Four-Dimensional Variational Data Analysis of Water Vapor Raman Lidar Data and Their Impact on Mesoscale Forecasts

MATTHIAS GRZESCHIK, HANS-STEFAN BAUER, AND VOLKER WULFMEYER
Institute of Physics and Meteorology, University of Hohenheim, Stuttgart, Germany

DIRK ENGELBART
German Meteorological Service, Lindenberg Observatory, Lindenberg, Germany

ULLA WANDINGER, INA MATTIS, DIETRICH ALTHAUSEN, RONNY ENGELMANN, MATTHIAS TESCHE, AND ANDREA RIEDE
Leibniz Institute for Tropospheric Research, Leipzig, Germany

(Manuscript received 12 January 2007, in final form 8 May 2007)

ABSTRACT
The impact of water vapor observations on mesoscale initial fields provided by a triangle of Raman lidar systems covering an area of about 200 km × 200 km is investigated. A test case during the Lindenberg Campaign for Assessment of Humidity and Cloud Profiling Systems and its Impact on High-Resolution Modeling (LAUNCH-2005) was chosen. Evaluation of initial water vapor fields derived from ECMWF analysis revealed that in the model the highly variable vertical structure of water vapor profiles was not recovered and vertical gradients were smoothed out. Using a 3-h data assimilation window and a resolution of 10–30 min, continuous water vapor data from these observations were assimilated in the fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5) by means of a four-dimensional variational data analysis (4DVAR). A strong correction of the vertical structure and the absolute values of the initial water vapor field of the order of 1 g kg⁻¹ was found. This occurred mainly upstream of the lidar systems within an area, which was comparable with the domain covered by the lidar systems. The correction of the water vapor field was validated using independent global positioning system (GPS) sensors. Much better agreement to GPS zenith wet delay was achieved with the initial water vapor field after 4DVAR. The impact region was transported with the mean wind and was still visible after 4 h of free forecast time.

1. Introduction
One outstanding goal of atmospheric research is the improvement of the skill of weather forecast models. Particularly critical is the prediction of precipitation and related extreme events. Only optimized models with well-specified performance can be applied to study the predictability of these meteorological situations. Corresponding research is the subject of several ongoing projects of the World Weather Research Program (WWRP) such as The Observing System Research and Predictability Experiment (THORPEX; see online at www.wmo.int/thorpex/), the Forecast Demonstration Project (FDP) Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine region (D-PHASE; see online at www.map.meteoswiss.ch/map-doc/dphase/dphase_info.htm), and the Research and Development Project (RDP) Convective and Orographically-induced Precipitation Study (COPS; see online at www.uni-hohenheim.de/spp-iop). Whereas THORPEX is focusing on “accelerating improvements in the accuracy of one day to two week high-impact weather forecasts,” D-PHASE is concentrating on the “entire forecasting chain ranging from limited-area ensemble forecasting, high-resolution atmospheric modelling (km-scale), hydrological modelling, and nowcasting to decision making by the end users,” (e.g., for improved flash flood forecasting; Arnaud et al. 2002). The overarching research goal of
COPS is to “advance the quality of forecasts of orographically-induced convective precipitation by 4D observations and modeling of its life cycle.”

Though the relevant scales in these projects are different, all activities have the following in common—improvements of forecasts can only be expected if progress is made on three areas simultaneously: handling of mathematical problems such as model coordinates in complex terrain, optimization of initial conditions, and improvement of model physics. Finally, studies of the predictability of precipitation will be possible by means of ensemble forecasting.

In this study, we are focusing on the determination of initial fields for mesoscale atmospheric modeling. Here, mathematical problems are indeed still an issue, as pointed out (e.g., in Rosatti et al. 2005; Steppeler et al. 2006). Furthermore, mesoscale forecasts are highly sensitive to the quality of model physics including land surface exchange (Cheng and Cotton 2004; Trier et al. 2004; Holt et al. 2006), boundary layer properties (Bright and Mullen 2002; Berg and Zhong 2005), as well as cloud and precipitation microphysics (Doms et al. 2002; Zängl 2004a,b). Regardless of these challenges, a major sensitivity remains to the accuracy of initial fields and boundaries by global forecasts as well as on their mesoscale variability (Richard et al. 2003; Faccani and Ferretti 2005). In various publications it was demonstrated that the representation of mesoscale dynamics and water vapor are particularly crucial (e.g., Crook 1996; Ducrocq et al. 2000).

Therefore, it is critical to study the impact of new water vapor observations on the quality of mesoscale forecasts. Taking into account the large spatial/temporal variability of water vapor, observation systems are required, which provide 2D or even 3D distributions of water vapor. This is only possible with advanced remote sensing systems such as passive IR or microwave sensors or active sensors such as the global positioning system (GPS) tomography (Flores et al. 2000; MacDonald et al. 2002) or ground-based lidar (Goldsmith et al. 1998; Wulfmeyer et al. 2003).

