PROJECT DESCRIPTION

Estimation of cloud properties in 4D from Cloud resolving Data Assimilation

Principal Investigator: Dr. Tomislava Vukicevic, PAOS, University of Colorado

Collaborator: Dr. Frank Evans, PAOS, University of Colorado

1. Introduction

Accurate estimates of cloud properties are required for improved understanding of feedbacks within the atmospheric system and the system's predictability on wide range of spatial and temporal scales from individual storms to climate. Quantitative observations of clouds and precipitation are typically obtained by indirect, remote sensing methods Although considerable progress has been made in remotely sensing and retrieving the cloud properties, complex 3D cloud structure and its interaction with the atmospheric environment is not well specified from observations alone. Cloud parameterizations in numerical weather prediction (NWP) and general circulation models (GCMs) and more explicitly cloud resolving models (CRMs) are used instead to study the 3D spatial variability of the cloud properties and interaction with the environment. While the NWP and GCM parameterizations and the CRMs produce physically realistic results they cannot simulate actual observed 3D cloud evolutions to aid in explicit analysis of the cloud properties and their interaction with the dynamic environment. This limitation results from solving an open dynamical system of equations for which initial and boundary conditions are required. The initial and boundary conditions are not currently available for the cloud properties and other atmospheric quantities on the cloud resolving scales.

An example of the significance of the initial conditions in the cloud resolving model simulations is shown in the study by Khairoutdinov and Randall (2003) where the model results were compared with the ARM (Atmospheric Radiation Measurement) observations. The study demonstrates that the uncertainties which are associated with a typical formulation of the cloud microphysics produce less sensitivity in bulk results of the CRM simulations such as monthly mean precipitable water or surface precipitation rate than the uncertainties associated with the initial conditions. This result directly implies that the CRM simulations require accurate initialization of the atmosphere on cloud resolving scales in 3D to simulate the observed cloud evolutions for the purpose of testing and improving the cloud parameterizations. The CRM initial conditions cannot be obtained directly from the high quality ARM observations because these are taken at only few locations and the atmosphere is not, as well known, spatially homogenous.

- **To facilitate accurate estimates of the evolving 3D cloudy atmosphere on the cloud resolving scales we propose that satellite remote sensing and ground based observations should be combined with a cloud resolving dynamical model such that they optimally constrain the model solution which is controlled by the initial and boundary conditions and the model parameterizations.**

This approach is known as data assimilation or optimal estimation of modeled state from observations (Tarantola, 1986; Cohn, 1997). The data assimilation methodology has many applications in the atmospheric sciences from the operational weather analysis and prediction (Kalnay, 2003; Parish and Derber, 1998), the ocean state estimation (Bennett, 2003), the atmospheric trace gas and emission analysis (Enting, 2002) to the modeling of carbon cycle (Vukicevic et al, 2001; Braswell et al, 2004). In all of these applications the parameters which control the modeled states, be it initial or boundary conditions or physical parameters, are estimated via a set of objective criteria that include knowledge about the information in the models and observations to render new quantitative representation of what is not directly observed from what is observed. The problem of estimating the atmosphere with clouds using the cloud resolving model is similar to these other applications in terms of underlying methodology. The specific properties of the problem would primarily result from the choice of observations which should be such that they are directly sensitive to the cloud properties in dynamic environment.

Perhaps the best set of observations to start with in exploration of the cloud resolving data assimilation is the satellite imaging and sounding by GOES (Geostationary Operational Environmental Satellites) and the ARM Southern Great Plains (ARM SGP) site measurements. The GOES visible and infrared measurements are well established long term observations of the atmosphere (Greenwald and Christopher, 2000; Menzel et al., 1998). The GOES imager measurements are traditionally used for the cloud identification and analysis of bulk cloud properties because of the spectral characteristics (Rossow and Schiffer, 1999). These measurements also have high horizontal spatial resolution (1- 4 km) and temporal frequency of 15 minutes over large areas including the SGP. The high spatial and temporal frequency of the GOES imager measurements is desirable for the cloud resolving data assimilation because it captures the spatial and temporal variability of the cloudy atmosphere.

