

ASSIMILATION OF DIAL DATA IN MESOSCALE MODELS: AN IMPACT STUDY DURING IHOP_2002

Volker Wulfmeyer⁽¹⁾, Hans-Stefan Bauer⁽¹⁾, Andreas Behrendt⁽¹⁾
Francois Vandenberghe⁽²⁾
Edward V. Browell⁽³⁾

⁽¹⁾Institute of Physics and Meteorology, University of Hohenheim, D-70593 Stuttgart, Germany,
wulfmeyer@uni-hohenheim.de, hsbauer@uni-hohenheim.de, behrendt@uni-hohenheim.de

⁽²⁾Research Applications Program, National Center for Atmospheric Research,
Boulder, CO 80307, USA, vandenb@ucar.edu

⁽³⁾NASA Langley Research Center, Hampton, VA 23681, VA, USA, edward.v.browell@nasa.gov

ABSTRACT

DIAL methodology is analyzed for developing a forward observation operator and an error covariance matrix for 4-d variational assimilation of water vapor data into mesoscale models. The operator is applied to data of the LASE system which were collected during the IHOP_2002 campaign. The assimilation effort will be performed using the MM5 model nested into an ECMWF analysis. The goal of this project is to investigate the impact of water vapor DIAL data assimilation on the prediction of water vapor, cloud and precipitation fields.

1. INTRODUCTION

Initiation of convection and precipitation are two atmospheric processes which to date are hardly understood and poorly predicted. It has often been argued that water vapor lidar data are ideal for data assimilation into mesoscale models. A significant impact on the quality of weather forecasts is expected, as a key atmospheric variable is measured with high temporal and spatial resolution as well as high accuracy. Additionally, a detailed error analysis can be performed.

In this connection, several important but still unresolved science questions occur such as:

- Can quantitative precipitation forecast (QPF) be improved by assimilation of water vapor lidar data?
- What is the best data assimilation strategy?
- Where do the observing systems have to be located and how many systems are required?
- Can regions be detected where observations are most critical for a good quality of the model forecast so that targeted observations can be performed?
- Is it better to apply a space borne DIAL system or a ground-based network?

These questions can only be answered if a suitable operator for lidar data assimilation is developed and an adequate model system is available. To the best of our knowledge only one result have been reported yet using real experimental data. In this work, the 3-d variational (3DVAR) analysis system of the ECMWF has been applied for improving track and intensity of hurricanes using LASE water vapor DIAL data [1]. A clear enhancement of track and hurricane intensity prediction has been found.

Corresponding results on the mesoscale are lacking, although due to the high resolution of the data particular on this scale an impact on short-range forecasts can be expected. Also 4DVAR systems have not been applied yet. There is no clear approach available how the data shall be assimilated and how the errors shall be considered.

Subject of this paper is the assimilation of water vapor DIAL data in the high-resolution mesoscale model MM5 [2]. The DIAL methodology is discussed for developing a most suitable forward operator for data assimilation and a corresponding operator is proposed. Assimilation is applied for data, which have been collected during the IHOP_2002 campaign [3]. A case is selected and the meteorological conditions are presented. DIAL data collected by the NASA LASE system, which shall be used for the data assimilation efforts, are shown and analyzed. A first comparison with MM5 results is presented.

2. DATA ASSIMILATION TECHNIQUES

In weather prediction models, the initial conditions of the fields of atmospheric variables have to be determined as accurate as possible. This procedure is called an analysis, which is usually achieved by combining a background field with a large set of observations. This information is accumulated in time

into the model state and propagated to all variables of the model. This concept is called data assimilation.