Currently, corresponding datasets are only available during dedicated field campaigns and/or specific case studies. Particularly, the International Lindenberg Campaign for Assessment of Humidity and Cloud Profiling Systems and its Impact on High-Resolution Modeling (LAUNCH-2005) performed by the German Meteorological Service in 2005 (Engelbart and Haas 2007) was a unique opportunity to investigate the role of advanced water vapor observations on mesoscale weather forecasting. During LAUNCH, what is to our knowledge the first time, a network of 13 water vapor Raman lidar systems was operated in an area covering central Europe so that their impact over a larger domain could be investigated by means of observing system experiments (OSEs). As Raman lidar systems have especially high vertical resolution and accuracy, they are expected to be well-suited as basis for future ground-based networks, which is one of the subjects of the European Research Action COST-720, “Integrated Ground-based Remote-Sensing Stations for Atmospheric Profiling.” Therefore, we are focusing on the impact of Raman lidar data assimilation.

Impact studies of new observation systems have to be separated in different steps. First of all, an extensive analysis of the performance of the Raman lidar systems is essential. As the corresponding inversion of the lidar radiation transfer equation for deriving water vapor is unique, this procedure is straightforward so that water vapor mixing ratio profiles can be retrieved with detailed error analyses with respect to noise and systematic errors.

Then, these profiles can be compared with initial mesoscale water vapor fields without any data assimilation efforts. This evaluation gives interesting insight into the accuracy of initial water vapor fields provided by standard mesoscale and global analyses.

Finally, these profiles can be assimilated using different techniques, which may be compared to each other. This comparison is not the subject of this work, as we are focusing on the assimilation technique, which is expected to provide the most accurate initial water vapor fields combining the analysis and the water vapor Raman lidar data: the four-dimensional variational data analysis (4DVAR). This is due to the fact that 4DVAR takes into account the physically consistent time evolution of the model system, the temporal and spatial resolution of the observations, and the error characteristics of the model and the observation system.

For the 4DVAR system, which has been developed for the fifth-generation Pennslyvania State University–National Center for Atmospheric Research Mesoscale Model (MM5), forward operators and their physical adjoints are already available for water vapor lidar observations. These have been developed within our first data assimilation study using airborne water vapor differential absorption lidar (DIAL) data (Wulfmeyer et al. 2006) where we found a strong impact on the initial water vapor field, convection initiation, and even precipitation fields. In this paper we are extending this work to the assimilation of a ground-based network of Raman lidar systems.

It has often been argued that the impact of the assimilation of high-resolution observations on the mesoscale is of limited value, as the skill of forecast is con-
strained by the strong nonlinearity of moist processes. Furthermore, 4DVAR studies are dependent on the choice of the model background error matrix (Hamill and Snyder 2002) as well as on how many, and in what way, variables are included in the cost function. Nevertheless, there is even a lack of information for weather forecasters about the variability of water vapor on the mesoscale and the impact of new observation systems. A clear positive impact of additional, high-quality water vapor observations alone can be expected, which are of equal importance as additional wind information, as pointed out, for example, in Fillion and Mahfouf (2000) and Errico et al. (2004). As we expect that better representations of the water vapor field will also improve the prediction of the location and timing of convection initiation, the impact should also remain after nonlinear moist processes take place. These studies will be the subject of future investigations taking advantage of recent progress in data assimilation and ensemble forecasting (Lorenc 2003; Caya et al. 2005).

This paper is organized as follows. The science goals and the set up of the LAUNCH campaign and an overview of the location and properties of the sensors systems during the intensive observation periods (IOPs) are presented in section 2. The meteorological conditions during IOP-7 selected for this study are discussed in section 3. In section 4, we demonstrate how the Raman lidar data were processed and prepared for 4DVAR. We also discuss how the data assimilation process was performed. Section 5 presents a comparison of initial fields and Raman lidar data. The impact of assimilated Raman lidar data on the initial fields is demonstrated and discussed. The results are summarized in section 6.

2. The LAUNCH-2005 campaign

a. Scientific objectives

Within the framework of the European Research Action COST-720, “Integrated Ground-based Remote-Sensing Stations for Atmospheric Profiling,” of the European Science Foundation (ESF; see also http://www.cost.esf.org) and in connection with the World Meteorological Organization (WMO) Global Energy and Water Cycle Experiment (GEWEX) Working Group on Cloud and Aerosol Profiling (GEWEX CAP), the Meteorological Richard-Aßmann-Observatory in Lindenberg (MOL), Germany, which is operated by the German Meteorological Service (DWD), organized the international campaign LAUNCH-2005 (Engelbart and Haas 2007) in late summer/early fall 2005. The campaign LAUNCH-2005 was designed to accomplish four major scientific objectives:

1) the assessment of new or improved profiling systems like water vapor lidars, cloud radar systems, various microwave profiler systems, a Doppler wind lidar, a new single-photon-counting high-range ceilometer, and the redesigned Fourier Transform Infrared Spectrophotometer (FTIR) spectrometer Emission-Infrared Spectrometer for Atmospheric Research (EISAR);
2) the assessment of various algorithms, combining different techniques for profiling of cloud parameters;
3) the provision of a dataset designed for validation and comparisons between measurements and NWP output; and
4) the provision of a dataset for data assimilation experiments using high-resolution water vapor profiling systems in regional NWP modeling.

This paper is related to objectives 1, 3, and 4.

b. The network for data assimilation experiments

It was intended to concentrate on approaching precipitation systems and to investigate their correct representation in numerical models. This required a large coverage of the observations. For the realization of a first OSE using high-resolution water vapor profiling systems, a network of water vapor Raman lidar systems was installed at a number of locations in Germany, the United Kingdom, and Italy. This network consisted of 13 Raman lidar systems plus two water vapor DIAL systems in order to get insight into the usefulness of lidar data assimilation for operational NWP forecasts.