The GOES sounder measurements are traditionally used to retrieve the atmospheric temperature and humidity at different vertical levels (Menzel et al., 1998). The GOES sounder radiance measurements have about 12 km horizontal resolution and 1 hour frequency, which is less than the imaging measurements, but they are also continuously available over the same large areas. The satellite sounding measurements are complementary to the imaging because the former provides the information about the atmospheric environment while the latter is mostly sensitive to the clouds when these are present. In standard operational weather analysis the GOES sounder temperature and humidity retrievals are used in the cloud free conditions only. This limitation is not necessary in theory. The sensitivity of the top of the atmosphere radiance to the clouds is well known from the point of view of radiation modeling. What is not well know is the cloud 3D spatial distribution and properties. This prevents the use of retrievals or measured radiances in the cloudy conditions. In the proposed cloud resolving data assimilation the uncertainty of cloud location and properties will be minimized simultaneously with the assimilation of the GOES sounding radiance in the cloudy atmosphere. There are numerous other imaging and sounding satellite measurements available today besides the GOES (Kidder and Vonder Haar, 1995). We propose to first address the use of the GOES measurements because of simultaneously good temporal and spatial coverage.

The ARM SGP measurements are the best local cloud measurements available today with nearly continuous temporal coverage of diverse and highly accurate measurements (Stokes and Schwartz, 1994). The ARM observations have limited spatial representation and cannot by themselves define the 3D cloudy atmosphere (Barnett et al., 1998; Long et al., 2002). In the proposed study they would provide: a) rich source of accurate data for verification of the GOES data assimilation and b) potentially strong additional constraint locally when assimilated.

We propose to explicitly address the following questions:

- **To what degree of accuracy can the 4D cloud properties and the associated atmospheric environment on the cloud resolving scales be retrieved from the satellite imaging and sounding measurements alone in the assimilation with a cloud resolving model?**
- **What is the value added by the high quality but spatially sparse ground based observations in the assimilation?**

The accuracy will be measured in three ways: a) statistical error analysis in the modeled space, b) statistical error analysis in the observation space with respect to the observations that are assimilated, and, b) verification against independent observations from the ARM archive.

To take advantage of the GOES and ARM observations with frequent temporal coverage which capture both the conditions prior to formation of the clouds and their evolution, the data assimilation should be perform with a technique that allows inclusion of multiple times within a time window. Such techniques are known as "smoothing" techniques. These are the 4D variational (4DVAR) and 4D ensemble smoother techniques (Vukicevic et al., 2004a and van Leeuwen , 2001, respectively). The cloud resolving 4DVAR data assimilation system was already developed under partial support from the NSF, Project #DEB-9977066 in the period 9/1/99 – 8/31/04 (see past support for Vukicevic). The assimilation of the GOES imager infrared radiances in Vukicevic et al. (2004b) shows that 15 minute measurements within 1 hour window strongly constrained simulated cloud properties for a case of ice cloud evolution when conditions for the cloud formation were already present in the model. The high accuracy of estimated ice cloud mass and thickness was confirmed by comparison with the ARM SGP cloud radar measurements. The studies in Vukicevic et al. (2004a and 2004b) provide initial explicit evidence of the validity and feasibility of the cloud resolving satellite data assimilation. The results in these studies were limited, however, to relatively simple observed cloud evolutions and to the GOES imager measurements only. We propose that additional observations and more complex cases must be tested to fully address the accuracy of estimation of the cloud properties in 4D.

We plan to use the same 4DVAR algorithm with the cloud resolving model that was used in Vukicevic et al. (2004a and 2004b). This CRM is a version of the Regional Atmospheric Modeling System (RAMS, Cotton et al, 2003) The cloud resolving version of RAMS includes a 2-moment cloud microphysics parameterization for 5 ice hydrometeor types and rain, and 1-moment scheme for the cloud liquid with an assumed size distribution for all (Walko et al., 1995). This type of CRM is similar to other CRMs with the bulk cloud microphysics (Xu and Coauthors, 2002). While the bulk microphysics parameterization are not most complete in terms of the microphysical cloud processes they explicitly distinguish between different hydrometeor types and capture key interactions among them and with the environment. The bulk cloud microphysics parameterizations are suitable for the assimilation of GOES measurements because these measurements are to the first order most sensitive to the mass and concentration.

Greenwald et al. (2003, 2004) demonstrated that the GOES imager measurements are most sensitive to the mixing ratio parameter for variety of simulated cloud cases. The other important property of the hydrometeor explicit bulk cloud parameterization in 3D is that the results of data assimilation could be used to define other integral cloud properties of interest such as for example the precipitating water, nonprecipitating water, cloud fraction, mass flux, or spatial distribution of cloud radiative forcing.

