Cloud and Precipitation Parameterizations in Modeling and Variational Data Assimilation: A Review

PHILIPPE LOPEZ
European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received 23 January 2006, in final form 16 May 2006)

ABSTRACT

This paper first reviews the current status, issues, and limitations of the parameterizations of atmospheric large-scale and convective moist processes that are used in numerical weather prediction and climate general circulation models. Both large-scale (resolved) and convective (subgrid scale) moist processes are dealt with.

Then, the general question of the inclusion of diabatic processes in variational data assimilation systems is addressed. The focus is put on linearity and resolution issues, the specification of model and observation error statistics, the formulation of the control vector, and the problems specific to the assimilation of observations directly affected by clouds and precipitation.

1. Introduction

Over the past 40 yr, our ability to forecast the weather and to simulate the climate of our planet with numerical models has strongly benefited from the growing understanding of the ocean–atmosphere system and from the constant progress in computer technology. Whether a given numerical model is able to properly forecast the state of the atmosphere up to 10 days or to produce realistic climatic simulations on monthly time scales strongly depends on its ability to represent the observed spatial distribution of clouds and their various effects on the environment, through latent heat release, radiative effects, and precipitation (Browning 1993). Taking advantage of the precious information about moist processes brought by observing systems and laboratory experiments, numerical weather prediction models (NWPMs) and (climate) general circulation models (GCMs) have become able to describe clouds and precipitation with an increasing level of realism and accuracy. However, current parameterizations of moist processes still rely on crude simplifying assumptions either for the sake of computational efficiency or because of uncertainties about individual processes, in particular microphysics and cloud-scale transport.

In the past 10 yr, some progress has also been achieved in the assimilation of observations affected by clouds and precipitation in NWPMs with the aim of producing more realistic initial atmospheric states (or analyses). Such measurements are already widely available with a large geographical coverage from various spaceborne instruments such as the Special Sensor Microwave/Imager (SSM/I), or platforms such as the Tropical Rainfall Measuring Mission (TRMM) or the Aqua satellite. Future missions including the Global Precipitation Measurement (GPM; planned in 2010) mission and CloudSat (launched in April 2006) are expected to provide better spatial and temporal samplings and more information about the vertical structure of clouds and precipitation thanks to their onboard radars. Observations from the ground are available from the national operational networks of precipitation radars and rain gauges as well as from several experimental sites equipped with microwave radiometers, cloud lidars, and radars, such as those of the Atmospheric Radiation Measurement Program (ARM; Stokes and Schwartz 1994).

So far four main methods have been developed to assimilate this kind of data: nudging (Macpherson 2001), diabatic or physical initialization (e.g., Puri and Miller 1990; Krishnamurti et al. 1993; Treadon 1996; Ducrocq et al. 2002), and n-dimensional variational...
data assimilation (nDVAR with \( n = 1, 2, 3, \) or 4). The alternative and interesting method of an ensemble Kalman filter was also investigated (e.g., Tong and Xue 2005) but has not been extensively applied in operational global data assimilation systems yet. Numerous feasibility studies using the variational approach on satellite or ground-based observations were conducted (e.g., Zupanski and Mesinger 1995; Treadon 1997; Tsuyuki 1997; Hou et al. 2001; Peng and Zou 2002; Marécal and Mahfouf 2002; Marécal and Mahfouf 2003; Moreau et al. 2004; Hou et al. 2004; Benedetti et al. 2005; Lopez et al. 2006). All these works identified the main issues to be solved and showed that the variational assimilation of precipitation observations could bring some improvement in the analyses as well as in the subsequent forecasts. Operational satellite rain-rate assimilation has become reality in 3DVAR at the National Centers for Environmental Prediction (NCEP; Treadon et al. 2002) and in 4DVAR at the Japan Meteorological Administration (JMA; Tsuyuki et al. 2002). Since June 2005, rainy 19- and 22-GHz brightness temperatures from the SSM/I are operationally assimilated in 4DVAR at the European Centre for Medium-Range Weather Forecasts (ECMWF; Bauer et al. 2006a,b).

The progress toward the inclusion of parameterizations of moist processes in variational data assimilation systems and toward the treatment of observations affected by clouds and precipitation has been slowed down by some important constraints such as linearity, an accurate description of moist processes, and the specification of proper error statistics for both model and observations. Many of these problems still await more accurate solutions (see also Errico et al. 2007).

This paper offers an up-to-date review of the use of moist physical parameterizations in weather forecast models and in variational data assimilation. It is a follow-up to the Joint Center for Satellite Data Assimilation (JCSDA) Workshop on the Assimilation of Satellite Cloud and Precipitation Observations that was held on 2–4 May 2005, near Washington, D.C. First, section 2 focuses on the approaches that have been proposed to forecast resolved and subgrid-scale moist processes in three-dimensional NWPMs and GCMs and discusses the main related issues. Note that parameterizations especially designed for idealized or single-column models will not be considered here. Section 3 presents how moist processes can be included in variational data assimilation systems, with a particular attention to the treatment of observations that are directly affected by clouds and precipitation. Both recent advances on this topic and the problems yet to be solved are presented. A summary in section 4 concludes the paper.

2. Forward modeling of clouds and precipitation

\subsection*{a. Generalities}

The time and space discretization involved in all NWPMs and GCMs provide a discrete representation of the atmosphere while in reality dynamical and physical processes operate on a continuous spectrum of spatial and temporal scales. In the case of clouds and precipitation, a rather artificial but unavoidable distinction needs to be made between resolved moist processes that can be explicitly calculated from the large-scale atmospheric conditions on the model grid, and subgrid-scale moist processes whose indirect effects (condensation/evaporation, transport) on the model variables must be estimated with some suitable parameterizations and added to the resolved scales. So far, resolved and subgrid-scale moist processes have often been referred to as stratiform and convective, respectively, a terminology that does not necessarily match the actual definition of the stratiform versus convective distinction in the real world. This becomes particularly obvious considering that typical horizontal resolutions used in current NWPMs and GCMs can range from several hundreds down to a few kilometers.

If \( s = C_p T + g z \) denotes the dry static energy (where \( T \) is the temperature, \( z \) the altitude, \( g \) the constant of gravity, and \( C_p \) the specific heat capacity at constant pressure), the equations governing the evolution of the grid box mean values of \( s \) and of specific humidity, \( q \), in the presence of moist processes can be written as (e.g., Anthes 1977)

\begin{align}
\frac{\partial s}{\partial t} &= - \mathbf{V} \cdot \nabla s - \omega \frac{\partial s}{\partial p} + LC - \frac{\partial \omega \bar{s}}{\partial p} + C_p Q_r, \tag{1} \\
\frac{\partial q}{\partial t} &= - \mathbf{V} \cdot \nabla q - \omega \frac{\partial q}{\partial p} - C - \frac{\partial \omega \bar{q}}{\partial p}, \tag{2}
\end{align}

where \( \mathbf{V} \) and \( \omega \) are the horizontal and vertical components of the wind, \( L \) is the latent heat of vaporization, \( C \) is the rate of large-scale condensation/evaporation, and \( Q_r \) is the radiative heating/cooling rate. Primes denote subgrid-scale fluctuations and overbars indicate an averaging of subgrid-scale fluxes over the model grid box. Resolved moist processes are summarized in \( C \) while the contribution of subgrid-scale moist processes is included in \( \frac{\partial \omega \bar{s}}{\partial p} \) and \( \frac{\partial \omega \bar{q}}{\partial p} \). Note that the two latter terms also include contributions from other subgrid-scale phenomena such as turbulence.