The part of the LAUNCH network, which is important for this study, is depicted in Fig. 1. Three Raman lidar systems were located in Lindenberg (52.21°N, 14.12°E), Ziegendorf (53.31°N, 11.84°E), and Leipzig (51.35°N, 12.43°E), Germany. The side length of this triangle amounted to 150 km (Leipzig–Lindenberg), 200 km (Lindenberg–Ziegendorf), and 220 km (Leipzig–Ziegendorf), respectively. Major questions to be investigated here were as follows:

- Is there a detectable impact from data of water vapor lidar systems on NWP forecasts?
- How long does this potential impact survive after the end of the data assimilation window?

c. Data conditioning for OSEs

During LAUNCH-2005, data were collected during seven IOPs with a duration of 2–3 days, respectively. The Raman lidar systems measured mainly during nighttime, which resulted in 10 nights of data collected in Germany.

Prior to the supply of all network lidar data to the
assimilation centers at the universities of L’Aquila, Italy, and Hohenheim, Germany, all water vapor profiles were prepared by using the same signal processing software and by securing data quality using collocated radiosondes at each of the different lidar sites. This procedure ensured the high-precision quality of water vapor measurements as well as its comparability, which are mandatory for assessing the different scientific questions listed in section 2b.

3. The selected case: IOP-7

To investigate the above questions, a definition of special synoptic situations was made—all network systems will be alarmed for initiating measurements. For this purpose, daily weather briefings were realized at MOL in Germany. During these briefings, a decision was made on the activities in the upcoming 24–48 h. This decision was generally coordinated between the two major network hubs in Lindenberg and L’Aquila. For the German network, special emphasis was laid on the area of northeastern Germany, as this region was covered by the three lidar systems (Ziegendorf, Leipzig, and Lindenberg; see Fig. 1), which are the subject of this study. Furthermore, additional validation systems, such as cloud radars, ceilometers, and microwave profilers were present at these sites.

For this study, IOP-7 (26–27 October 2005) was chosen because all high-priority systems of the network in eastern Germany were in operation and weather conditions ahead of a sharp frontal rainband approaching from the west were favorable for the lidar systems. Figure 2 shows the surface analysis of the German Weather Service that provides an overview of the synoptic situation for 0000 UTC 27 October 2005. Two low pressure systems were located over the eastern North Atlantic and northwestern Russia. In between the two systems, a northwesterly flow occurred along a boundary that connected the warm front of the western low with the cold front of the eastern low. The surface front separated warm and moist air to the west and southwest from drier and cooler air over northern and eastern Germany. During the IOP, the front moved slowly eastward and diluted during 27 October.

The National Centers for Environmental Prediction (NCEP) analysis of the 500-hPa geopotential height...
(not shown) shows that a westerly to northwesterly flow was present up into the midtroposphere. However, eastern Germany was already located beneath a northward-extending ridge that slowly moved to the east. In the region of the ridge, warm air was transported northward.

The Meteorological Satellite (Meteosat) Second Generation (MSG) channel 6 image (water vapor, 7.6 μm; see Fig. 3) provides information of the water vapor content around 500 hPa. Over southern-central Europe and over the Alps, southerly of the warm front, a region of very dry air was observed in the region where very warm air from the Mediterranean and northern Africa was transported northward in the western flank of the warm sector. Over the locations of the Raman lidar systems, a complex structure along the warm front was
observed with indications of cirrus clouds and increased humidity. Since the surface front was located over eastern Germany, this suggests that the warm and moist air in the warm sector of the low pressure system approaching from the west was already gliding up onto the cooler and drier air to the east of the surface front. The MSG infrared image (not shown) for the same time step showed that large parts of central Europe were cloud free or only covered by low-level clouds. In the broad northwesterly flow along the surface front optically thicker clouds with lower cloud-top temperatures were transported into the region where the lidar systems were located.

In such a situation it is expected that high values of humidity occur on the cold side of the front at the surface with a sharp humidity gradient above the boundary layer. With the approach of the warm air and its up-gliding on the colder near-surface air, the moist region is expected to rise and thicken with time.

4. Raman lidar data assimilation

a. Lidar system overview

Since the first demonstration of Raman lidar measurements, this technique became an operational active water vapor remote sensing technique. This is due to the fact that huge technological advances have been made with respect to high-power laser transmitter development at suitable wavelengths, optical receiver technology, and detector technology. Details of the setup and water vapor retrievals of state-of-the-art Raman lidar systems can be found in Whiteman (2003) and Wandinger (2005). The Cloud and Radiation Test bed (CART) Raman lidar (CARL) demonstrated since the 1990s that routine measurements of water vapor profiles with the Raman lidar technique are possible during daytime and nighttime (Turner et al. 2002).

1) RAMSES IN LINDENBERG

Since the summer of 2005, the DWD has run the Raman lidar for Atmospheric Moisture Sensing (RAMSES) at MOL in Lindenberg. RAMSES supplies quasi-operational, high-precision water vapor profiles both for climate monitoring of the troposphere [e.g., in the frame of the Global Water Vapor Project (GVaP; Randel et al. 1999)] and as a reference for the assessment and validation of new ground-based and/or spaceborne sensors [e.g., for the European Climate Monitoring Satellite Application Facility (CM-SAF; Deutscher Wetterdienst 2000)]. Because these main objectives are long-term activities and require continuous operation of the lidar, RAMSES was specifically designed for unattended and dependable operation. During its test phase, RAMSES was restricted to nighttime measurements. Following this test phase, RAMSES will be upgraded later to water vapor measurements during daytime conditions. RAMSES is the second fully operational worldwide lidar system after CARL and the first one with a fully autonomous data retrieval scheme (Mattis and Jaenisch 2006).

