

# A generalized approach to parameterizing convection combining ensemble and data assimilation techniques

Georg A. Grell<sup>1</sup> and Dezső Dévényi<sup>1</sup>

Received 11 April 2002; revised 3 June 2002; accepted 12 June 2002; published 25 July 2002.

[1] A new convective parameterization is introduced that can make use of a large variety of assumptions previously introduced in earlier formulations. The assumptions are chosen so that they will generate a large spread in the solution. We then show two methods in which ensemble and data assimilation techniques may be used to find the best value to feed back to the larger scale model. First, we can use simple statistical methods to find the most probable solution. Second, the ensemble probability density function can be considered as an appropriate “prior” (a priori density) for Bayesian data assimilation. Using this “prior”, and information about observation likelihood, measured meteorological or climatological data can be directly assimilated into model fields. Given proper observations, the application of this technique is not restricted to convective parameterizations, but may be applied to other parameterizations as well. *INDEX TERMS:* 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 3314 Meteorology and Atmospheric Dynamics: Convective processes; 3329 Meteorology and Atmospheric Dynamics: Mesoscale meteorology; 3309 Meteorology and Atmospheric Dynamics: Climatology (1620)

## 1. Introduction

[2] Properly parameterizing the effects of convection is still a challenging problem for numerical weather prediction (NWP). There are many different parameterizations for deep and shallow convection that exploit the current understanding of the complicated physics and dynamics of convective clouds to express the interaction between the larger scale flow and the convective clouds in simple “parameterized” terms. These parameterizations often differ fundamentally in closure assumptions and parameters used to solve the interaction problem, leading to a large spread and uncertainty in possible solutions. In past studies, these uncertainties have led to many discussions regarding which assumptions are the proper ones to use under what conditions.

[3] In this paper we offer a generalized approach to make use of these uncertainties by combining ensemble and data assimilation techniques. First a parameterization is developed that can employ a large ensemble of closure assumptions and parameters. These closures and parameters are taken from cumulus parameterizations which are currently used in various three-dimensional models. This is described

in section 2. Statistical techniques may then be applied to find the proper feedback to the three-dimensional model. Such techniques have already been successfully applied and verified in several operational centers. We discuss the statistical methods that we use for our application in section 3. We offer an additional solution by combining data assimilation with an ensemble-type parameterization. This is discussed in section 4. Finally, conclusions are provided in section 5.

## 2. The Parameterization Framework

[4] The parameterization framework is a simple scheme that is based on a convective parameterization developed by Grell [1993, G1] and discussed in more detail by Grell *et al.* [1994, G2]. For our application, the simple scheme was expanded to allow for a series of different assumptions that are commonly used in convective parameterizations and that have proven to lead to large sensitivity in model simulations. In addition, values for the assumed parameters are perturbed (see section 3). Because of the limited scope of this paper, we refer the reader to G1 and G2 for details, and this paper only discusses the most important aspects that we use in our new ensemble approach. Following G1, we will use the same terminology of dynamic control (the modulation of the convection by the environment), feedback (modulation of the environment by the convection), and static control (the cloud model that is used to determine cloud properties).

### 2.1. Static Control and Feedback

[5] Many cumulus parameterizations use some type of simplified cloud model to calculate cloud properties. Despite the simplicity of these cloud models, assumptions and parameters chosen by these 1-d cloud models can lead to large sensitivities within the framework of a cumulus parameterization. Here we choose to implement and test assumptions that directly influence the vertical redistribution of heat and moisture or the rainfall rate. Following G1, we introduce the symbol  $\lambda$  to denote an ensemble type, and rewrite the entrainment hypothesis as

$$\mu_{ue}(z, \lambda) - \mu_{ud}(z, \lambda) = \frac{1}{m_u(z, \lambda)} \frac{\partial m_u(z, \lambda)}{\partial z}, \quad (1)$$

where  $\mu_{ue}$  is the gross fractional entrainment rate,  $\mu_{ud}$  is the gross fractional detrainment rate (subscript  $u$  designates an updraft property), and  $m$  is the mass flux. Following G1, each subensemble is normalized by the mass flux at cloud base ( $m_b$ ) to give

$$m_u(z, \lambda) = m_b(\lambda) \eta_u(z, \lambda). \quad (2)$$

<sup>1</sup>Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado/NOAA Forecast Systems Laboratory, Boulder, Colorado, USA.