Before discussing the ingredients that are needed to build efficient moist physical parameterizations, it should be recalled that an accurate modeling of the dynamical forcings is, of course, an absolute prerequi-
site to any realistic simulation of clouds and precipitation.

b. Resolved moist processes

1) Existent parameterizations

Many prognostic large-scale cloud schemes suitable either for NWPMs or GCMs have been proposed in the literature during the last 30 yr; these can be divided into three main categories. The list of references provided here is not meant to be exhaustive.

The first type that can be referred to as all-or-nothing schemes corresponds to schemes that do not account for subgrid-scale variability inside the model grid box. Cloud fraction is either 0% or 100% depending on whether the model grid box mean relative humidity is below or above 100%, and any supersaturation is instantaneously removed through cloud and precipitation formation. This assumption is valid when the model resolution is better than a few hundred meters but is more questionable at coarser resolution, especially in GCMs. For instance, Fowler et al. (1996) developed a large-scale condensation scheme for the Colorado State University (CSU) GCM with a separate prognostic treatment of cloud liquid water, cloud ice, and snow contents. Their scheme also included physically based parameterizations of autoconversion of cloud condensate, depositional growth of snow, multiphase collection of cloud condensate by precipitation, and the Bergeron–Findeisen mechanism. However, they acknowledged that not allowing partial cloud cover in their scheme could be problematic in their highly detailed radiative and microphysical calculations in which in-cloud condensate amounts (i.e., grid box mean amounts divided by cloud fraction) ought to be used.

The second type of parameterizations is based on the specification of a critical relative humidity threshold (lower than 100%) above which clouds start to form, which permits the calculation of a fractional cloud cover in each model grid box. This approach seems more appropriate than the previous one at resolutions coarser than 5 km used in NWPMs and GCMs. Sundqvist’s (1978) pioneering cloud scheme, later refined in Sundqvist et al. (1989), was capable of predicting the evolution of nonconvective cloud water with crude parameterizations of autoconversion and collection processes and a diagnostic relationship between cloud fraction and grid box mean relative humidity. Under the basic hypothesis that stratiform cloud formation can be determined from the temporal evolution of saturation specific humidity, Tiedtke (1993) proposed a scheme able to predict both cloud condensate and cloud fraction. He also included a contribution of convective activity to large-scale cloud formation through the detrainment of updraft liquid water into the environment ( anvils). Lohmann and Roeckner (1996) and Rasch and Kristjánsson (1998) designed their own cloud schemes with prognostic cloud water and ice, diagnostic cloud fraction, and individual parameterizations of autoconversion, accretion, aggregation, freezing, and riming. Zhao et al. (1997) developed a scheme inspired from Sundqvist et al. (1989) with prognostic cloud condensate for the NCEP Eta Model.

The third category of schemes, often referred to as statistical schemes, was pioneered by Sommeria and Deardorff (1977) and Mellor (1977) who assumed that the subgrid-scale variability of a synthetic thermodynamical variable can be accounted for in a statistical point of view through the specification of some suitable probability density function (PDF). Most commonly, schemes have been based on the PDF of total water, \( q_t \), that is the sum of specific humidity, \( q \), and cloud condensate, \( q_c \). In this case, the cloud fraction, \( f \), and \( q_c \) can be calculated through a partial integration of the PDF, denoted \( G \), above the saturation specific humidity, \( q_{sat} \) (see Fig. 1):

\[
f = \int_{q_{sat}}^{\infty} G(q_t) \, dq_t
\]

\[
q_c = \int_{q_{sat}}^{\infty} (q_t - q_{sat})G(q_t) \, dq_t.
\]

Le Treut and Li (1988) assumed a uniform distribution of total water to derive a simple parameterization with prognostic cloud water content for use in the Laboratoire de Météorologie Dynamique’s (LMD, France) GCM. Starting from the latter scheme, Boucher et al. (1995) added a prognostic variable for rainwater with a more detailed representation of the autoconversion and collection processes. As they mentioned it, the main limitation of their work was the crude treatment of ice that was supposed to reach the ground within each 30-min time step. Smith (1990) used the same statistical framework for developing a large-scale condensation scheme with prognostic cloud condensate and a triangular PDF for total water. Starting from a similar assumption, Rotstyn (1997) developed a stratiform cloud scheme with prognostic cloud liquid water and cloud ice for the Commonwealth Scientific and Industrial Research Organization (CSIRO, Australia) GCM, while Wilson and Ballard (1999) introduced a scheme with two prognostic variables (cloud liquid water plus vapor on the one hand, and total ice on the other hand) in the Met Office’s Unified Model. Lopez (2002) proposed a parameterization of large-scale moist processes...
with one prognostic variable for cloud condensate (liquid plus ice) and another for precipitation (rain plus snow) that can be used in NWPMs and data assimilation.

In contrast to the previous statistical schemes in which the assumed total water distribution was independent of meteorological conditions, Ricard and Royer (1993) and Lohmann et al. (1999) allowed their respective PDF variance to be modulated by turbulence inside the planetary boundary layer. Tompkins (2002) added even more complexity in his scheme by including the effect of both turbulence and convective activity on his beta distribution through two additional prognostic equations for variance and skewness.

In cloud-resolving models (CRMs) that have been used only for research purposes so far, the expected level of details, the small time step (less than 1 min), and the high horizontal resolution (less than 1 km) have led to the development of complex and computationally expensive parameterizations of moist processes. Schemes of this type would usually include separate prognostic variables for cloud liquid water, drizzle, rainwater, small ice particles, graupel, snow, and hail; different PDFs for each variable; and a detailed representation of microphysical exchanges between categories (e.g., Ferrier 1994; Pinty and Jabouille 1998). A detailed review of parameterizations in CRMs is, however, beyond the scope of the present paper.

An alternative approach to the prognostic description of the main statistical moments of a PDF is offered by spectral bin models (SBMs) in which the particle size distribution of each hydrometeor type is divided into spectral bins and the number of particles in each bin is treated as a prognostic variable. A recent example of the use of a SBM in the Pennsylvania State University–National Center for Atmospheric Research (Penn State–NCAR, United States) Mesoscale Model (MM5) can be found in Lynn et al. (2005), who considered 33 bins for each one of their 8 particle categories. However, the excessive cost of SBMs prohibits their use in any operational model for the moment.