- **We propose to evaluate integral cloud properties such as the liquid/ice water content profiles, cloud advective tendencies, vertical velocity and heating rate profiles. These quantities are not measured and could not be verified directly. Instead, we propose to analyze the sensitivity of the integral properties to the assimilated observations to evaluate physical consistency of the observations' impact.**

The bulk cloud microphysics in the CRM is also suitable for practical reasons. The more explicit cloud physics representations such as binned microphysics are *several orders* of magnitude more expensive computationally that the bulk microphysics. This property makes the explicit microphysics parameterizations unfeasible for application in the data assimilation or in NWP applications with the current and near future computational resources.

It the data assimilation it is obviously assumed that the model has skill, although not perfect, in representing the relationship between the observed and unobserved quantities. The model is defined here as a mapping from a space of control parameters (i.e., what is to be estimated) to the observed quantities. In general, the accuracy of data assimilation results must be controlled primarily by the strength of constraint or information content that the observations provide in the estimation because it is the observations that are by definition the best representation of the unknown true state. This property implies that not all of the control parameters in the model can be estimated equally well (Braswell et al, 2004; Vukicevic et al., 2004). The role of the model is to provide physically based dynamic correlations which result from the best knowledge of the governing laws. The model error in the data assimilation is not to be neglected, but is assumed that it would not dominate the data assimilation accuracy for the quantities to which the observations are most sensitive over short periods. If this is not the case the model would have no value in the assimilation. The analysis of possible model errors should be included in the data assimilation as discussed in sections 2 and 3.

The 4D data assimilation methodology and the already developed cloud resolving data assimilation system are described in section 2. The proposed research is detailed in section 3. The management plan is outlined in section 5.

2. Cloud resolving data assimilation

2.1 Short review of the estimation methodology

To describe the state, evolution and feedback mechanisms of a complex system such as the cloudy atmosphere, both models and observations are needed. The optimal specification of the state from both sources is obtained when the observations are included in the estimation of modeled state quantities. This defines the methodology of data assimilation.

The relationship between an estimate of the state to be derived from the observations and a prior modeled state is given by Bayes' rule

$$p(x_t / y) = \frac{p(y / x_t) p(x_t / x_t^f)}{\int p(y / x_t) p(x_t / x_t^f) dx_t} \tag{1}$$

where $p(x_t / y)$ is the probability density function of the true state x_t at time t given the observations y, $p(y / x_t)$ is the probability density of the observations given the true state, and $p(x_t / x_t^f)$ is the probability density of the true state given the prior knowledge of the state based on a modeled (forecast) state x_t^f. The states denoted by x are vectors of physical quantities with values given on a model grid.

The probability density of the observations given the true state is assumed to be Gaussian function

$$p(y / x_t) = (2\pi)^{-\frac{m}{2}} |R|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left(y - H(x_t) \right)^T R^{-1} \left(y - H(x_t) \right) \right] \tag{2}$$

where $H(x_t)$ is called the observational operator that simulates the observation vector y from the model state vector, R is the observation error covariance matrix, and there are m observations in y. For example the observational operator could be a radiative transfer model that maps from the cloud resolving model forecast of profiles of temperature, water vapor and cloud properties into satellite radiances across the doman at different wavelengths.

Assuming that best estimate of x_t is the state which maximizes (1), the prior and observation probability functions are Gaussian, and observation errors are not correlated with the state errors, t**hen the best estimate is obtained by minimizing the cost function**

$$J = \frac{1}{2}(x - x^f)^T B^{-1} (x - x^f) + \frac{1}{2}\left[y - H(x^f) \right]^T R^{-1} \left[y - H(x^f) \right] \tag{3}$$

where B is the modeled state (background or prior) error covariance matrix.

The minimization of (3) can be performed several ways. The most common methods of solution for a state x that is function of space (3D) and time are (Cohn, 1997; van Leeuwen, 2001): 3D variational analysis (3DVAR), 4D variational analysis (4DVAR; 4D = 3D + time), and Ensemble Kalman Filtering (EnKF). The first two methods are used in daily operational weather analysis: 3DVAR at the National Center for Environmental Prediction (NCEP), and 4DVAR at the European Centre for Medium Range weather Forecast (ECMWF).