RAMSES is based on an injection-seeded frequency-tripled Nd:YAG laser with a total pulse energy of up to 1.6 J. Only third-harmonic radiation at 354.7 nm is emitted into the atmosphere. The typical pulse energy at this wavelength chosen for operational conditions is 300 mJ. The pulse repetition rate is 30 Hz. The laser beam is expanded tenfold and directed onto the axis of the far-field telescope with three beam-folding mirrors.

The receiving optics is presently optimized for nighttime water vapor measurements throughout the troposphere. RAMSES is operated with two receiver telescopes with a diameter of 0.2 m for the near-range observations and a diameter of 0.8 m for far-range observations simultaneously. Two nearly identical receiver boxes for the far-field and the near-field channels are deployed. After the beam collimation, dichroic beam splitters and interference filters separate the elastically backscattered light at 355 nm and the vibrational–rotational Raman signals of water vapor at 408 nm and of nitrogen at 387 nm. The ratio of the corresponding backscatter signals at these wavelengths is already proportional to the water vapor mixing ratio. RAMSES is supplied with a combined analog and photon-counting data acquisition system. In the framework of this study, only signals based on the photon-counting technique were evaluated.

2) IFTR Raman Lidar in Leipzig

RAMSES benefited strongly from the in-depth and long-term experience of Raman lidar developments at the Leibniz Institute for Tropospheric Research (IfT; Mattis et al. 2002; Wandinger 2005). In Leipzig, the stationary three-wavelength Raman lidar of the IfT contributed to the LAUNCH-2005 campaign. This powerful instrument applies a Nd:YAG laser and emits radiation at 355, 532, and 1064 nm with an overall pulse energy of 1.6 J. The backscattered light is collected with a 1-m telescope. Ten independent return signals are evaluated. Water vapor measurements are performed on the basis of the vibration–rotation Raman signals of nitrogen and water vapor at 387 and 408 nm, respectively, for the primary wavelength of 355 nm. The signals are acquired in the photon-counting mode. Details of the system are given by Mattis et al. (2002).
3) IFT RAMAN LIDAR IN ZIEGENDORF

The IfT multiwavelength Raman lidar was set up, together with the IfT Doppler wind lidar, next to the DWD wind profiler in Ziegendorf. The system has been described in detail by Althausen et al. (2000). Two Nd:YAG lasers and two Ti:Sapphire lasers generate radiation at 355, 400, 532, 710, 800, and 1064 nm. At the receiver end, the light collected with a 0.56-m telescope is divided into 15 detection channels. The water vapor mixing ratio is determined from the vibration–rotation Raman signals of water vapor and nitrogen at 660 and 607 nm, respectively, which are generated by laser radiation of 532-nm wavelength. The photon-counting technique is applied in case of the Raman signals.

4) OPERATIONAL DATA ANALYSIS

A common data evaluation algorithm developed at IfT (Mattis and Jaenisch 2006) was used to analyze all data measured in Lindenberg, Leipzig, and Ziegendorf during LAUNCH-2005. The water vapor to dry air mixing ratio is calculated from the ratio of the water vapor Raman signal to the nitrogen Raman signal (Wandinger 2005). The calibration factor needed to convert the relative to an absolute measure is derived by comparing the lidar profile for a selected time and height range with a local radiosonde measurement performed parallel to the lidar observation. At Lindenberg, the on-site operational soundings at 0000, 0600, 1200, and 1800 UTC were used. Both IfT lidar systems were equipped with radiosonde stations, so that individual profiles for calibration were also available in Leipzig and Ziegendorf. Consequently, the overall accuracy of Raman lidar water vapor measurements was limited by the accuracy of the radiosondes used for calibration. Extended comparisons demonstrated that the systematic calibration error is of the order of 5% (Wandinger 2005).

The statistical measurement error results from the signal noise in the Raman signals and depends on height as well as on spatial and temporal averaging. These noise errors were derived for each profile using the analytical error propagation of photon statistics in the water vapor Raman lidar equation. For details, see Whitman (2003) and Wandinger (2005). Typically, the errors are <5% in the lower troposphere and of the order of 10%–20% in the upper troposphere for averaging periods of 15–30 min.

5) SYSTEM PERFORMANCE

Figure 4 shows a typical comparison of Raman lidar and radiosonde water vapor profiles performed on 30 October 2005 using resolutions of 10 min and 67.5–307.05 m, respectively. The agreement is excellent and no indication of systematic errors larger than 0.1 g kg⁻¹ between 500 and 7500 m was found.

These data also demonstrate the large vertical variability of water vapor, which was observed during all IOPs. Several dry layers were found (e.g., at a height of about 3500 m), which corresponded to about 10% relative humidity, whereas at a height of 6000 m the relative humidity increased to 60%. These structures changed continuously confirming the importance of high-resolution observations for model evaluation and water vapor data assimilation.