2) MAIN ISSUES IN THE MODELING OF LARGE-SCALE CLOUDS AND PRECIPITATION

The first capability that should be expected from a parameterization of large-scale condensation is to produce realistic profiles of cloud properties and precipitation, especially in terms of fractional coverage and amount, given their strong impact on radiation.

Issues that arise when developing a new parameterization concern the choice of a prognostic formulation and the choice of the basic variables (e.g., cloud fraction, cloud liquid water, cloud ice, rain, snow, and so forth). Very different approaches have been adopted in the literature (see section 2b), ranging from a fully diagnostic representation of clouds and precipitation to the prognostic treatment of up to eight separate variables. In fact, the inclusion of prognostic variables for cloud properties is probably not essential unless accurate long-term radiative budgets are expected (climate scenarios) or unless high-resolution (say, less than 5 km) NWPM simulations are to be performed [limited-area models (LAMs) or CRMs]. The advection of cloud properties, which is sometimes taken as an argument in favor of their prognostic treatment, would only be relevant in the unlikely case of a high-resolution model run with a large time step or in some rare occurrences of very strong winds. Furthermore, a realistic propagation of cloud systems can be obtained with a diagnostic scheme through the sole advection of specific humidity. Prognostic partitioning between cloud liquid water and cloud ice is probably not crucial either and a simple

![Figure 1. Utilization of a specified PDF, \( G \), to describe the subgrid-scale variability of total water, \( q_t \), inside a model grid box. Here \( f \) denotes fractional cloud cover obtained by partial integration of the PDF over the gray shaded area.](image-url)
diagnostic treatment through a simple temperature dependence (e.g., Ryan 1996) can suffice. As far as precipitation variables are concerned, raindrops’ terminal fall velocities are on the order of 5 m s⁻¹, which supports the idea that a diagnostic treatment is sufficient (i.e., all rain formed reaches the ground within one time step), while the fact that snowfall speeds seldom exceed 0.5 m s⁻¹ advocates a prognostic treatment of this type of hydrometeors in most models. This issue will be discussed further in section 3c.

As for microphysics, the prognostic representation of a given process is not necessary as long as the model time step is kept larger than the time scale of this process. Condensation, autoconversion and collection in warm clouds (above freezing point) can be assumed to be fast processes compared to typical model time steps used in NWPMs and GCMs (~10–60 min). On the contrary, this assumption is more questionable for ice clouds, especially cirrus clouds, because heterogeneous nucleation and deposition/sublimation processes can exhibit time scales of up to several hours (Khvorostyanov and Sassen 1998).

In the current NWPMs and GCMs, the inclusion of subgrid-scale variability in statistical cloud schemes by means of a PDF usually provides reasonable forecasts of clouds and precipitation. However, when accurate radiative calculations are expected from either a GCM or a NWPM (e.g., for assimilating radar reflectivities), prognostic equations that describe the evolution of the PDF characteristics due to turbulence and convective processes become highly recommendable (Tompkins 2002). The relevance of considering a given moment of the PDF decreases with increasing order of the moments: predicting the mean is essential, variance is important, skewness can be beneficial in some convective situations (Wood and Field 2000), and so on. Besides, as mentioned throughout section 2b, various basic PDF shapes (uniform, triangular, normal, bimodal, etc.) have been tested in the literature, but there is currently no clear consensus on the optimal type and number of basic shapes that ought to be represented in parameterizations. In fact, the uncertainty on PDFs is large even from the observational point of view. On the one hand, in situ observations episodically performed by aircraft or balloons over limited portions of selected cloud systems suffer from an obvious lack of representativeness (undersampling both in time and space). On the other hand, indirect retrievals of PDF parameters from satellite are still inaccurate because of their rather coarse resolution and because they rely on predefined PDF shapes. Nonetheless and despite its nadir-only field of view, the 94-GHz cloud radar of the CloudSat mission (launched in 2006; Stephens et al. 2002) is expected to provide more accurate and higher-resolution information on the horizontal and vertical distributions of cloud condensate (Austin and Stephens 2001; Benedetti et al. 2003) when combined with other passive instruments of the A-Train satellite constellation (Stephens et al. 2002). Such data might help to develop more universal cloud parameterizations.

Another relevant issue relates to the so-called overlap assumption that describes the way partially cloudy model layers are arranged in the vertical inside a given grid box. Atmospheric forecast models currently assume maximum-overlap, random-overlap, or combined maximum-random-overlap (Morcrette and Jakob 2000). These are illustrated in Fig. 2 and the corresponding formulae to compute the total cloud cover, \( f_{\text{TOT}} \), from the cloud fraction, \( f_k \), at each model level number \( k \) are also given. According to the maximum-random overlap assumption, maximum overlap is applied to adjacent layers while random overlap is assumed for separate cloud layers. The cloud overlap assumption has a significant impact on the radiation calculations through the computation of total cloud cover, but also on the simulation of precipitation evaporation that depends on the location of subsaturated subcloud layers in the vertical (Jakob and Klein 2000).

Clearly, all these questions deserve to be answered, but even bigger uncertainties remain on how to parameterize subgrid-scale moist processes.

c. Subgrid-scale moist processes

1) EXISTING PARAMETERIZATIONS

With the current horizontal resolutions of at least several kilometers, operational NWPMs as well as GCMs are not able yet to resolve convective activity explicitly. In nature, the main role of convection is to remove vertical instability in the atmosphere through a vertical redistribution of moisture, through the heating of upper-tropospheric levels due to condensation in the updrafts and to large-scale subsidence outside convective cells, through the large-scale compensating subsidence around convective areas, and through the low-level cooling caused by precipitation–evaporation induced downdrafts. Weather and climate modelers soon realized that some way of including the impact of convection on their model’s large-scale variables was needed to obtain realistic simulations of the atmosphere and of the hydrological cycle. Many parameterizations have been proposed in the literature; these can be classified into three categories.

In the first category, the so-called moist convective adjustment schemes, the initial atmospheric profiles of
T and q at a given grid point are simply relaxed to some predefined reference profiles (often pseudoadiabatic). Among those, Kuo (1965, 1974) assumed a fixed partitioning of the local moisture convergence (due to large-scale advection and surface evaporation) into moistening of the environment and convective precipitation release. Betts (1986) and Betts and Miller (1986) developed an adjustment scheme in which the latter partitioning is no longer fixed but depends on the subsaturation level of the environment, and in which the relaxation time scale is set to about two hours.