We have implemented 4DVAR for the problem of explicit cloud analysis. The 4DVAR method is most suitable for this purpose because

1) It allows for simultaneous assimilation of all observations taken at irregular time points within a prescribed assimilation time interval. The 3DVAR and EnKF methods do not have this feature.

2) The modeled cloud physics and dynamics are mutually correlated in space and time through the modeled governing equations that are integrated over the assimilation period. This feature is also characteristic of the EnKF, but not of 3DVAR.

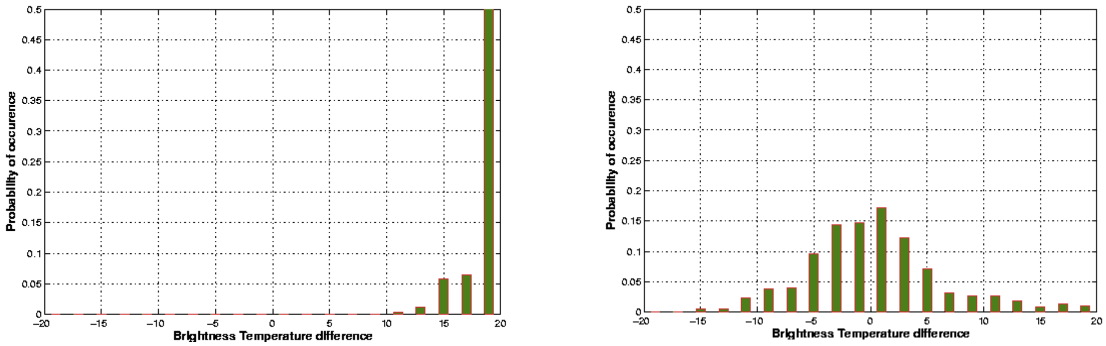
3) Unlike the EnKF method, the 4DVAR does not require pre-specification of a number of perturbations to be included in the adjustment of the modeled state to the observations. An actual number of required perturbations is not known and is likely very large for the problems with a large number of degrees of freedom such as the cloud analysis with dynamics. This would render the EnKF method currently computationally unfeasible.

The 4DVAR methods' undesirable features are:

1) Error estimates for the new state after the assimilation of observations are not directly available. Post processing to obtain the error estimates is required. In contrast to the 4DVAR method, the error estimates are available from the EnKF methods by design, but are limited to a subspace of states that is represented with the ensemble members. The Gaussian error distribution is typically assumed (van Leeuwen, 2001). We propose to estimate the posterior errors (i.e., the errors of the final estimate) in a post processing of the 4DVAR assimilation results using the approximations that are equivalent to those in the EnKF methodology (i.e., the errors are evaluated in a subspace and are normally distributed). We also propose to evaluate the posterior errors relative to independent observations that are not used in the assimilation. The proposed error analysis is detailed in section 3.

2) So called adjoint models are required for the observational operators and the cloud resolving model (RAMS in this proposal). These models are well defined in theory but are time consuming to develop. This feature is not a problem for this proposed research because the adjoint model for RAMS and the observational operators (GOES and standard meteorological station observations) have been already developed (Vukicevic et al., 2004ab; Greenwald et al., 2002, 2004; Zupanski et al., 2004). The entire 4DVAR data assimilation system with RAMS cloud resolving version is designated RAMDAS. This 4DVAR system was developed under partial support from the NSF, Project #DEB-9977066 in the period 9/1/99 – 8/31/04 (see past support for Vukicevic).

The 4DVAR cost function is minimized using the gradient of the cost function (computed via the adjoint models in RAMDAS) in an iterative preconditioned quasi-Newton algorithm (Zupanski et al, 2004). The 4DVAR algorithm is guaranteed to converge to the approximate minimum of the cost function, and hence the maximum of the posterior probability density function, if the probability function is Gaussian. If the first guess is suitably close to the global minimum, then the probability function is locally Gaussian, and the 4DVAR iterations will converge. The observation part of the cost function measures how close the model state is to the observations, and thus one can easily determine if convergence has been achieved. An example of the converged cost function is shown in Figure 1 from the RAMDAS experiments in Vukicevic et al (2004b) with the GOES imager infrared measurements. The observation part of the cost function in this example is defined as the difference between the observed and modeled brightness temperatures in the infrared window channels. Figure 1 shows that the associated brightness temperature errors at the end of assimilation has nearly zero mean value and small standard deviation around that mean. This result implies well behaved assimilation with achieved desired convergence to the Gaussian final error distribution.



