining mesoscale model and downscaling model easily outperforms pure mesoscale model in prediction of ozone concentrations in location of measurement stations. Second boxplot shows that also classical statistical downscaling even with very simple statistical model brings some improvement over the pure mesoscale forecast. Improvement is seen in last two days of episode but not elsewhere. Last plot compares errors of classical statistical downscaling with statistical downscaling inside of data assimilation framework. The latter method is better in all days except the last one.

4. Conclusions and future work

Test case of data assimilation and downscaling in air quality model showed that for local forecasts even simple downscaling model is much better than mesoscale model alone. Downscaling within data assimilation framework exhibited better skill than model output statictics alone. The best results were obtained with predictor containing also the forecast of analysis increment. This could be attributed to the systematic part of air quality model bias. Combination of downscaling and mesoscale model also allowed us to use local observations that are usually unusable in traditional data assimilation due to large representativity errors. The testing period was quite short for a statistical verification but the first results are encouraging.

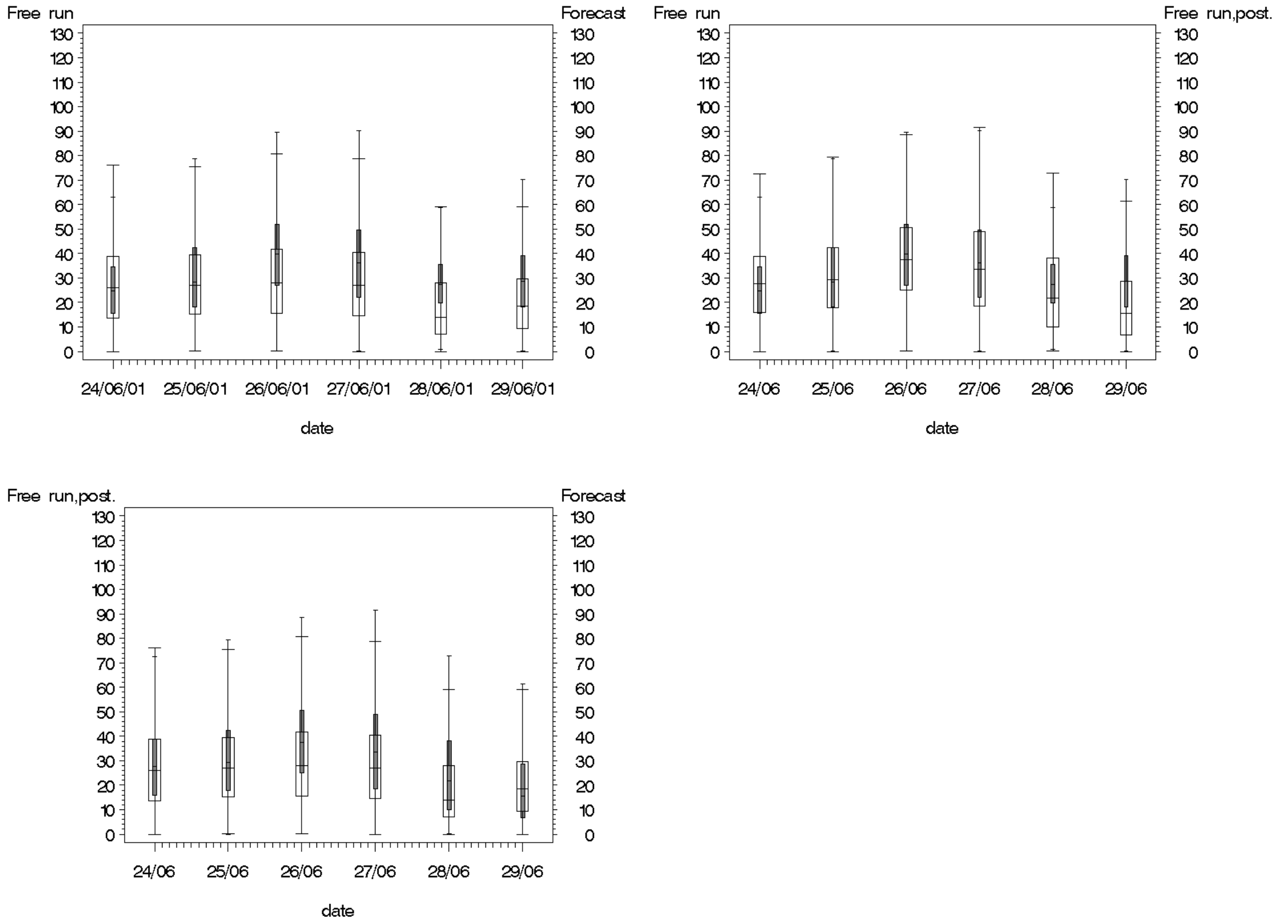

Figure 6: Boxplots of mean absolute errors of forecasts. Left side descriptions belongs to thin full boxes, right side belongs thick empty boxes. Center resp. bottom and top of boxes represent median resp. 1st and 3rd quartile of the error distribution. *Free run* is 1 day ahead forecast of mesoscale model started from analysed state. *Free run,post.* is 1 day ahead forecast of mesoscale model with separate downscaling from data assimilation. *Forecast* is 1 day ahead forecast of mesoscale model with the models of local component and analysis increment.

Our case demonstrated that combination of downscaling model, mesoscale model and data assimilation is beneficial even for very simple downscaling model. There are many possibilities for further development and improvement of presented methods. Downscaling model can be much more sofisticated. Explicit forecasts of the error covariances and confidence intervals which were omitted in our testcase is another topic that should be investigated. Analysis scheme can be improved to analyze simultaneously local and mesoscale component and provide better approximation of the best estimate in sense of the least squares or maximum likelihood. Careful modelling of error covariances is important in this case, since downscaling and mesoscale model can interact and compete for the explanation of residual variance.