In the second category, the so-called mass-flux convection schemes, the final profiles of model variables are calculated by explicitly describing the properties (including the mass flux) of a single or multiple convective updrafts. In nature, however, a whole spectrum of convective clouds reaching different heights are often observed so that the single updraft described in such parameterizations is actually meant to represent the bulk effect of an ensemble of cumulus clouds at the scale of the model grid box. A closure assumption is required to link the actual amount of convection (mass flux, precipitation) to the model control variables. In most parameterizations, the closure of the scheme is achieved by relating the cloud-base mass flux to the vertically integrated moisture convergence (Kuo 1965), the low-level mass convergence (Frank and Cohen 1987), or the release of convective available potential energy (CAPE; Fritsch and Chappell 1980). Arakawa and Schubert (1974) proposed a mass-flux scheme based on a spectrum of cloud types that are assumed to be in quasi equilibrium with their environment. Tiedtke (1989) and Gregory and Rowntree (1990) developed their respective mass-flux schemes using a bulk representation of a cumulus cloud ensemble (i.e., a single-cloud model) and a closure based on moisture convergence in the former and on the stability of the lowest convective layers in the latter. Kain and Fritsch (1990) introduced an original estimation of entrainment/detrainment based on Fritsch and Chappell’s (1980) statistical parameterization of the turbulent mixing of environmental and updraft air in deep convection. Using the same framework, Bechtold et al. (2001) designed a numerically efficient mass-flux parameteriza-
tion of both shallow and deep convection based on a CAPE closure that is applicable to a wide range of horizontal resolutions.

The third category of convective parameterizations is based on the concept of buoyancy sorting (Raymond and Blyth 1986), in which subcloud-scale drafts move from one level to the next where they mix with the environment, thereby creating mixtures that can in turn initiate one or several new drafts. These are then allowed to rise or descend according to their buoyancy. Emanuel (1991) applied this concept to create his own parameterization and closed his scheme through precipitation efficiency and the fraction of precipitation falling in unsaturated air. Contrary to Emanuel (1991), Hu (1997) allowed each draft to give birth to multiple mixtures at a given model level. Parameterizations of this category are rather expensive to run and therefore have only been used in rather coarse-resolution models so far.

In the case of CRMs, the horizontal resolution is usually below 1 km and convection is assumed to be resolved explicitly with no need of a specific parameterization.

2) MAIN ISSUES IN THE MODELING OF CONVECTION

All convective parameterizations previously described have their own limitations. Besides, several uncertainties remain about the actual characteristics and behavior of convection in the real world.

For instance, except for indirect reconstructions from a few field experiments that were conducted in regions of deep or shallow convection, such as the Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment (GATE), Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE), or Barbados Oceanographic and Meteorological Experiment (BOMEX), or from large-eddy simulations (e.g., de Roode et al. 2000), little is known about entrainment/detrainment that is the mixing of convective updrafts and downdrafts with environmental air (e.g., Yanai et al. 1973; Cohen 2000). In mass-flux schemes, entrainment and detrainment rates that directly determine the mass-flux vertical profile are always specified in a highly simplified and therefore highly uncertain way, based on laboratory experiments or through some simple dependence on height (de Roode et al. 2000) or on cloud radius (Simpson and Wiggert 1969).

Similarly, the formulation of a physically sound closure assumption to link convective activity (e.g., cloud-base mass flux) to the grid box mean model variables is far from simple. For deep convection, Grell (1993) showed that a CAPE closure works better than a moisture convergence closure. The latter type of closure is more appropriate in the case of shallow convection, which is mostly governed by boundary layer processes.

Furthermore, given the increasing horizontal resolutions (better than 10 km) used nowadays in mesoscale limited-area models and the simultaneously growing complexity of physical parameterizations, the usual underlying assumption that convective activity responds to large-scale forcings through a sequence of quasi equilibria (Arakawa and Schubert 1974) is no longer valid. At this stage, a prognostic treatment of convection should be considered (e.g., Gérard and Geleyn 2005) to properly account for the delayed response of convection, for the fluctuations of the large-scale forcings in time, and for the fact that the average lifetime of convective clouds (~1 h) becomes much longer than the model time step (~1 min).

The one-dimensional approach used to describe convection in current parameterizations is also questionable since real convective circulations are often three-dimensional. For instance, the strong updrafts that occur at the front of squall lines often tilt significantly toward the rear of the system, thereby supplying heat, moisture, and cloud condensate to its trailing stratiform precipitation region (e.g., Montmerle et al. 2000). In the same respect, observations have shown that the horizontal spreading of convective downdrafts close to the surface (gust fronts) can favor the development of new cells and thereby promote the horizontal propagation of mesoscale cloud systems. This level of detail is currently not available in operational forecast models, although some propagation of mesoscale cloud systems can be already simulated through the large-scale lifting that results from the strong horizontal temperature gradients associated with downdrafts. This also emphasizes the importance of including downdrafts in convective parameterization.

Obtaining a proper representation of the diurnal cycle and triggering of convection is also a crucial point. Parameterizations usually initiate deep convection as soon as CAPE or low-level moisture convergence is diagnosed in the model state, while in the real world this stage of convection is only reached when sufficient amounts of vertical instability and midlevel moisture become available. This observed buildup stage typically materializes through the presence of cumulus humilis and congestus clouds from morning to early afternoon. This explains why numerical models tend to predict both the triggering of convection and the time of maximum precipitation too early (by a couple of hours) compared to observations (Betts and Jakob 2002), especially over land areas where the diurnal cycle of
boundary layer properties and CAPE exhibit a large
amplitude. Attempts to overcome this problem in
three-dimensional models such as in Bechtold et al.
(2004) were of limited success. It might be worthwhile
to investigate whether a prognostic treatment of some
of the convective properties (e.g., the mass flux) can
improve the representation of the diurnal cycle.

3. Parameterizations of moist processes in
variational data assimilation

a. Generalities

To produce three-dimensional atmospheric states or
analyses that can be used to initialize forecasting mod-
els and that are optimally consistent with a huge set of
observations available from ground stations, ships,
buoys, aircraft, and satellites, most operational weather
forecasting services have adopted the variational data
assimilation approach (e.g., Lorenc 1986; Daley 1991),
although other methods such as nudging, diabatic ini-
tialization, or ensemble Kalman filtering are sometimes
used. To obtain a given analysis in the context of vari-
tional data assimilation, observations can be assumed
to be valid at a given time (3DVAR) or they can also be
assimilated at the individual time of each observation
(4DVAR; e.g., Courtier et al. 1994).