The atmospheric variables handled by the model are projected onto the measured variables using suitable forward operators. Then the multi-dimensional distance between the projected and the measured fields are reduced or minimized using different techniques [4]. We are focusing on 4DVAR, as this is a continuous and physical assimilation technique, which will take major advantage of the high time and spatial resolution of DIAL data. In 4DVAR, the cost function J between the observations and the model fields is minimized which reads:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \sum_{i=0}^n [\mathbf{y}_i - \mathbf{H}_i(\mathbf{x}_i)]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \mathbf{H}_i(\mathbf{x}_i)] \quad (1)$$

\mathbf{x} and \mathbf{x}_b are the state vectors of the model and background field variables, respectively. \mathbf{B} is the background error covariance matrix. \mathbf{y}_i are the observations and \mathbf{x}_i the model forecasts both valid at times i . \mathbf{H}_i is the corresponding model forward operator and \mathbf{R}_i is the observation error covariance matrix.

From Eq. (1), requirements for a DIAL operator can be derived. A forward operator must be constructed which generally includes interpolation from the model discretization to the observation points as well as conversions from the model variable to the observed parameters. Furthermore, an observation error covariance matrix must be provided which includes all types of known errors at each observation point.

3. DIAL FORWARD OPERATOR

3.1 Data processing

In order to construct a suitable DIAL operator, we investigate the DIAL methodology in the case of space borne operation. Modifications of the data processing for ground-based or airborne systems are straightforward. Figure 1 shows the data analysis process proposed for the WALES mission [5].

Lidar backscatter signals are collected at three different online and one offline wavelengths [6]. The SNR of the data is estimated using analytical error propagation. Horizontal averaging is performed until the $\text{SNR} > 5$ in each range bin for minimizing biases in the retrieval [7]. After taking the logarithms of the ratios, the data are vertically averaged and differentiated.

Using external temperature and pressure information such as from NWP forecasts, profiles of the water

vapor absorption cross sections are calculated. A first estimate of a water vapor profiles using the DIAL approximation can be determined. Using propagation of noise errors, a composite water vapor number density profile is calculated using weighted averages.

Rayleigh-Doppler correction of the data is applied and, after iteration, a composite number density profile is derived including a noise error profile. This data processing requires another external data set, the lidar ratio profile S . Details of the inversion processes are discussed in [7].

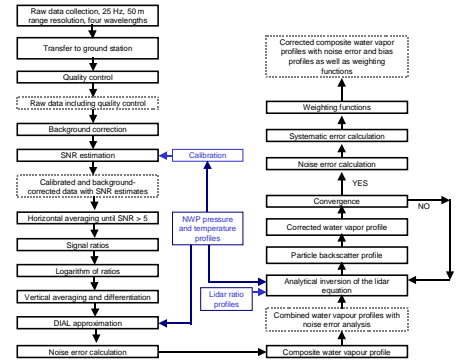


Figure 1. Processing of DIAL data for the WALES mission. The gray dotted boxes indicate certain levels of the data at which they could be delivered to a data assimilation center. The boxes with thin gray lines indicate external information required for the water vapor retrievals. If self-broadening of water vapor must be considered, the loop must be started at DIAL approximation.

The maximum absolute bias in single water vapor retrievals can also be estimated using error propagation. Special attention needs to be devoted to gradients in the particle backscatter profile. Finally, a water vapor profile is provided which includes a noise error as well as a systematic error profile.

3.2 Forward operator and error covariance matrix

Water number density N_w is the measured variable of a DIAL system. If water vapor mixing ratio m is used in the models, the forward observation operator \mathbf{H} reads

$$N_w = \frac{N_L}{M_w R_L} \frac{p}{T} \frac{m}{1 + 1.608 m} \quad (2)$$

where N_L is Loschmidt's number, M_w is the molecular weight of water vapor and R_L is the gas constant of dry air. p is pressure and T is temperature. The observation operator defined by (2) is non-linear. Since m is most of the time much smaller than 1kg/kg, Eq. (2) is

practically linear in m . Data assimilation studies with atmospheric refractivity measurements from GPS have also shown that the effects of the non-linear term p/T^*m are in general weak for typical analysis increments.