(a) (b)

Figure 1: Prior (a) and posterior (b) errors of the estimated ice cloud brightness temperature in K in a domain of about 150 km ^{2} (from Vukicevic et al., 2004b).

2.2 Model constraint

If the background term in the cost function includes only the initial condition then the associated 4DVAR algorithm is called the *strong-constraint,* which assumes that the model is exactly error free. In the current research applications of the EnKF and 4VAR methods (e.g., van Leewen, 2001 and Lee and Lee 2003, respectively) this assumption is typically relaxed by adding a linear model error term to the model governing equations. This is equivalent to adding a linear forcing term to the prognostic tendencies. Assuming that the model errors are uncorrelated with the background and observation errors (the typical assumption), the cost function then becomes

$$J = \frac{1}{2}(x - x^f)^T B^{-1}(x - x^f) + \sum_t \frac{1}{2}\left[y_t - H(x_t^f)\right]^T R^{-1}\left[y_t - H(x_t^f)\right] + \sum_t \frac{1}{2}\varepsilon_t^T Q^{-1}\varepsilon_t$$

where ε_t is the linear model error at time t and Q is this error's covariance matrix. The linear model error approach was shown successful in Zupanski et al (2003) , Lee and Lee (2003) in applications with synoptic scale models. In the recent application with RAMDAS for the cloud resolving data assimilation in Vukicevic et al (2004b) the use of linear forcing model error was only partially successful: A faster convergence was achieved in some of the experiments with the linear model error but the assimilation results were less accurate than without this model error. The convergence was not achieved with the linear model error when the high spatial variability in the observed cloud features was assimilated. These results suggest that the specification of the model error needs to be refined for the application in the cloud resolving data assimilation . We propose to take an alternative approach where the model error will be treated as the error in the cloud microphysical parameterizations (section 3.4).

2.3 Regional Atmospheric Modeling and Data Assimilation System (RAMDAS)

The RAMDAS algorithm was developed at the Cooperative Institute for Research in the Atmosphere under major support by the Department of Defense's "Center for Geoscience/

Atmospheric Research" program and partial support by the NSF (see past support for Vukicevic). RAMDAS consists of 4 major components: 1) the nonlinear forecast model, RAMS, 2) observational operators, 3) the adjoint of the forecast model, and 4) the minimization algorithm. All of these components are described in detail in Vukicevic et al. (2004a), Zupanski et. al. (2004) and Greenwald et al., (2002, 2004). Only brief summary of key features relevant to the proposed work is presented.

Clouds and precipitation in RAMS are explicitly predicted via a microphysical parameterization that features a one-moment scheme (mixing ratio) for cloud liquid water and two-moment (mixing ratio and number concentration) for six other hydrometeor types, including pristine ice, aggregates, snow, graupel, hail and rain. The hydrometeor size distribution is approximated by a gamma distribution with a prescribed width parameter. Longwave and shortwave radiative fluxes are calculated with a two-stram model that allows radiative heating to influence the condensational growth of water droplets and ice particle vapor deposition. The adjoint model in RAMDAS is an adjoint of the true tangetnt linear RAMS: The linearization was performed with respect to the full model nonlinear solution at every time step (typically 10 seconds). The accuracy of the adjoint model is confirmed in Vukicevic et al. (2004ab) and Zupanski et al. (2004).

The observational operators in RAMDAS are

1) VISIROO: forward and adjoint models for computation of visible and infrared radiances in both clear and cloudy plane-parallel conditions. The models are based on the Spherical Harmonics Discrete Ordinate Method(SHDOM; Ebans 1998) and a delta-Eddington approach (Deeter and Evans, 1998). Extinction by gases is computed from the Optical Path TRANsmittance (OPTRAN) method (McMillen et al., 1995). The VISIROO is desribed in detail in Greenwald et al. (2002, 2004).

2) 3DVAR-WRF for assimilation of standard meteorological observations that are available from the NCEP operational data assimilation system (see Zupanski et al., 2004, for further detail).

The minimization algorithm described in Zupanski et al. (2004) is state of the art efficient limited memory quasi-Newton type algorithm with a restart procedure. Empirical so called Hessian preconditioning is employed to reduce the number of minimization iterations. The model variables that are adjusted to the observations are: the potential temperature, Exner perturbation function, vertical wind, velocity potential, stream function, total water mixing ratio, cloud hydrometeor mixing ratio and number concentrations.