In Fig. 4, the thin red lines around the lidar profiles correspond to the noise error estimates. The error profiles agree with the estimates made above. The errors are about 5% or 0.1 g kg⁻¹ (whichever is lower over the entire range), which can be transferred to other vertical and temporal resolutions using noise error propagation. Based on these results, the following conclusions for the data assimilation efforts could be drawn: no bias correction of the Raman lidar data was considered necessary in this study, as the systematic errors were considerably less than the background errors of the model. The noise error analyses were stable and reasonable and could be ingested in the observation error covariance matrix.

6) TEMPORAL EVOLUTION OF WATER VAPOR PROFILES

All Raman lidars were able to provide high-quality water vapor profiles up to the midtroposphere. RAMSES demonstrated the best signal-to-noise-ratio (SNR) with a time resolution of 10 min due to high average power of the laser transmitter and high efficiency of the receiver. However, the RAMSES measurements were limited to heights >500 m. To obtain similar noise errors, the time resolution of the IfT and Ziegendorf Raman lidar water vapor profiles were reduced to 30 min, but they provided water vapor observations down to the ground.

Figure 5 presents the temporal development of the moisture profiles from sunset to sunrise on 26–27 October 2005 as observed with RAMSES in Lindenberg. Similar vertical resolutions were used as in Fig. 4. Noise error profiles are also shown and demonstrate that the noise errors was typically <0.5 g kg⁻¹ over the entire range, which sets strong constraints for data assimilation.

The extraordinary vertical resolution of the Raman lidar water vapor measurements allowed the detection of thin dry layers up to 8 km alternating with moist layers carrying relative humidities up to 85%. During the observation period, subsidence of the dry layers was
observed by about 2 km. This was probably due to the effect that the approaching warm front, which glided up to the cold sector of the low over Russia. Furthermore, the approach of the warm front appeared as an overall increase of humidity by about 1–2 g kg$^{-1}$ in the lower troposphere starting at about 0100 UTC, as expected by the weather analysis in section 3.

b. The MM5 4DVAR system

For the assimilation experiments the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model was used. It is a limited-area, nonhydrostatic, terrain-following sigma-coordinate model designed to simulate or predict mesoscale and regional-scale atmospheric circulation. The MM5 is the latest in a series developed from a mesoscale model used by Anthes at Penn State in the early 1970s that was later documented by Anthes and Warner (1978). Since that time, it has undergone many changes designed to broaden its usage. These include the following:

- a multiple-nesting capability (one-way, two-way, moving nests);
- nonhydrostatic dynamics, which allows the model to be used at kilometer scales;
- multitasking capability on shared- and distributed-memory machines;
- different methods for data assimilation [4D data assimilation (FDDA), 3DVAR, and 4DVAR]; and
- various physics options of different complexity.

A detailed overview of the MM5 is given in Grell et al. (1995).

Since MM5 is a regional model, it requires an initial condition as well as lateral boundary conditions to run. To produce lateral boundary condition for a model run, one needs gridded data to cover the entire time period that the model is integrated. For our experiments the European Centre for Medium-Range Weather Forecasts (ECMWF) operational analysis provides the necessary data.

The MM5 model has been chosen for data assimilation because it provides convenient tools for ingesting measurements of different observation systems, which is a good starting point for lidar data assimilation. The different data assimilation schemes are described in Barker et al. (2004) 3DVAR, Ruggiero et al. (2001) 4DVAR, and Stauffer and Seaman (1994) FDDA. With the MM5 3DVAR system, only several sequential analyses are possible. All observations are shifted to the next analysis time, or observations are only accepted at analysis time. Because of the spinup of the
MM5 system, we think the minimum step size between 3DVAR analysis is about 3 h. The major disadvantage of FDDA is that the temporal and spatial variations of observation errors are not considered. Especially for lidar systems, this is very important (e.g., in the presence of clouds). Therefore, neither 3DVAR nor FDDA were alternatives for us. We selected 4DVAR, as we expected the major impact of a data assimilation effort. 4DVAR takes advantage of the high temporal and spatial resolution of ground-based Raman lidar data. This is particularly important in regions with high spatial–temporal variability in the water vapor field such as near frontal boundaries. 4DVAR considers the error characteristics of each instrument by using the provided error profile to define the observation error covariance matrix $R_i$. Additionally, this continuous data assimilation technique takes into account the physics of the atmospheric processes while minimizing the cost function $J$:

$$J(x) = (x - x_b)^T B^{-1} (x - x_b)$$

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

(1)

where $x$ and $x_b$ are the state vectors of the model and background field variables, respectively; $B$ is the background error covariance matrix; $y_i$ are the observations and $x_i$ the model forecasts both valid at times $i$; $H_i$ is the corresponding model forward operator; and $R_i$ is the observation error covariance matrix.

4DVAR finds the initial condition $x$ where the cost function has a minimum. The minimization is done via an iterative scheme using the gradient of the cost function. For the propagation of the cost function gradient at observation time to the analysis time, the adjoint model is necessary. The adjoint model of the MM5 was derived from a linearized version of the MM5 (Nehrkorn et al. 2001). The iterative process of finding the minimum of the cost function has to be started from a first guess, in our case the ECMWF analysis.