In practice, variational data assimilation aims at find-
ing the optimal initial model state that minimizes its
distance to a set of observations and to an a priori
background model state, in a least square sense. This
distance or cost function to be minimized is usually de-
defined as

\[ J(x) = \frac{1}{2} (x - x_0)^T B^{-1} (x - x_0) \]

\[ + \frac{1}{2} [H(x) - y_o]^T R^{-1} [H(x) - y_o]. \]  

(5)

where \( x \) denotes the model’s state vector, and su-
perscript \( b \) denotes the background model state (i.e.,
the initial model state to be corrected). Matrix \( B \) con-
tains the background error covariances of the model control
variables, \( y_o \) denotes the vector of available observa-
tions, and \( R \) is the observation error covariance matrix
that accounts for instrumental errors and for the simu-
lation and interpolation errors in operator \( H \). Here, \( H \)
is the usually nonlinear observation operator that per-
mits the conversion of the model variables into equiva-
lent observed quantities. For instance, when observa-
tions directly affected by cloud and/or precipitation
(e.g., brightness temperatures, radar reflectivities, ra-
diative fluxes) are to be assimilated, \( H \) has to include
some parameterizations of resolved and subgrid-scale
moist processes and possibly a radiative transfer model
as well. In 4DVAR, \( H \) also involves the integration of
the forecast model, the so-called trajectory, over the
assimilation window (usually 6 or 12 h). The result of
the minimization of Eq. (5) (i.e., the analysis) can there-
fore be seen as the combination of the background state
and the observations, weighted by the inverse of their
respective statistical errors. In practice the minimiza-
tion of the cost function involves the computation of its
gradient through

\[ \nabla_x J = B^{-1} (x - x_0) + H^T R^{-1} [H(x) - y_o]. \]  

(6)

In this equation, \( H^T \) is the adjoint of operator \( H \), which
includes the adjoint of a set of moist physical parameter-
izations that are usually simplified versions of the
ones used in the forecast model (see section 3b). More
details on the adjoint technique and its applications can
be found in Errico (1997). It is also essential to recall
that the two fundamental assumptions in the variational
assimilation approach are that 1) errors described by \( B \)
and \( R \) are normally distributed and 2) operator \( H \) is
linear or at least quasi-linear for perturbations with a
size typical of analysis increments (\( \approx 1 \) K for tem-
perature, \( 10^{-3} \) kg kg\(^{-1} \) for specific humidity, and 1 m s\(^{-1} \) for
wind). Only if both conditions are fulfilled, the result of
the assimilation is optimal and equal to the so-called
best linear unbiased estimator (BLUE; e.g., Talagrand
1997).

b. Current status

Since the introduction of variational data assimila-
tion in meteorology in the 1980s, the underlying as-
sumption that \( H \) is quasi-linear has been one of the
major obstacles to the inclusion of moist processes in
the adjoint calculations involved in the minimization of
Eq. (5). Indeed, moist processes such as condensation/
evaporation or autoconversion of cloud condensate
into precipitation are often nonlinear since they involve
on-off switches or threshold values such as saturation
or critical particle size. Convective parameterizations
can be even more problematic: their usually rather
noisy behavior and the disregard of regime changes can
lead to inaccurate tangent-linear (TL) approximation
and convergence problems during the 4DVAR minimi-
(zation (e.g., Zupanski 1993; Vukićević and Errico 1993;
Zou 1997; Vukićević and Bao 1998). In other respects
the significant computational overhead required to run
the adjoint version of moist physical parameterizations
had long remained inhibitory for operational purposes,
especially in 4DVAR.

However, with the permanent increase of computer
resources during the last decades, the computational limitation has become less of an issue. Besides, special efforts have been dedicated to the development of simplified yet rather realistic parameterizations of moist processes that are computationally cheap and easy to linearize and “adjoint,” and in which strong nonlinearities are smoothed out in various ways. Improved linearity can usually be achieved either by using smooth functions to describe each physical process or by artificially reducing or neglecting the perturbations of problematic quantities involved in the parameterization. Of course, it is absolutely essential to check that the TL version of the simplified schemes still provides a reasonable approximation to the finite difference of two nonlinear integrations with the full parameterizations (see example in the following).

Simplified linearized parameterizations of moist processes have been developed, for instance, by Janisková et al. (1999a) who derived a simplified mass-flux convection scheme inspired from Bougeault (1985), with a reformulation of detrainment/entrainment to provide smoother mass-flux profiles as well as a simple diagnostic parameterization of large-scale clouds and precipitation based on local moist adjustment.

Mahfouf (1999) proposed two simplified parameterizations meant for data assimilation. The first scheme describes deep convection using a mass-flux formulation, and perturbations of all convective cloud properties including the mass flux are neglected. Only perturbations of the model control variables and output tendencies are considered in the TL and adjoint versions. The second parameterization represents large-scale condensation processes by assuming a simple local moist adjustment and includes the melting of snow but neglects precipitation evaporation. These two parameterizations have been recently replaced in 4DVAR at ECMWF by the more detailed schemes of Tompkins and Janisková (2004) and Lopez and Moreau (2005; see below).

Lopez (2002) showed that the function used to describe the autoconversion of cloud condensate to precipitation had to be sufficiently smooth in order to improve the linearity of his large-scale condensation scheme in Météo-France’s Action de Recherche Petite Echelle Grande Echelle (ARPEGE) model. Fillion and Bélaire (2004) demonstrated how the linear behavior of the linearized version of the Kain-Fritsch (1990) convection scheme became questionable for input perturbations with sizes typical of analysis increments, which made its application difficult in variational data assimilation.

Tompkins and Janisková (2004) designed a simplified diagnostic large-scale cloud and precipitation parameterization that is consistent with Tiedtke’s (1993) original prognostic scheme used in ECMWF’s operational forecast model. A uniform PDF of total water is assumed so that cloud cover and cloud condensate are diagnosed from relative humidity and a novel treatment of precipitation evaporation is included. A reduction of cloud cover perturbations and a regularization of the autoconversion function had to be applied to make the scheme more linear.

Lopez and Moreau (2005) proposed a simplified mass-flux convection scheme that is based on Tiedtke’s original parameterization (1989) and in which perturbations of most convective quantities are retained, in opposition to Mahfouf (1999). In addition, all types of convection are now accounted for with a CAPE closure for deep and midlevel convection and a moisture convergence closure for shallow convection. They found that the perturbations of buoyancy and of the updraft initial vertical velocity had to be reduced and neglected, respectively, for the sake of linearity.

Finally, Mahfouf (2005) demonstrated that perturbations of detrainment and cloud properties had to be disregarded so as to make his modified Kuo-type convection scheme applicable in the Canadian Global Environmental Multiscale model.