**Table 1.** Overview of Ensembles Used in this Study

Name	Part of Parameterization	Varied Parameter	Number of Variations
Edyn1	dynamic control	larger-scale forcing tendencies	3
Edyn2	dynamic control	$A$	4
Edyn3	dynamic control	$dtc$	3
Edyn4	dynamic control	$b$	3
Edyn5	dynamic control	$l_r$	3
Ef1	static control/feedback	$\beta$	6
Ef2	static control/feedback	$\nu_{ud}(z, \lambda)$	4
Ef3	static control/feedback	$\nu_{ue}(z, \lambda)$	6
Ef4	static control/feedback	$\nu_{dd}(z, \lambda)$	6

The 16 *Edyn* closures are allowed to interact with any of the other closures, giving a total of 13824 ensemble members ( $16 \times 6 \times 4 \times 6 \times 6$ ).

where  $\eta_u$  is the normalized mass flux. Given initial conditions and closures for entrainment, as well as detrainment rates, Equations (1) and (2) can be used together with the steady state plume equation (see G1) to estimate model-cloud properties such as normalized mass flux, normalized condensation and evaporation profiles, moist static energy, and liquid water content for each ensemble member. Following G1, the equations for the downdraft mass budget would be analogous to (1) and (2). The choice of entrainment and detrainment rates characterizes subensembles *Ef2*, *Ef3*, and *Ef4* (see Table 1) in this study.

[6] Results have also proven very sensitive to the dependency of the downdraft mass flux on the updraft mass flux. Following G1, by assuming that part of the total condensate is evaporated in the downdraft, this dependency may be expressed as

$$m_0(\lambda) = \beta(\lambda) \frac{I_1(\lambda)}{I_2(\lambda)} m_b(\lambda). \quad (3)$$

Here  $(1 - \beta(\lambda))$  would be the precipitation efficiency (usually a function of wind shear and subcloud humidity),  $m_0$  the downdraft originating mass flux, and  $I_1(\lambda)$  and  $I_2(\lambda)$  are the normalized condensate and evaporate of the updraft and downdraft of ensemble  $\lambda$ . This closure is usually very sensitive to the parameterization of the precipitation efficiency, and has been chosen for subensemble *Ef1* (Table 1).

## 2.2. Dynamic Control

[7] A multitude of different closures exist in the literature to determine the amount and location of convection. Many of these closures may be used in this scheme to determine  $m_b$ , the cloud base mass flux.

[8] The first type of closure is based on some type of stability equilibrium. We chose to use the definition of the cloud work function  $A$ , an integral measure of the buoyancy force associated with a cloud of subensemble  $\lambda$ , which was first given by *Arakawa and Schubert* [1974, AS]. In the original implementation of the Grell scheme (G1), an equilibrium between the larger scale forcing and the cloud response was assumed

$$-\frac{A'(\lambda) - A(\lambda)}{dt} = \frac{A''(\lambda) - A(\lambda)}{m'_b(\lambda)dt} m_b(\lambda), \quad (4)$$

where  $A'(\lambda)$  is the cloud work function that was calculated using thermodynamic fields that were modified by forcing terms, and  $A''$  is the cloud work function that was calculated using thermodynamic fields that were modified by a cloud with arbitrary unit mass  $m'_b(\lambda)dt$ . Equation (4) can be easily solved for  $m_b$ . In G1,  $A$  is calculated locally (subensemble *Edyn1*). To more closely follow AS, a climatological value for  $A$  can be used (subensemble *Edyn2*).

[9] In a third implementation (subensemble *Edyn3*), to simulate a closure in which the stability is simply removed by the convection (as assumed in similar form by *Kain and Fritsch* [1992]), we simply assume

$$-\frac{A(\lambda)}{(dt)_c} = \frac{A''(\lambda) - A(\lambda)}{m'_b(\lambda)dt} m_b(\lambda), \quad (5)$$

which has the effect of making  $m_b(\lambda)$  strong enough to remove the available instability within the specified time period  $(dt)_c$ . Naturally (5) is sensitive to the choice of the parameter  $dtc$ .

[10] Another group of widely used closure assumptions is based on moisture convergence (*Kuo* [1974], *Molinari* [1982], *Krishnamurti et al.* [1983], to name a few). While there are many different choices, here we chose an assumption first introduced by *Krishnamurti et al.* [1983], where the total rainfall  $R$  is assumed to be proportional to the integrated vertical advection of moisture  $M_{rv}$  using

$$R = M_{rv}(1 + f_{emp})(1 - b). \quad (6)$$

Here  $b$  is the *Kuo* moistening parameter, and  $f_{emp}$  is an empirical constant. In addition, it can be shown (see G1) that the subensemble rainfall is defined as

$$R(\lambda) \equiv I_1(\lambda)(1 - \beta)m_b(\lambda). \quad (7)$$

Equations (6) and (7) can be used to calculate  $m_b$  in terms of  $M_{rv}$ . This closure is used for subensemble *Edyn4*. Any other moisture convergence closure may be employed by redefining  $M_{rv}$ .