There are further possibilities to investigate that are more specific for different cases. For example in case of nowcasting of accidental release (4th case in introduction), the model for downscaling can be a simple model accounting for terrain and landuse forcing of wind field. This model can adapt and refine its parameters with an increasing number of observations. Or the model for downscaling can be deterministic large-eddy simulation model. Data assimilation in traditional sense can be then performed in both mesoscale and large-eddy model.

References

[1] R. G. Hanea, G. J. M. Velders, A. Heemink, "Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model", *J. Geophys. Res.*, vol. 109, D10302, doi:10.1029/2003JD004283, 2004.

[2] A. Segers, A. Heemink, M. Verlaan, M. van Loon, "A modified RRSQRT-filter for assimilating data in atmospheric chemistry models", *Environ. Mod. Software*, vol. 15, pp. 663–671, 2000.

[3] K. Eben, P. Juruš, J. Resler, M. Belda, E. Pelikán, B. C. Krüger, J. Keder, "An Ensemble Kalman filter for short term forecasting of tropospheric ozone concentrations", *Q. J. R. Meteorol. Soc.*, in print

[4] K. Eben, J. Resler, P. Juruš, "Accounting for representativity of ground-level air pollution monitoring stations in data assimilation systems", *Geophysical Research Abstracts* vol. 8, 10015, 2006.

[5] J. S. Whitaker, T. M. Hamill, "Ensemble data assimilation without perturbed observations", *Mon. Weather Rev.*, vol. 130, pp. 1913–1924, 2002.

[6] T. M. Hamill, J. S. Whitaker, "Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches", *Mon. Weather Rev.*, vol. 133, pp. 3132–3147, 2005.

[7] R. Daley, "Atmospheric Data Analysis", Cambridge University Press, New York, 1991.

[8] "The project MEDARD", Inst. of Computer Sci. Czech Acad. of Sciences, Prague. Website www.medard-online.cz

[9] W. P. L. Carter, "Documentation of the SAPRC-99 Chemical Mechanism for VOC Reactivity Assessment", Final Report to California Air Resources Board Contract No. 92-329, and 95-308., 2000.

[10] "AirBase - the European air quality database", European Environment Agency, Website air-climate.eionet.europa.eu/databases/airbase