To illustrate the impact of the inclusion of moist processes on the TL approximation after a 12-h T159 L60 integration with the ECMWF global model, Fig. 3 displays vertical cross sections of the zonal mean changes (relative to an adiabatic TL run) of

$$\epsilon = M(x + \delta x) - M(x) - M'(x)\delta x$$  \hspace{1cm} (7)

for temperature, when more and more detailed simplified physical parameterizations are gradually added to the TL model, $M'$. Formally, $\epsilon$ can be seen as the residual of the first-order Taylor development of the nonlinear version of the ECMWF model with full physics, $M$, which includes Tiedtke’s (1989, 1993) parameterizations of large-scale and convective precipitation. The global mean change in $\epsilon$ (expressed in percent) relative to an adiabatic TL run is given at the top of each panel. Figure 3a shows that the inclusion of dry physics (vertical diffusion, gravity wave drag, and radiation) reduces $\epsilon$ by 7.09%, mainly below model level 40 ($\approx$3 km height). Figure 3b indicates that adding Mahfouf’s (1999) simplified moist physics leads to an even larger reduction of the error (13.65%), and that the impact now spreads over the entire atmosphere. Error changes in the tropical stratosphere (above model level 25) are due to the feedback of clouds on radiation. Figure 3c exhibits a further reduction of $\epsilon$ (16.83%) when the more detailed simplified moist physics of Tompkins and Janisková (2004) and Lopez and Moreau (2005) is sub-
FIG. 3. Impact of the inclusion of simplified physics on TL approximation after a 12-h T159 L60 integration with the ECMWF global model: vertical cross sections of the zonal mean changes (relative to an adiabatic TL run) of the TL approximation error $\epsilon$ (see text) for temperature (in K) (a) when dry physics is included, (b) when Mahfouf’s (1999) simplified moist physics is also added, and (c) when the new simplified moist physics of Tompkins and Janisková (2004) and Lopez and Moreau (2005) is substituted for Mahfouf’s (1999). The global mean change in $\epsilon$ relative to the adiabatic TL run is given at the top of each panel (in %). x axis: latitude; y axis: model vertical levels from 1 to 60 (surface).
stituted for Mahfouf's (1999). Such improvement of the TL approximation associated with the inclusion of moist processes can only be beneficial to the overall behavior of a 4DVAR assimilation system, as suggested in Vukičević and Bao (1998) for instance.

c. Main issues in data assimilation

From all the studies mentioned above, it becomes clear that one of the major difficulties when designing a new parameterization suitable for variational data assimilation is to find the optimal trade-off between physical realism, linearity, and computational affordability. The two major assumptions of 1) normal distributions of errors for background (matrix B) and observations (matrix R), and 2) linearity of H are somewhat difficult to satisfy in the context of clouds and precipitation assimilation. If either of these assumptions turns out not to be satisfied, then the analysis obtained from the minimization of Eq. (5) is not equal to the expected BLUE, and the resulting analysis error distribution is not Gaussian (Errico et al. 2000). Also, Lorenc (2003) showed that nonlinearities in the observation operator H can lead to non-Gaussian error distributions in R. Besides the important question of linearity mentioned in section 3b, other issues that will be addressed here relate to the specification of error statistics, to the inclusion of moist processes, and to temporal and spatial resolutions.

1) MODEL AND OBSERVATION ERROR STATISTICS

It has already been emphasized that observation and model error statistics play a fundamental role in the minimization of Eq. (5). Therefore the error statistics of the observations to be assimilated ought to be known as accurately as possible. Unfortunately, the errors of direct observations (i.e., instrument and representativeness errors) or indirect retrievals of clouds and precipitation properties are very often so poorly documented that modelers have to specify them in a crude conservative way. However, there is hope that the recent interest taken in new types of observations (e.g., reflectivities or derived products) will incite observation providers to tackle this issue. In other respects, in the case of rain observations, imperfect ways of dealing with non-Gaussian errors such as suitable changes of variable (e.g., logarithm, square root) have been proposed in the literature (e.g., Hou et al. 2004; Lopez et al. 2006).

Another shortcoming in the current variational assimilation of clouds or precipitation observations is the crude representation or even disregard of forward modeling errors in matrix R (i.e., the errors in operator H). Therefore modelers need to properly document the error statistics of simulated microwave brightness temperatures, cloud and precipitation fields, and even radar reflectivities. Furthermore, the contribution of moist physics parameterizations to the total error in H is expected to be much larger than that of the errors in radiative transfer calculations. Hollingsworth and Lönberg (1986) proposed a method able to disentangle observation and model errors. Spatial covariances of model–observation departures are computed for a selected forecast time. If observations can be assumed to be spatially uncorrelated then it is possible to get an estimation of model error spatial covariances, model error standard deviation, $\sigma_b$, as well as observation error standard deviation, $\sigma_o$, as illustrated in Fig. 4. This method was successfully applied to microwave brightness temperatures in Bauer et al. (2006a). An example of its application to the computation of error statistics for the daily accumulated surface precipitation simulated with the ECMWF model compared to rain gauge observations over Europe from the European Land Data Assimilation System (ELDAS) Project (Rubel et al. 2004) is presented here for January 2000. Maps of horizontal correlations of forecasted precipitation errors for distances ranging between 50 and 200 km are displayed in Fig. 5. Figure 6 shows mean horizontal covariances of the precipitation error as a function of distance as well as the PDF of model–observed precipitation departures. Both figures indicate a fast drop of
the spatial correlations of model precipitation errors with distance and the absence of any strong bias in the daily amounts of precipitation on average over Europe. The same kind of plots for the month of July 2000 (not shown) led to similar conclusions, with the difference that the span of horizontal correlations is even shorter, that the PDF of model–observation departures is broader, and that the mean bias is slightly larger, as expected in the presence of convective activity. This kind of information could be valuable to improve the definition of matrix $R$. An alternative solution might be found in a weak constraint formulation of 4DVAR that involves the inclusion of systematic and random model errors in the assimilation control vector (e.g., Zupanski 1997; Trémolet 2005), but this requires further investigation.

Fig. 5. Maps of spatial correlations of ECMWF model–ELDAS daily precipitation departures in January 2000 and for a horizontal distance of (a) 50, (b) 100, (c) 150, and (d) 200 km. Solid (dotted) contour lines are used for positive (negative) correlations.
Although it is reasonable to assume that background error statistics ($B$) for the basic control variables (namely temperature, specific humidity, wind, and surface pressure) have Gaussian distributions, more dependence on the meteorological situation at each model grid point (local flow, presence of clouds, or precipitation) ought to be introduced in current data assimilation systems. This seems particularly desirable in the vicinity of frontal areas where spatial gradients are pronounced and where horizontal error correlations or structure functions are expected to be highly anisotropic. This kind of refinement is more difficult to implement when $B$ is defined in spectral space, as it is the case in many global assimilation systems, although a wavelet formulation of the $J_b$ term (Fisher 2004; Deckmyn and Berre 2005) or a suitable coordinate change (Desroziers 1997) can provide some degree of flow dependence.

2) INCLUSION OF MOIST PROCESSES

Which moist physical processes must be included in the data assimilation process depends on the spatial and temporal resolutions at which the problem is supposed to be solved and on the complexity of the observations to be handled. Indeed it can be expected that the finer the resolution or the more detailed the observation is (e.g., a reflectivity profile versus a 2-m temperature measurement), the more detailed the representation of clouds and precipitation processes needs to be. But at the same time, whether the linearity assumption involved in variational data assimilation still holds at resolutions of a few kilometers is most uncertain, especially in situations of convective activity (Walser et al. 2004). Shortening the assimilation window to less than an hour can help to limit the effects of strong nonlinearities, as for instance in the experimental 10-min 4DVAR assimilations of radar reflectivities and radial velocities performed by Sun (2005) with the Variational Doppler Radar Assimilation System (VDRAS; Sun and Crook 2001).