3. Proposed Research

3.1 Assimilation of the GOES imager and sounder radiances

As stated in the introduction, the GOES imager visible and infrared measurements are traditionally used for the cloud identification and analysis because of the spectral characteristics and high horizontal spatial resolution (1- 4 km) and temporal frequency (15 minutes) over large areas including the North America (e.g., Greenwald and Christopher, 2000). The GOES sounder measurements are routinely used in NWP to retrieve the atmospheric temperature and humidity at different vertical levels in the cloud free conditions and for the upper level cloud information (Menzel et al., 1998). The GOES sounder radiance measurements have about 12 km horizontal resolution and 1 hour frequency. In the RAMDAS observational operator for the visible and infrared wavelengths (VISIROO) the radiative transfer models are capable of computing both the GOES imager and sounder radiances.

The unique feature of the VISIROO models is that the GOES sounder radiances are computed for both the clear and cloudy plane-parallel atmosphere unlike in the current NWP applications (Menzel et al, 1998). This feature will allow that the GOES sounder measurements will be applied in the cloud resolving data assimilation experiments simultaneously with the GOES imager in every grid point in the model and over the assimilation time interval. The desired effect will be consistent improvement of the cloud properties, temperature and humidity in the entire 3D domain and over the evolution of the observed cloudy atmosphere within the given time interval. **The assimilation of GOES radiances in the proposed study will not require new algorithm development.**

We propose to investigate how much each of the GOES channels would contribute to the estimation of the cloud properties and the associated environment and the dependence of this contribution on the type of cloud evolution. This analysis will help to evaluate independent information content of this type of measurements and will aid in selection, if any, of a subset of measurements to propose for the application in the future NWP analysis in the cloudy conditions.

3.2 Addition of ARM data

Our strategy is not to use every ARM data stream, but to focus on those that are most relevant for cloud properties and are available at multiple facilities across the Southern Great Plains (SGP) domain. Table 1 lists those instruments whose data we propose to assimilate in addition to the GOES visible and infrared measurements.

The surface observations and radiosondes (BBSS) are the usual datasets assimilated for mesoscale modeling and are important source of temperature, water vapor, and wind information. The data assimilation system already handles surface and sonde measurements. A large amount of water vapor data is available from the dense network of GPS stations in the SGP area. For each GPS station the integrated water vapor along the slant paths to an average of 8 GPS satellites is available every 30 seconds. We propose to assimilate GPS retrievals of the slant path water vapor for which the forward model is simply integration of the water vapor density along the satellite ray paths. The GPS data is dense enough to allow reconstruction of the 3D water vapor field when surface and some limited profile information is introduced (e.g., MacDonald et al., 2002).

The five microwave radiometers provide precipitable water vapor and cloud liquid water path. We will assimilate the two brightness temperatures directly, rather than PWD and LWP products. The high resolution infrared radiance measurements (AERI in Table 1) will be use to constrain the cloud base temperature when the clouds are optically thick while for the optically thin clouds these measurements will help constrain optical depth and particle size information.

The radar measurements (MMCR) will provide cloud profile information, though only at the central facility. The MMCR will initially be used for verification purposes as in Vukicevic et al. (2004). The assimilation of the MMCR will be tested but it is expected that this observation's constraint will not have large impact. The broadband flux measurements made by NIP, shaded PSP and PIR instruments will be used for verification of the modeled surface radiative fluxes. Direct and difuse shortwave flux is mainly related to cloud (and aerosol) optical depth and cloud cover, though this relation is complicated by cloud heterogeneity which will be partially resolved in the cloud resolving model vertical domain. The assimilation of these measurements may contribute to further constrain these cloud parameters (after the assimilation of GOES and MWR and GPS) and will be tested. The longwave flux is affected by the temperature profile, precipitble water vapor and cloud cover, emissivity, and base height. These measurements will also be used in the verification and additional assimilation.

Table 1: ARM and other data streams proposed for assimilation with RAMDAS1

Instrument	Measured quantities	Locations	Frequency
SMOS	p, T, q_v, u, v	CF + 14 EF	1 min
OK Mesonet	p, T, q_v, u, v	50 sites in SGP	5 min
BBSS	p, T, q_v, u, v profiles	CF + 3 IF + NOAA	6 h
GPS	Slant path water vapor	24 near CF + 20 over SGP	30 sec
MWR	T_b at 23.8 and 31.4 GHz	CF + 4 BFs	20 sec
AERI	IR radiance spectra	CF + 4 EF	20 sec
IRT	9.6-11.5 μm	CF + (≥ 6 in 2005)	20 sec
MFRSR	Direct and diffuse flux	CF + 21 EF	20 sec
MMCR	35 GHz radar reflectivity	CF	10/40 sec

CF is central facility, BF are the boundary facilities, EF are extended facilities and IF are intermediate facilities. T is air temperature, p is pressure, q is waer vapor, u and v are horizontal wind components and T_b is brightness temperature.