In the error covariance matrix \mathbf{R} , three kinds of errors have to be considered: the bias ϵ_b , the instrumental noise error ϵ_n and the representativeness error ϵ_r . Bias errors are handled in the data assimilation system by subtracting it from the original data. Therefore only instrumental noise and representativeness errors show up in \mathbf{R} . In the retrieval process, system noise gets correlated by the same weighting function W of the water vapor data. The noise error covariance matrix corresponds to the variance profile multiplied with the autocorrelation matrix of W . This approach was also applied within a WALES impact study by means of linear optimal estimation theory [8].

Representativeness errors are very difficult to estimate. As DIAL data provide a cross section through model boxes, the representativeness error will be smaller than that of radiosondes. For optimizing the data assimilation effort, we suggest two measures. Firstly, the horizontal model grid should be adapted to the horizontal resolution of the data. Secondly, the representativeness error can be included in \mathbf{R} by simply multiplying the noise error matrix by a certain enhancement factor α . Hence,

$$R_{mk} \cong \alpha \epsilon_{n,m}^2 C_{mk} \quad (3)$$

where $\epsilon_{n,m}$ is the noise variance at model level m . C_{mk} is the autocorrelation function of W between model levels m and k , hence $C_{mk} = C(z_m - z_k)$ where z_m and z_k are the heights of the model levels.

4. CASE STUDY DURING IHOP_2002: MAY 24, 2002

Several cases during IHOP_2002 have been analyzed with respect to a potential impact of DIAL data. On the one hand, DIAL data with good quality, coverage and long duration must be available. On the other hand, interesting atmospheric processes must occur such as initiation of convection and precipitation. The DIAL measurements must be performed in a region, which is sensitive to a better forecast of these processes.

The weather conditions and the measurements performed on May 24, 2002 were most suitable. During this day, three major features dominated the large-scale situation over the mid-west region. In the upper troposphere a well-defined short-wave trough

moved from west to east over the Great Plains during the day. It was associated with a cold front that moved southward across Oklahoma during the evening and night. At the same time a strong southerly low-level jet transported hot and moist air from the Gulf of Mexico. Together this led to strongly increased gradients of geopotential height, temperature and deep-layer shear. The associated large-scale ascent and the already strong mid-level vertical temperature gradients of 8-9 K/km resulted in high values of CAPE (convective available potential energy). Moreover, a significant dryline moved eastward from New Mexico into Texas during the day. Together these features lead to a classical situation for the development of severe thunderstorms in the central and southern Great Plains.

5. LASE DATA ANALYSIS

LASE is an extensively characterized airborne DIAL system which provides water vapor profiles in the entire troposphere using two online and one offline wavelength [9].

A similar data processing scheme as presented in Fig. 1 was used, except that the inversion of the lidar equation for Doppler correction was performed by calibrating the offline lidar signal in a region with negligible particle backscatter. The noise errors in the data were about 8 % over the entire vertical range, which has been estimated using auto covariance analysis of high-resolution data. The maximum bias is of the order of 3-5%, which will be used for sensitivity studies.

During IHOP, data were collected with a horizontal resolution of 14 km and a vertical box average of 330 m. On May 24, 2002, LASE performed measurements between 5:30-10:30 pm UTC. For data assimilation, a sequence between 5:30-8:00 pm will be used whereas the other sequence will be applied for validation. LASE performed several west east transects so that it probed the complex water vapor distribution on both sides of the dryline. The coverage was only partly limited by clouds. This gives rise to optimism that the data have impact on model forecasts.

6. ASSIMILATION OF LASE DATA IN MM5

In MM5, coding of \mathbf{H} and \mathbf{R} according to Eqs. (2) and (3), respectively, has already been performed. The triangle weighting function W of a vertical box average has been calculated. In the first assimilation attempt, an average noise variance profile will be used so that the covariance matrix \mathbf{R} will be time independent. For including the representativeness error, the noise error variances can be scaled by the enhancement factor α .

Currently, this factor is set to unity. Figure 2 shows an example of \mathbf{R} .