The choice of the prognostic variables to be included in the trajectory calculations and in the minimization is still an open question, especially when observations affected by clouds and/or precipitation are to be assimilated. An illustration of the potential problems arising from the diagnostic treatment of precipitation in trajectory calculations can be found in Lopez et al. (2006), who assimilated ARM cloud-radar reflectivity profiles within a 2DVAR method based on the single-column version of the ECMWF model. It was realized that the diagnostic treatment of snow in the trajectory could increase the risks of failure of the 2DVAR minimization. This issue will be presented here in the idealized case of a midlatitude winter stratiform ice cloud located, say, between 6 and 8 km as shown in Fig. 7. The cloud location is supposed to be correct in the initial model state but cloud ice content is assumed to be underestimated. Simulated reflectivities are therefore too weak compared to those obtained from the cloud radar. After the first iteration of the minimization, the assimilation process increases the model reflectivities through a cooling and above all a moistening of the initial model state inside the cloud (not shown), thereby producing

![Fig. 6. Error statistics of ECMWF model–ELDAS daily precipitation departures over Europe in January 2000: (a) mean spatial covariances (mm day$^{-1}$) as a function of horizontal distance (km) and (b) PDF of departures (mean bias indicated in the title).](image)
more cloud ice. Part of the latter is spontaneously converted by the microphysical parameterization into snow, which is in turn assumed to fall and reach the ground within a single time step (say $\Delta t = 30$ min here). The 2DVAR increments therefore lead not only to the desired increase of in-cloud reflectivity but at the same time to an unrealistic subcloud enhancement of reflectivity caused by the falling of snow. This increase of model reflectivities below cloud base depends on the method employed to convert model output snow fluxes ($F_S$ in kg m$^{-2}$ s$^{-1}$) into snow contents ($q_s$ in kg m$^{-3}$) inside each model layer. Commonly, it can be written $q_s = F_S / V_F$, which supposes a steady snow sedimentation at speed $V_F$ throughout the model time step. To use $q_s = \max[(\partial F_S / \partial z)\Delta t, 0]$ could also be imagined, but this is equivalent to making the drastic assumption that $V_F = 0$. The truth lies in between those two extreme approaches. The assumption of fast falling implied by a diagnostic treatment of snow is not justified because observed values of $V_F$ are usually well below 0.5 m s$^{-1}$, which means that within $\Delta t = 30$ min the snow should not descend by more than 1 km. With a prognostic treatment in the trajectory, the falling of snow would be represented with higher accuracy, and besides snow contents would be directly available to calculate reflectivities (no need for questionable assumptions to estimate $q_s$ from $F_S$).

It should be stressed that an inadequacy of the microphysical parameterization could also partly account for an excessive conversion of cloud ice into snow. More generally, the current distinction between cloud ice ($V_F = 0$) and snow ($V_F > 0$) in most NWPMs is somewhat very crude and a prognostic treatment of the ice particle size distribution (e.g., its lower-order moments or its representation in multiple bins) might improve our description of microphysical processes in general and make the assimilation of cloud and precipitation observations easier.

Adding cloud properties (mixing ratios, fractional coverage) or even precipitation to the control vector $x$ of the minimization itself (including TL and adjoint calculations) is not straightforward because it requires the specification of corresponding modeling error statistics contained in matrix $\mathbf{B}$. These fields are highly heterogeneous in space and time and their error spatial correlations are often anisotropic and flow-dependent, which makes their estimation problematic. It is also very likely that the distribution of the background error statistics for such fields is far from being Gaussian. Few attempts have been actually made to assess the $\mathbf{B}$ matrix for clouds and precipitation variables (e.g., Moreau et al. 2003; Lopez 2001) but large uncertainties remain. It should be noted that the inclusion of prognostic variables for clouds or precipitation in adjoint calculations also permits the determination of the time-evolved sensitivities of any aspect of the forecast to these variables (e.g., Lopez 2003).

An attractive alternative solution could be to substitute a total water prognostic variable for the usual water vapor variable in the assimilation control vector, which would provide a smoother field with nearly Gaussian error distribution regardless of the saturation level. The only requirement in this case is to find a proper way for splitting the total water variable into its water vapor and cloud condensate components in the physics calculations, which can be achieved by specifying some suitable PDF of total water.

The joint inclusion of parameterizations of resolved and subgrid-scale moist processes in the data assimilation procedure is also problematic because nothing guarantees that the right scheme will be activated during the minimization. For example, in their 1DVAR rain assimilation experiments, Fillion and Mahfouf (2000) and Moreau et al. (2004) showed that their large-scale condensation parameterization was much more active than their convective scheme during the minimization, every time the initial state was close to saturation at some vertical levels. In this case precipitation can indeed be readily modified through large-scale condensation by local adjustments of temperature and moisture, while nonlocal adjustments of these variables to alter vertical instability are required in the convection scheme to obtain the same precipitation change. In other words, a large-scale condensation scheme is more sensitive than a convection scheme in nearly saturated conditions. Unsuccessful attempts were made to solve...
this problem via the addition of a penalty term in the cost function of Eq. (5), either to control supersaturation or to include additional information about precipitation nature from observed $T$ and $q$ profiles (Fillion and Mahfouf 2000).

Inside convective regions, using a different treatment of deep and shallow convection (e.g., closure assumption, entrainment/detrainment rates) may sometimes cause a failure of the minimization if the model state switches from one convective regime to another during the iterative solving. But at the same time, in my own experience, distinguishing between deep and shallow convection in the linearized parameterization can improve the validity of the TL approximation.

3) SPATIAL AND TEMPORAL RESOLUTIONS

The assimilation of clouds or precipitation can also be affected by large mismatches between model and observations in terms of geographical location, timing and nature (convective versus stratiform) of the cloud systems. Such problems are more likely to happen when the resolution of the data assimilation system is increased as well as when the available observations become more frequent in time. An illustration of the latter case is given in Lopez et al. (2006), who found that their 2DVAR assimilation of half-hourly ARM cloud-radar reflectivities over a 12-h window could not converge because the model–observation departures exhibited too much variability on a 30-min time scale. The diagnostic nature of the convection schemes currently used in most NWPMs largely explains the spurious variations that are sometimes seen in model outputs from one time step to the next. The assimilation of instantaneous observations in clouds or precipitation from satellites or ground-based radars can be particularly affected by this problem. One possible solution is to reduce the amount of information coming from the observations either by thinning the data in time or by assimilating only statistical moments of the observed time series as in Lopez et al. (2006). It can also be beneficial to accumulate or average the model outputs over several time steps prior to the assimilation as for example in the “1 + 1DVAR” assimilation of rain observations developed by Hou et al. (2000), or even to combine both approaches.