[11] A further dynamic closure that is easily implemented was first introduced by *Brown* [1979], who assumes that the cloud mass flux at the cloud base is proportional to the environmental mass flux  $\tilde{M}$  at some lower tropospheric level  $l_r$ . This level may be taken as the top of the PBL height, or the level of the updraft originating air. *Brown's* closure was modified by *Frank and Cohen* [1987], by assuming

$$m_b(\lambda) = m_u(l_t, \lambda) = \tilde{M}(l_t) - m_d(l_t, t - \Delta t). \quad (8)$$

Here  $m_d(l_t, t - \Delta t)$  is the downdraft mass flux at the previous time step. This closure simulates a time lag between updraft and downdraft, envisioning the downdraft of a thunderstorm forcing another updraft at a later time. This closure builds the foundation for subensemble *Edyn5*.

## 3. Ensemble Statistics

[12] Table 1 summarizes the set of ensembles used in this study. For subensembles *Edyn1*, *Edyn3*, *Edyn4*, *Ef1*, *Ef2*,

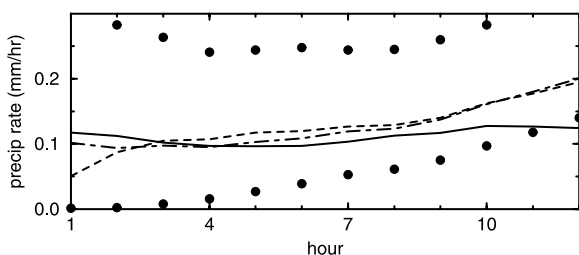
$Ef3$ , and  $Ef4$  parameters are perturbed within a range (uniform distribution) bounded by different values. For  $Edyn2$  the spread of climatological values of  $A$  has been chosen based on work from Lord and Arakawa [1980]. For  $Edyn5$ ,  $l$ , was taken as either the level of free convection, the level of updraft originating air, or the level with the maximum upward vertical velocity below the level of free convection. In current implementations of versions of this scheme [Grell and Dévényi, 2001], we use the ensemble mean of  $m_b$  at each time step and grid point to determine the feedback to the 3-d model. However, here our goal is to find a method to feed back the “best” values, where “best” is defined in the context below. The “best” values do not necessarily bear any relationship to this mean.

### 3.1. Basic Statistics

[13] As a first step, each ensemble and subensemble was submitted to detailed statistical analysis following a strategy close to one given in Stephenson and Doblas-Reyes [2000]. This approach is somewhat similar to ones applied in large scale weather forecasting, unifying forecasts from different ensembles, different forecast models, and even from different weather forecasting centers (see Ebert, [2001] and references therein).

[14] The most basic statistics we compute are mean (average), standard deviation, skewness, and flatness (kurtosis). These estimations were performed at each time step and at each gridpoint individually for individual subensembles and also for unified ensembles.

[15] In order to illustrate the simplest application of this scheme, we collected statistics from two real time MM5 experiments over a one month period. The model was run for a forecast length of 12 hour, twice a day, during August of 2001, using a horizontal resolution of 27 km (60 runs each). The domain with size of about 3000 km  $\times$  3600 km was centered over the central/eastern USA. For the first experiment (R1) to preserve the full spread with respect to the rainfall rates but reduce the computational costs, we limited the feedback ensemble size (using only 3 variations of  $Ef1$  and 3 variations of  $Ef2$ ), but left the dynamic closure size unchanged. For the second experiment (R2), the number of ensembles was further reduced by keeping only  $Edyn1$  (as used in G2). Results are displayed in Figure 1. It can be seen that the use of ensembles improves the



**Figure 1.** Domain averaged precipitation rates from MM5 real time runs. The model was run twice daily for 12-hour forecasts during August of 2001. Displayed are results from experiments R1 (dotted) and R2 (dashed), as well as observed precipitation rates (solid). Bracketing maximum and minimum precipitation rates from R1 are represented by the filled circles.