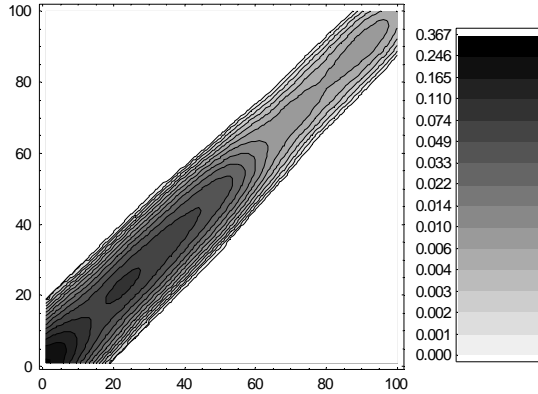


Figure 2. Error covariance matrix \mathbf{R} for 3000 m of LASE data on May 24, 2002, at 21:09:37 UTC. The units are $(\text{g/kg})^2$. The matrix values between 0 and 100 correspond to a height from 1477-4477 m with a resolution of 30 m.

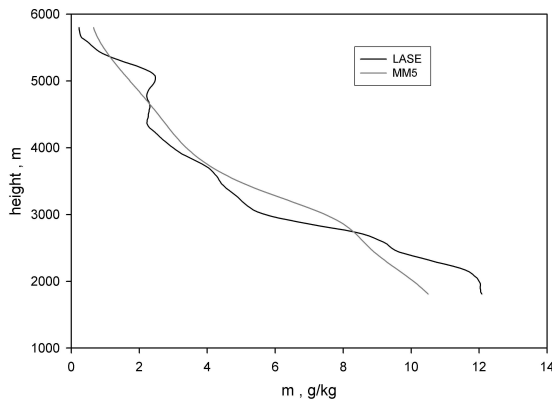


Figure 3. Comparison between LASE and MM5 profile at 17:39 UTC.

MM5 runs have also been performed for investigating the boundary conditions provided by ECMWF. Figure 3 show a comparison of a LASE mixing ratio profile calculated using Eq. (2) with the MM5 profile at the corresponding grid box. In average, the absolute values are similar, however, the profiles differ largely in their fine structures.

7. SUMMARY AND OUTLOOK

LASE DIAL data have been investigated and prepared for data assimilation in MM5. The water vapor number density will be interpolated on the model sigma levels and converted into mixing ratio using the model pressure and temperature valid at that level at the time of the observation. This approach allows to readily use

the MM5-4DVAR software without modifications and shall provide a good preview of the potential value of the LASE data for mesoscale forecasting. Then, the full operator \mathbf{H} and its adjoint will be coded and implemented in the MM5 4DVAR systems. Data assimilation experiments and impact studies will follow. First results will be presented at the conference.

REFERENCES

1. Kamineni, R., et al. Impact of high resolution water vapour cross-sectional data on hurricane forecasting. *Geophys. Res. Lett.*, Vol. 30, 1234-1237, 2003.
2. Zou, X., et al. Introduction to Adjoint Techniques and the MM5 Adjoint modeling system. *NCAR Technical note*, NCAR/TN-435-STR, 110 pages, NCAR, Boulder, 1997.
3. Weckwerth, T. M., et al. An overview of the International H₂O Project (IHOP_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.*, submitted, 2003.
4. Kalnay, E., *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, Cambridge, 2002.
5. European Space Agency, WALES Water Vapour Lidar Experiment in Space, *Report for Assessment*, ESA SP-1257(2), 91p., 2001.
6. Bauer, H., et al. End-to-end simulation of the performance of WALES: Forward module, 22nd International Laser Radar Conference. This issue, 2004.
7. Di Girolamo, P., et al. Simulation of the performance of WALES based on an end-to-end model. 22nd International Laser Radar Conference. This issue, 2004.
8. Gérard, É., et al. Major advances foreseen in humidity profiling from WALES, accepted for publication in *Bull. Amer. Meteor. Soc.*, 2004.
9. Browell, E.V., et al. LASE validation experiment. *Advances in Atmospheric Remote Sensing with lidar*, edited A. Ansmann, R. Neuber, R. Rairoux and U. Wandinger, Springer-Verlag, New York, Berlin, Heidelberg, 289-295, 1996.