4DVAR assimilation can also suffer from mismatching locations of cloud systems at the different resolutions used for integrating the model trajectory and for solving the minimization. For instance, the current ECMWF operational 4DVAR system (Courtier et al. 1994) involves four different horizontal resolutions: T799 (~25 km) for the main forecast model trajectory, T95 (~210 km) for the first minimization trajectory (without moist physics), and T159 (~130 km) and T255 (~80 km) for the second and third minimization trajectories (with moist physics). In the past, the lower-resolution trajectories of the minimizations were entirely recalculated over the entire 12-h assimilation window. In this case, nothing could guarantee that the T159 cloud systems would be collocated with the T511 ones, which was particularly undesirable when trying to assimilate clouds or precipitation observations. To avoid this problem, the low-resolution trajectory is now interpolated from the high-resolution one.

4. Summary

Clouds and precipitation are major contributors to the earth’s radiative, thermal, and hydrological budgets and important rulers of human activities. Therefore parameterizations of both resolved and subgrid-scale atmospheric moist physical processes have become essential components of current NWPMs, GCMs, and data assimilation systems. Although significant improvements have been achieved in the forecasting of clouds and precipitation over the past 40 yr, some uncertainties still arise from the discrete nature of our representation of the atmosphere, from limitations in computer resources and from our imperfect theoretical and observational knowledge of moist processes. It can reasonably be claimed that large-scale clouds and precipitation are currently better simulated than convective ones because the latter involve parameterizations of phenomena that are highly variable in space and time. Even though the description of slow ice phase processes (e.g., nucleation, ice supersaturation) and subgrid-scale variability of total water in stratiform clouds does require further improvements (prognostic ice variable and total water PDF, more detailed microphysics), most of our efforts should aim at getting a better understanding and ability to simulate convective activity. The simplifying assumptions on which current convection parameterizations are based (namely bulk representation of cloud ensemble, quasi-equilibrium closure) should be revised. More statistical information about subgrid-scale variability could be inferred from ground-based or spaceborne radar observations (TRMM, CloudSat, and future GPM) and included in new parameterizations. A prognostic treatment of major convective properties such as mass flux and cloud condensate amount might improve the simulation of the diurnal cycle and in particular of the gradual buildup of deep convection over land. It is also essential to improve our knowledge and thus our representation of entrainment/detrainment in convective clouds, of the coupling between convective and stratiform clouds, and
of the interaction between precipitation and land/sea surface processes (heat fluxes, orography, triggering of convection).

In the context of variational data assimilation, the two major questions are: 1) whether it is useful to include moist processes in the assimilation process itself (regardless of the type of observations to be assimilated), and 2) whether it is beneficial to assimilate observations that are directly affected by clouds and/or precipitation. Theoretically, the inclusion of large-scale condensation and convective parameterizations in the 4DVAR minimization is expected to be beneficial because it improves the overall validity of the underlying TL approximation and should also provide better-balanced analysis increments. However, the basic assumption of quasi linearity and the computational requirements of variational data assimilation imply that the inclusion of moist physical parameterizations in the minimization of the cost function can only be achieved through some substantial simplifications and partial regularizations of the code. All sources of spurious large growth of perturbations in time need to be identified, by means of singular vectors computations for instance, and to be removed for any new parameterization to become usable in data assimilation. The development of (moist) physical parameterizations for the purpose of data assimilation therefore consists of finding the best possible compromise between realism, computational cost, linearity, and approximation of the parameterizations used in the nonlinear forecast model. This means that some degree of inconsistency always exists between the physical parameterizations used in the nonlinear forecast model and those used in the linearized calculations of the minimization. Trying to reduce this inconsistency should be regarded as a priority among the data assimilation community, and we can only fear that the unrelenting trend toward higher and higher model resolution will render this problem even more acute, especially as regards the treatment of convection. The implementation of a more statistically based (and thus smoother) treatment of convection might alleviate the problem of high temporal and spatial variability associated to this process, but this would need to be tested. More efforts should also be dedicated to constrain the stratiform/convective partitioning of analysis increments more efficiently during the iterative minimization of the cost function.

In spite of the latter limitations, some studies have shown that the inclusion of properly regularized simplified moist physics in 4DVAR can have a positive impact on analyses and forecasts (Tsuyuki 1997; Janišková et al. 1999b). Other works have demonstrated the feasibility of the 4DVAR assimilation of observations of rain rates and microwave brightness temperatures in rainy regions, which can result in some improvements of analyses and forecasts (Marécal and Mahfouf 2002; Hou et al. 2001; Hou et al. 2004). Such observations are already being operationally assimilated by some weather services with either a slightly positive or a neutral impact on analyses and forecast scores (e.g., Bauer et al. 2006b). Although nudging and diabatic initialization techniques are sometimes also used operationally to extract information from observations of clouds and precipitation, 4DVAR offers more consistency through the use of the dynamics throughout the assimilation window combined with the use of each observation at their appropriate time.

These are encouraging advances but, in addition to the aforementioned question of linearity, several other important issues will require further investigation in the near future. Whether to treat clouds and precipitation as control variables in the trajectory or even in the minimization itself is still unclear. Substituting total water for specific humidity in the minimization control vector might make the assimilation smoother since it would eliminate the discontinuity around saturation and provide error distributions closer to normality. This needs to be tested. Adding a prognostic variable for snow (slow fall speeds) in the trajectory calculations (but not necessarily in the minimization) might be helpful when observed vertical profiles of precipitation or reflectivities are to be assimilated.

Our knowledge and specification of background error statistics (B) for standard control variables (T, q, wind) inside and around cloudy and rainy areas clearly need to be improved, especially anisotropies of model error correlations that result from flow dependence. Similarly, more studies should aim at documenting the error statistics of observation operators (R) that simulate precipitation, brightness temperatures, radar reflectivities, or lidar backscattering. In particular these statistics should include the errors of moist physics parameterizations. It is essential to understand and maybe alter the way the additional measurements of clouds or precipitation are blended with clear-sky and conventional observations (especially moisture data) via structure functions during the assimilation process. It is also important to assess the degree of consistency between these various sources of information. The potential of a weak constraint formulation of the variational data assimilation problem to represent model errors better should be investigated further.

Acknowledgments. I am very grateful to Drs. Peter Bechtold and Adrian Tompkins from ECMWF for their help and discussions about the current issues in
the modeling of convection and large-scale condensation. Drs. Marta Janisková and Peter Bauer from ECMWF should be thanked for their early reviewing of this paper. The two anonymous reviewers should also be acknowledged for their constructive comments about this paper.

REFERENCES


Gregory, D., and P. R. Rowntree, 1990: A mass flux convection