3.3 Accuracy analysis

Evaluating the accuracy of the estimated cloud properties and surrounding atmosphere that will be produced in the assimilation is critical to the proposed project. As described in section 2, the 4DVAR framework does not traditionally provide error estimates directly. We propose to compute the error estimates in the following way:

1) Posterior error variance estimate in the control parameter space
2) Domain error statistics in observation space
3) Comparison with independent measurements

1) Posterior error variance estimates in the control parameter space will be produces using an estimate of the Hessain matrix that is associated with the cost

function in (3). In the framework of the assumed Gaussian error distributions in (1) the error distribution of the estimate is also Gaussian with the covariance matrix

$$P = \left[B^{-1} + H^T R^{-1} H\right]^{-1} = \left(\frac{\partial^2 J}{\partial x^2}\right)^{-1} \tag{4}$$

where the second order derivative matrix is the Hessain of the cost function J in (3). It is not possible to evaluate the entire Hessian matrix for the huge cloud resolving model state vector which is of the order of 10^7 elements. Instead we propose to estimate the errors of each variable and at each model level using a preconditioning procedure in Zupanski et al. (2002). This procedure transforms the original model variables so that the errors in the transformed state (relative to the unknown truth) at the observation times are directly proportional to the perturbations in the state at the initial time. The preconditioning is normally applied in the minimization at the beginning of the procedure. The same transformation can be, however, performed also at the end of assimilation (i.e., at the minimum of the cost function). This would then render an estimate of the posterior model state errors at the observation time. The transformation procedure involves evaluating the change of the cost function at the minimum and the associated cost function gradient with respect to the model variables. The error estimates using the preconditioning procedure would involve very small amount of additional computations because the cost function and its gradient are already computed as part of the 4DVAR algorithm. To test the validity of the error estimates that would be obtained with the preconditioning method we propose to also perform an ensemble evaluation of the errors using the Monte-Carlo approach. In this approach an ensemble of random perturbations is made in the model initial conditions around the minimum solution to produce an ensemble of states at all observation times by forward integration of the model (e.g., Tarantola, 1997).

2) The domain error statistics in the observations space will be produced using so called residual at the end of assimilation which is defined as: point by point differences between the model solution after assimilation and the GOES radiance observations. The posterior error defined in this way is simply the 2D field in the radiance space for each observed narrow band of wavelengths (i.e., for each GOES channel). The standard error statistics such as mean, median, standard deviation, root mean square error and empirical probability density function will be then computed. The example of this error analysis is shown in Figure 1. The residual errors could also be categorized by either measurement characteristics such as high, mid and low wavelengths or by the cloudy conditions such as the cloud type to represent the measure of success as it would depend on the state characteristics.

3) The comparison with the ARM SGP measurements will be performed for all the measurements that are discussed in the previous section (2.3). The integral cloud properties such as the liquid/ice water content, cloud fraction, the precipitable water and precipitation rate will be also compared to the equivalent measurements or site retrievals. **The comparison to the ARM measurements will provide strongest measure of the accuracy but will be limited to the locations and quantities that are observed.** The information content of

this error measure will be extended by evaluation of the time series and time averages, similarly to what is done traditionally with the ARM site measurements to extend their representativeness in the large domain.

3.4 Updates in the data assimilation system

During the course of the project few minor upgrades of the existing state of the art research data assimilation system will be made.

- The first modification will be to convert the cloud control parameters error statistics from the standard symmetric Gaussian to the log-normal. This change will directly include the property of positive-definiteness in the data assimilation instead of the current configuration which includes the positive-definite property in the cloud resolving model only. This change in RAMDAS is straightforward to implement as the existing Gaussian framework can be used by substituting $\ln(r_c)$ in r_c (the cloud mixing ratio or number concentration).
- The second minor modification in RAMDAS will include adding the post-processing of the errors that is described in the previous section.
- The last modification will include adding some physical parameters from the cloud parameterization to the set of control parameters in the data assimilation. In this way the model error as represented with uncertainties in the particular parameters will be treated explicitly in the assimilation. To determine what parameters may be of most interest, a sensitivity analysis will be performed using the nonlinear forward and also adjoint CRM version of RAMS.