**Table 2.** Correlations Between Main Groups of Closures

Closure Group	Edyn1	Edyn3	Edyn4	Edyn5
Edyn1	1.00	0.85	0.88	0.19
Edyn3	0.85	1.00	0.72	0.26
Edyn4	0.88	0.72	1.00	0.25
Edyn5	0.19	0.26	0.25	1.00

domain averaged precipitation comparison, even in this simple application. Figure 1 also shows the bracketing maximum and minimum values of the precipitation rates, indicating that more improvement may be possible with an appropriately trained scheme. This will be discussed in the following sections.

### 3.2. Correlation Between Subensembles

[16] An important issue regards how much information is contributed from different subensembles to the unified ensemble. In an ideal case all subensembles are statistically independent, which maximizes the contribution to a unified ensemble. Because subensembles are constructed under similar but not identical physical hypotheses, complete independence cannot be expected. As an illustration of a method that tests the degree of inter-dependence of the subensembles, we generated correlations among the four main groups of closures for one arbitrarily selected convectively active grid point. As expected, Table 2 indicates various degrees of independence among the subensembles. These correlations are driven by the character of the convection and are therefore a function of gridpoint and time. On average we expect to maintain an appropriate estimation of spread.

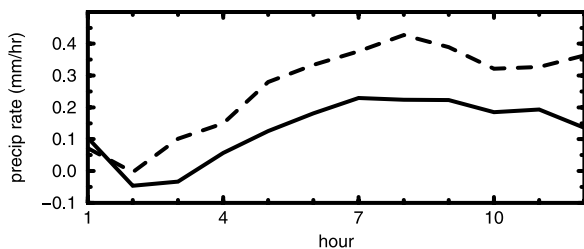
[17] Using inter-subensemble correlations, a statistically optimal mixture of subensembles may be derived, trained on observational data. One simple and efficient way to do this may be the application of linear regression techniques as was done by Krishnamurti *et al.* [1999]. This is also an option in our scheme and will be explored for global and regional climate modeling applications as well as for weather forecasting. However, a regression trained on a climatological data set may be less effective compared to some local methods.

### 3.3. Probability Density Estimation

[18] In order to visualize ensemble probability distribution functions (PDFs) for operational weather forecasters or data assimilation studies, appropriate probability density estimation methods should be employed. For our purposes we found the Epanechnikov kernel method from Härdle [1990] satisfactory. The PDFs estimated with this approach may then be used in the data assimilation technique described in the next section.

## 4. Data Assimilation

[19] The large size of the cumulus ensembles (see Table 1) and the application of different controls and closures provide a unique opportunity for assimilating data into model fields where and when corresponding measurement data are available. To realize this opportunity we should go beyond the standard methods of data assimilation. In our case of highly nonlinear systems of convection a full description of PDFs is required and a general Bayesian framework should be employed. We formulate our data assimilation method in the Bayesian framework of conditional probability distribu-



**Figure 2.** Domain averaged bias from experiment R1, averaged over all 60 runs wherever observed and forecasted precipitation was non zero (dashed line). Shown is also the bias (solid line) from the same experiment after diagnostically applying the data assimilation method.

tions (Miller *et al.* [1999]). According to the Bayes theorem we can write the *posterior* density  $f(x|y)$  as

$$f(x|y) = \frac{f(y|x)f(x)}{\int f(y|x)f(x)dx}, \quad (9)$$

where  $f(x)$  is the *prior* density (PDFs described in the previous section) deduced from the actual ensemble,  $y$  is an observation, and  $f(y|x)$  is the *likelihood*. We suppose the observation process results at a given time and location in a value  $y_k = h(x) + \sigma\epsilon$ , where  $h$  is the observation operator (could contain interpolation and physical processes) and  $\epsilon$  is Gaussian white noise with  $\sigma$  standard deviation. If we accept that the observation noise is Gaussian, we can compute the likelihood as

$$f(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-h(x))^2}{2\sigma^2}\right). \quad (10)$$

This way we have all the ingredients (prior from the ensembles, likelihood by Equation (10)) to apply Bayes Theorem, and can compute the posterior density using Equation (9). Mean or median of posterior distribution could be applied as feedback to model fields.

[20] Figure 2 presents an example of how much improvement may be possible when using this method. Shown is the domain averaged bias for experiment R1, averaged over all 60 runs wherever observed and forecasted precipitation was non zero (dashed line). Shown is also the bias (solid line) from the same experiment after diagnostically applying the data assimilation method. The bias is generally reduced substantially, except during the first two hours of the simulation.