In its current release, the MM5 4DVAR offers diagonal background error covariance matrices $B$ only. This approximation, however, has proven to work well for most studies conducted with the system (Zou et al. 1995; Xiao et al. 2000). This can be explained by the ability of 4DVAR to self-generate physically consistent structure functions as the model integration continues in time. Therefore, the specification of $B$ at the initial time does not seem crucial. For each control variable, the background error variances (diagonal elements of the matrix) were specified by constructing the differences between the 3-h forecasted and the initial values at each grid point. At each vertical level, the maximum value of the differences is found and assigned to all grid points on that level. This creates a vertical profile of forecast errors valid at all geographical locations of the model. The forecast errors are then squared to produce

---

**Fig. 5.** Observations of RAMSES between 1900 UTC 26 Oct 2005 and 0400 UTC 27 Oct 2005. (top) The observed water vapor mixing ratio in a time resolution of 10 min. (bottom) The estimated absolute error of the observations.
the diagonal elements of the background error covariance matrix.

The system was recently considerably improved for the assimilation of water vapor DIAL data (Wulfmeyer et al. 2006). Airborne DIAL measurements were assimilated on a convective active day during the International H2O Project (IHOP_2002) field campaign. The assimilation study shows that the inclusion of the water vapor profiles in a 3-h assimilation window clearly improved the prediction of precipitation. Therefore, we were optimistic to observe similar effects by the extension of the MM5 4DVAR system to water vapor Raman lidar data assimilation.

For this purpose, only minor changes had to be added to the work of Wulfmeyer et al. (2006):

- The model background errors had to be derived (see above).
- The observation operator $H$ was even simpler, as the Raman lidar delivers directly the observable of the cost function (viz., the water vapor mixing ratio), whereas in the DIAL data assimilation absolute humidity had to be transformed to a mixing ratio.
- The error characteristics of the Raman lidar systems had to be derived [see sections 4a(5) and 4a(6)].
- The vertical and temporal resolutions of the time steps were chosen so that acceptable error profiles were achieved. Table 1 summarizes the corresponding values for each lidar system.
- Mean values of the measured mixing ratio for each corresponding model box were calculated for each time step. This interpolation procedure was explained in Wulfmeyer et al. (2006).
- For the diagonal elements of the observation error covariance matrix, the provided errors were used, as long the absolute error was larger than 0.1 g kg$^{-1}$. Otherwise, an overall error of 0.1 g kg$^{-1}$ was selected. We assumed that this approach also included the representativeness error. The off-diagonal elements of the observation covariance matrix were set to zero.

After all these efforts, we set up the required infrastructure for present and future data assimilation studies using ground-based networks and airborne or even space-borne water vapor lidar systems.

c. Model configuration

The assimilation experiments were split into two major steps. First, the assimilation is done using coarse horizontal resolution and simplified physics. The restriction to a simplified physical package is necessary because adjoint versions of the parameterizations are required, which are to date only available for the simplest ones. For this reason, the state vectors in the cost function include the atmospheric wind, temperature, and water vapor fields but no surface variables. For our assimilation run, a horizontal resolution of 27 km was used in a model domain with $82 \times 74$ grid points and 36 vertical levels with layer thicknesses ranging from 70 m near surface to about 1 km at the top of the model.

Result of the assimilation is an optimal initial condition at the beginning of the assimilation window in the course model domain. We consider this approach valid, as we expect that the major uncertainties in the initial fields are, to a large extent, due to initial conditions and less by model physics. In our study, we chose a data assimilation window from 2300 UTC 26 October to 0200 UTC 27 October 2005.

Afterward, we performed two forecasts. One was done with the modified initial state from the assimilation and the other one with the original initial state. For these forecasts most sophisticated parameterizations available and 4 two-way interactive nested domains were used. All domains have the same vertical structure as in the assimilation. The location of the domains is shown in Fig. 1 and their size and resolutions are summarized in Table 2.

### Table 1. Vertical and temporal resolutions of the lidar systems used for the assimilation.

<table>
<thead>
<tr>
<th>Lidar system</th>
<th>Vertical resolution (m)</th>
<th>Temporal resolution (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAMSES (Lindenberg)</td>
<td>60–300</td>
<td>10</td>
</tr>
<tr>
<td>IfT Leipzig</td>
<td>60–300</td>
<td>30</td>
</tr>
<tr>
<td>IfT Ziegendorf</td>
<td>60–300</td>
<td>30</td>
</tr>
</tbody>
</table>

### Table 2. MM5 domains used in this study.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>27 km</td>
<td>$82 \times 74 \times 36$</td>
</tr>
<tr>
<td>9 km</td>
<td>$82 \times 82 \times 36$</td>
</tr>
<tr>
<td>3 km</td>
<td>$100 \times 100 \times 36$</td>
</tr>
<tr>
<td>1 km</td>
<td>$82 \times 82 \times 36$</td>
</tr>
</tbody>
</table>

5. Results and discussion

a. Evaluation of initial fields

Before data assimilation is performed, a comparison of the time–height cross sections of the lidar data with
the ECMWF control data gives insight into the performance of the model with respect to the representation of the water vapor field.

Figures 6–8 present the model results at the grid boxes of the lidar systems in comparison with the lidar data (see the top and middle panels). The coarse structure of the water vapor field, a humid layer up to about 4 km in Lindenberg and Ziegendorf and 2 km in Leipzig, is present and the model also shows the moistening of the lower troposphere due to the approaching warm front. However, in all cases, the control fields are not able to reproduce fine structures in the vertical variability of the water vapor field and their strong vertical gradients. The dry subsidence layers, which are particularly visible in Lindenberg and Ziegendorf, are missing and the humidity gradients above the humid layer are largely underestimated. These deviations may be due to a coarse vertical resolution of the model.