3.5 Cases and experiment design

The cloud resolving data assimilation computational requirements are very heavy for large grids and long time periods. We plan to use the supercomputing facilities at the National Center for Atmospheric Research in Boulder. The RAMDAS algorithm is already in use at NCAR IBM system in the ongoing NSF supported project on the analysis of carbon fluxes from the field campaign data in complex terrain (see current funding for Vukicevic). The NCAR supercomputers are powerful but the computational resources are not unlimited. To make best use of the computational resources we will make careful selection of the observed few cases lasting 1-3 days each. The cases will include examples of typical stratiform or forced convective cloud evolutions over the SGP domain. The periods will be selected such to capture precursors to the cloud formation, the growth and possibly dissipation phase.

The model horizontal grid resolution will range 1-4 km depending on the case. The model domain will be also selected based on the case characteristics that would include consideration of the advective time scales to reduce the errors from the lateral boundaries by design. The model lateral boundary conditions will be produced from the currently best regional weather analysis that is available from NCEP. The EDAS grided weather analysis with 32 km horizontal resolution will be used. This is the same type of weather analysis data are already used in the carbon flux project with RAMDAS, thus the application in the proposed cloud resolving data assimilation will not require additional software development. The lateral boundary conditions in the data assimilation are included in the control parameter set (i.e. the set of variables that are changed by the observations) and would be also improved in the assimilation. The uncertainties/errors at the lateral boundaries will be monitored carefully to diagnose the effect of downscaling in the regional data assimilation.

4. Management plan

4.1 Schedule

Year 1: Implement log-normal cloud statistics and perform sensitivity of the cloud microphysics to select parameters to be included in the control set in the data assimilation experiments. Code and test the forward and adjoint parts of the observational operator for the new ARM data streams (GPS slant path water vapor, MWR and AERI radiances, MFRSR flux, and MMCR reflectivity profiles). Select some of the time periods and prepare the GOES and ARM data for assimilation.

Year 2: Develop algorithm for the posterior error analysis. Begin assimilation of the GOES and ARM data for one of the selected cases.

Year 3: Complete the assimilation experiments and the error analysis.

4.2 Personnel

The proposed project will be joint effort between Dr. Tomislava Vukicevic and a graduate research assistant in collaboration with Dr. Frank Evans in the Program in Atmospheric and Oceanic Sciences (PAOS) at the University of Colorado, Boulder. Dr. Vukicevic will lead the development of the proposed necessary upgrades in the data assimilation system and will supervise a graduate student in PAOS starting in August 2005. Dr. Evans will lead the development of observational operators for the ARM data stream and will aid in their implementation in RAMDAS with Dr. Vukicevic. The graduate student will perform data assimilation experiments and innovative error/accuracy analysis .

Dr. Vukicevic also has affiliation with the Cooperative Institute for Research in the Atmosphere at the Colorado State University where she is Head of the Data Assimilation Research Group. Her ongoing collaboration with the colleagues in this group which focuses on further development of the data assimilation with satellite remote sensing in the hydrologic cycle research will complement her participation in the proposed project.

Long Term Measure of Scientific Advancement

Cloud feedbacks are the largest source of uncertainty in predictions of the climate sensitivity, and hence in determining acceptable levels of greenhouse gases for a given climate change. Another major source of uncertainty in prediction on short to long time scales is intensity and distribution of precipitation which is naturally linked to the cloud formation and evolution . The major goal of satellite and ground based observing systems is to provide data for the analysis of the cloud properties and the environment but the full complexity of the 3D cloud structure within dynamic atmosphere cannot be determined from the observations alone. Our proposed research will investigate integration of the satellite and ground based data streams with a cloud resolving model to produce cloud properties and the environment in 3D over time. We will evaluate the accuracy of estimated cloud properties and the atmospheric environment to provide measure of observability by currently best systematic cloud observations. Dependence of the accuracy on modeling errors will be also addressed to provide directions for modeling improvements. The results will be directly relevant to evaluating facility of new generation of Numerical Weather Prediction Models which are rapidly approaching the resolution of the Cloud Resolving Models.