## 5. Conclusions

[21] We have developed a new convective parameterization framework that is able to use a large ensemble of assumptions and can make use of ensemble as well as data assimilation techniques to determine the optimal value for feedback to three-dimensional models. The model output fields that were generated by the described statistical methods may also aid forecasters or scientists in diagnosing model predictions or simulations.

[22] While in this paper we only show how the parameterization may be trained with precipitation data, a similar procedure may be applied to train the vertical redistribution of heat and moisture, or the detrainment of hydrometeors and their interaction with radiation, if observations become available. Results from process resolving simulations may of course also serve as observations. In addition, we want to emphasize that our approach using ensembles of (or within) parameterizations in combination with data assimilation techniques may also be used for other physical parameterization schemes. Any available observed data, not handled in traditional data assimilation schemes but related to the physics, could be assimilated directly into the model fields.

[23] This parameterization is currently being used in the operational 20-km RUC model [Grell and Dévényi, 2001] used at the National Centers for Environmental Predictions (NCEP), in a version with 144 ensembles. Work is in progress to fully use its capabilities (data assimilation techniques) within several NWP models.

[24] **Acknowledgments.** The authors are grateful to Drs. John Brown, Thomas Hamill, Steve Koch, and Mariusz Pagowski as well as an anonymous reviewer for their helpful scientific reviews of this paper. Thanks also to Nita Fullerton for her editorial review.

## References

- Arakawa, A., and W. H. Schubert, Interaction of a cumulus cloud ensemble with the large scale environment, Part I, *J. Atmos. Sci.*, 31, 674–701, 1974.
- Brown, J. M., Mesoscale unsaturated downdrafts driven by rainfall evaporation: A numerical study, *J. Atmos. Sci.*, 36, 313–338, 1979.
- Ebert, E. E., Ability of a poor man’s ensemble to predict the probability and distribution of precipitation, *Mon. Wea. Rev.*, 129, 2461–2480, 2001.
- Frank, W. M., and C. Cohen, Simulation of tropical convective systems, Part I: A cumulus parameterization, *J. Atmos. Sci.*, 44, 3787–3799, 1987.
- Grell, G., Prognostic evaluation of assumptions used by cumulus parameterizations, *Mon. Wea. Rev.*, 121, 764–787, 1993.
- Grell, G., and D. Dévényi, Parameterized convection with ensemble closure/feedback assumptions, *9th conference on Mesoscale Processes*, AMS, Ft. Lauderdale, pp 12–16, 2001.
- Grell, G. A., J. Dudhia, and D. R. Stauffer, A description of the Fifth-generation Penn State/NCAR Mesoscale Model (MM5), *NCAR Tech Note TN-398 + STR*, 122 pp., 1994.
- Härdle, W., *Applied nonparametric regression*, Cambridge University Press, Cambridge, 333 pp., 1990.
- Kain, J. S., and J. M. Fritsch, The role of the convective “trigger function” in numerical forecasts of mesoscale convective systems, *Meteorol. Atmos. Phys.*, 49, 93–106, 1992.
- Krishnamurti, T. N., S. Low-Nam, and R. Pasch, Cumulus parameterizations and rainfall rates II, *Mon. Wea. Rev.*, 111, 815–828, 1983.
- Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, Improved weather and seasonal climate forecasts from multimodel superensemble, *Science*, 285, 1548–1850, 1999.
- Kuo, H. L., Further studies of the parameterization of the effect of cumulus convection on large scale flow, *J. Atmos. Sci.*, 31, 1232–1240, 1974.
- Lord, S. J., and A. Arakawa, Interaction of a cumulus cloud ensemble with the large-scale environment, Part II, *J. Atmos. Sci.*, 37, 2677–2692, 1980.
- Miller, R. N., E. F. Carter Jr., and S. T. Blue, Data assimilation into non-linear stochastic models, *Tellus*, 51A, 167–194, 1999.
- Molinari, J., A method for calculating the effects of deep cumulus convection in numerical models, *Mon. Wea. Rev.*, 110, 1527–1534, 1982.
- Stephenson, D. B., and F. J. Doblas-Reyes, Statistical methods for interpreting Monte Carlo ensemble forecasts, *Tellus*, 52A, 300–322, 2000.

G. A. Grell, NOAA/OAR/FSL, Skaggs Research Center, 325 Broadway, Boulder, CO 80305, USA. (georg.a.grell@noaa.gov)

D. Dévényi, Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado/NOAA Forecast Systems Laboratory, Boulder, Colorado USA.