Indication for biases in the model field are found in the midtroposphere between 2 and 6 km. In all cases, the midtroposphere is slightly too dry. In the atmospheric boundary layer (ABL) the result is unclear. In Lindenberg, no ABL data were available. In Leipzig, the model ABL was too humid whereas it was too dry in Ziegendorf. In any case, the data indicate that both biases in the model field are significant and the vertical distribution of water vapor can strongly deviate from reality.

b. Impact of 4DVAR on the analysis and the results in the data assimilation window

Figures 6–8 also present the comparison of water vapor fields of the control run and the run after data

![Fig. 6. Direct comparison of the assimilated observations with the forecast from the first guess and the improved forecast. (top) Observations with RAMSES, (middle) forecast from the ECMWF initial state used as first guess, (bottom) and forecast from the modified initial state by 4DVAR.](image-url)
assimilation (see the middle and bottom panels). The improvement of the initial fields after data assimilation is substantial. In all cases, the vertical structure and gradients in the water vapor fields are corrected. Furthermore, the above-mentioned biases in the midtroposphere and deviations in the ABL are strongly reduced. These improvements were produced by a decrease of the value of the cost function $J$, which is shown in Fig. 9.
The expected behavior of a successful reduction of the cost function is visible. The observational part is generally decreasing with the iterations, so the changed initial fields now produce a forecast for the assimilation window, which fits better to the observations. Only the cost function part of Ziegendorf develops differently. Already for the first guess, the cost value is small so that it changes only slightly. The background part of the cost function increases, which indicates a change of the forecast with respect to the first guess.

Figure 10 shows differences in the mixing ratio and temperature for 850- and 500-hPa heights for the initial state at the beginning of the assimilation window at 2300 UTC 26 October 2005. This is the time step that is changed by the 4DVAR run. In the region of Lindenberg the amount of moisture was increased at 850 hPa and decreases slightly at 500 hPa, whereas the moisture in Ziegendorf and Leipzig were hardly influenced at the beginning of the data assimilation window.

The small impact in the region of Ziegendorf can be explained by Fig. 9. The cost value at iteration step zero was already small for Ziegendorf and during the 4DVAR this value changed very little. Therefore, the first guess from the ECMWF analysis was close to the observation that the 4DVAR had no reason to change the initial field of the model in this region.

However, a strong impact on the water vapor field was produced in a region that was even larger than the Raman lidar triangle. In the lower to the midtroposphere, a redistribution of moisture of up to 1 kg m\(^{-1}\) occurred. This amount can have significant consequences for the timing and location of convection initiation (Crook 1996). The vertical impact reached also the 500-hPa level.

The shape of the impact region indicates that mainly the water vapor content upstream of the Leipzig and Lindenberg Raman lidar system was reduced. This is consistent with the prevailing wind direction from the northwest. Obviously, the water vapor field was adjusted by the modification of the advection of water vapor. The extent of the impact region is promising, as it is shown that a coarse grid of three Raman lidars can produce a nearly homogeneous impact over an area of 200 km × 200 km.

Although only water vapor observations were assimilated, the temperature was also influenced. Figure 10 shows temperature changes of −1 to 0.5 K in 850- and 500-hPa heights. This illustrates the dynamical coupling of the humidity and temperature fields in the assimilation system.

c. Validation of the improvement of the water vapor field by 4DVAR

If observations with high accuracy are present, they are strongly constraining the resulting water vapor field after data assimilation. The 4DVAR system minimizes the cost function by combining changes of the water vapor, wind, and temperature fields. It can be expected that a minimization of the cost function will always be possible if the observations do not deviate too much from the control fields.

However, the question arises whether the 4DVAR results are in better agreement with reality. This must not necessarily be the case considering the limitations of the 4DVAR system, namely, diagonal error covariance matrices, coarse horizontal resolution, and simple physical parameterizations. Furthermore, the impact area and number of observations are limited. Therefore, it is reasonable to validate the 4DVAR run using an independent water vapor dataset with good accuracy and coverage. We are using the German GPS network, which is well suited for this purpose (Gendt et al. 2004).

The GPS stations are determining the zenith wet delay according to the following equation:

\[
ZTD_{\text{wet}} = 3.37 \times 10^{-1} \frac{K^2}{\text{hPa}} \int_{\text{ground}}^{\text{top of model}} \frac{e}{T^2} dh, \quad (2)
\]

which is proportional to the integrated water vapor in the column. In this equation, \(e\) is water vapor pressure, \(T\) is temperature, and \(h\) is height. Consequently, if strong changes in water vapor profiles occur during the assimilation process, this should be reflected in \(ZTD_{\text{wet}}\).

We determined the observed delay \(ZTD_{\text{wet,obs}}\) at GPS stations close to the impact region at Potsdam (POTS), Dessau (DESS), and Staßfurt (STAF), which are also indicated in Fig. 10. Then we calculated their difference with the delays \(ZTD_{\text{wet,mod}}\) produced by the model re-
sults with and without data assimilation. For this purpose, we used Eq. (2) and evaluated the integral via the trapez formula from the levels of the model.

Figure 11 presents the results. It clearly shows a better agreement of the model with assimilation to the observed wet delay during the time when the region of influence of the assimilation covers the location of the GPS measurement. This impact reduces after several hours because the impact region is advected to other locations. Therefore, we demonstrated not only an impact of Raman lidar data assimilation but also the fact that the results agreed much better with reality. Unfortunately, no GPS stations were available for Poland and the Czech Republic, to validate the impact of the advanced forecast. This will be a subject of future activities.

d. Impact on the forecast

The next sequence of plots, Fig. 12, shows the differences at the same levels as in Fig. 10, but 4 h after the end of the assimilation window, corresponding to a 7-h forecast. A clear transport of the region influenced by the assimilation corresponding to the synoptic situation is seen. The region changed by the assimilation is transported to the southeast. The difference increases with height due to increasing wind speeds with height in the strong northwesterly flow behind the cold front of the cyclone. The temperature differences indicate a larger spread of the information in the temperature field, while the amplitude of the disturbance is decreasing. The reason for this different transport behavior is under investigation.

6. Summary and outlook

In this study, for the first time, water vapor Raman lidar data were assimilated in the MM5 4DVAR system. For this purpose, a triangle of Raman lidar systems covering an area of about 200 km × 200 km in north-
eastern Germany was operated during the LAUNCH-2005 campaign. The triangle consisted of the RAMSES lidar of DWD and two Raman lidar systems of the IfT.

The performance analysis of the Raman lidar systems indicated that these systems are capable of delivering water vapor profiles with a vertical range of up to 8 km and resolutions of 60–300 m and 10–30 min, respectively. Intense comparisons with radiosondes demonstrated that the corresponding bias is <5% over the entire range. Consequently, bias correction of these data for data assimilation was not considered necessary. The noise errors were of the order of 5% using the same resolutions so that the vertical structure of the water vapor profiles could be revealed in detail. Currently, the operation of these Raman lidar systems is constrained to nighttime operation. Also the data assimilation effort presented in this work was performed using nighttime data. However, future improvement will enable daytime operation of these lidar systems, too. This potential has already been demonstrated by Turner et al. (2002).

Fig. 11. Differences between observed and modeled GPS total zenith path wet delay for the three GPS stations: POTS, DESS, and STAF. Dotted lines are the differences between observation and the forecast from the ECMWF first guess (CONTROL). Solid lines are the differences to the forecast with data assimilation.

Fig. 12. As in Fig. 10, but at 0600 UTC or 4 h after the end of the assimilation window.
The evaluation of initial water vapor fields derived from ECMWF analysis without any data assimilation efforts revealed that the model was not able to recover the highly variable vertical structure of water vapor profiles and their strong vertical gradients. Furthermore, the model fields were consistently drier in the midtroposphere. In the ABL, the deviations were also significant.

Using the MM5 4DVAR system with a 3-h data assimilation window, continuous water vapor data from these observations were assimilated. In the observation error covariance matrix, the diagonal elements were filled with the noise error variances, which can directly be determined with the Raman lidar signals in real time.

Afterward, a strong improvement of the water vapor field was achieved not only at the locations of the lidar systems but also in a region that had at least the size of the triangle. The correction of the vertical structure and the absolute values of the initial water vapor field were of the order of 1 g kg\(^{-1}\). The improvement occurred mainly upstream of the RAMSES and Leipzig Raman lidar systems. This indicates that mainly within the data assimilation process the advection of the water vapor field to the measurement sites was corrected.

Particularly promising is the fact that we were able to validate these improvements with independent GPS measurements. At three GPS stations, which were in the range of the impact region, we found a considerably better agreement to GPS zenith wet delay of the initial water vapor field after 4DVAR than the uncorrected field. This was visible as long as the impact region was advected over the GPS sites. This region was transported with the mean wind and was still visible after 4 h of free forecast time.

In summary, water vapor Raman lidar systems hold great promise to be used for mesoscale data assimilation efforts. Their high resolution and accuracy set strong constraints to the data assimilation system so that their data cause a high and correct impact on the initial water vapor field. Taking advantage of 4DVAR, a network density of 200 km was sufficient to cause a nearly homogeneous impact region between the Raman lidar locations. As the magnitude of the correction of up to 1 g kg\(^{-1}\) is critical for the location and timing of convection initiation, it can be expected that a network of continuously operated Raman lidar systems would be able to improve the skill of short-range weather forecasts with respect to cloud formation and precipitation.

This first study is meant as a qualitative study to check whether there is an overall larger-scale impact of Raman lidar observation on the 3D water vapor field. Furthermore, we use only data from one case, since only the data from all three Raman lidar systems were available simultaneously.

In the future, this study will be extended to a variety of meteorological conditions and to the comparison of different data assimilation techniques. Furthermore, the fast analysis of Raman lidar data permits real-time data assimilation studies. First attempts will be made during the WWRP RDP COPS in collaboration with the FDP D-PHASE.

Acknowledgments. We thank Florian Zus at our institute for developing the tools for GPS validations. The provision of GPS data by Galina Dick and Jens Wickert from Geoforschungszentrum Potsdam is highly appreciated.

REFERENCES


 Ducrock, V., J.-P. Lafore, J.-L. Redelsperger, and F. Orain, 2000: Initialization of a fine-scale model for convective-system pre-


