**The SCM concept and creation of forcing datasets**

Minghua Zhang, Richard Somerville, and Shaocheng Xie

* 1. The concept of SCM (Richard Somerville)
  2. Early studies using SCMs (Richard Somerville)
  3. Requirement on large-scale forcing data (Minghua Zhang)
  4. Forcing data from field experiments prior to ARM (Minghua Zhang)
  5. The ARM variational analysis of forcing data (Minghua Zhang)
  6. Input data in the ARM variational analysis (Shaocheng Xie)
  7. Available ARM forcing datasets (Shaocheng Xie)
  8. Application of ARM forcing data to understand and improve models (Minghua Zhang)

1.3 Requirement on large scale forcing data

In the simplest setting, the SCMs calculate the time evolution of the vertical distributions of temperature and water paper, schematically written as:

 (1)

, (2)

where subscript “*m*” denotes model values; “*LS*” stands for prescribed large-scale fields; “*phy*” denotes physical parameterizations;  and *q* are potential temperature and water vapor mixing ratio; other symbols are as commonly used. In the vertical advection term of the above equations, the simulated profiles of control variables of  and *q* are used. Therefore, the horizontal advective tendencies  and the vertical velocity  are the large-scale forcing (*Randall and Cripe*, 1999). In another formulation, the observed profiles of  and *q* are used in the vertical advection term. Therefore  is prescribed as the large-scale forcing. Equations (1) and (2) retain some feedback of the simulated fields to the forcing fields. It is sometimes referred as “2-D” forcing, while the second formulation is referred as “3-D” forcing, in which there is no feedback between the simulated fields and the large-scale forcing.

SCMs that use the “2-D” forcing are numerically unstable if a central difference scheme is used in the vertical advection term and a forward difference scheme is used in time integration. The SCMs need to use either a central time difference scheme or upstream vertical difference scheme that also meets the CFL condition for integration in time. An upstream vertical difference scheme provides an effective negative feedback to damp model bias at a:

/(

Therefore the SCMs are less likely to drift too far from observations. When high vertical resolution is used, the integration time step needs to be short. SCMs that use the “3-D” forcing are not subject to numerical instability in the use of the forcing terms, but instability can arise in the physical parameterization terms, such as the term of subsidence warming from the Arakawa-Schubert type of convective parameterizations (Arakawa and Shubert 1974). Because of the lack of feedbacks between model fields and the large-scale forcing, these SCMs are more likely to drift away from observations.

The same prescribed forcing fields can be also applied to Cloud Resolving Models (CRM) and to Large Eddy Simulation (LES) models. The CRMs and LESs simulate  and *q*, or more often their corresponding conservative variables of liquid water potential temperature and total liquid water (*Siebesma et al*. 2004; *Stevens et al.* 2005), at each horizontal and vertical grids. The forcing terms in Equations (1) and (2) are applied to each of these grids, with the model fields  and in the vertical advection term replaced by domain averaged values from the CRMs and SCMs.

Clouds, radiation, and precipitation are products of the physical parameterization terms. With the use of the large-scale forcing, SCMs become a convenient tool to understand and evaluate how the physical parameterizations behave. They also enable better comparison of model results with observations since the prescribed large-scale dynamics allow the physics biases to be isolated from biases in the dynamics.

The horizontal advection and the vertical velocity are required to specify the forcing data of SCMs in Equations (1)-(2). The horizontal advective in turn requires the calculation of the spatial gradient of temperature and water vapor. Their derivations from field measurements however are subject to large errors. These errors may induce compensating to amplified errors in the physical parameterization terms. Errors in the forcing data originate from two sources. One is the instrument and measurement errors. The second is errors from scale aliasing, or sampling errors. Both depend on scales because horizontal derivatives are involved. The smaller the scale is, the larger the errors.

If the accuracy requirement of the physical parameterization terms in Equations (1) and (2) are 1 (K/day) and 1 (g/kg/day), the comparable accuracy requirements on the errors of the horizontal differences of temperature () and humidity () over a distance can be estimated as the followings:





Given | and over a distance of 200 *km*, the above inequalities require that

 (3)

 (4)

Requirement on the pressure vertical velocity error is:



.

The requirement on the difference of horizontal winds across the domain is:

.

Assuming a vertical layer of 100 mb, we get

  (5)

The error bounds of the spatial differences in Equations (3)-(5), corresponding to accuracy requirement of 1 (K/day) and 1 (g/kg/day) in the forcing data, need to be scaled proportionally if the horizontal scale is different from 200 km. When the scale is smaller, the bounds are smaller. These magnitudes are close to absolute instrument errors (Zhang and Lin 1996), but since they are relative errors, the systematic instrument errors are removed if the same equipment is used across the domain, while the random errors can be suppressed by averaging over vertical levels. The more problematic errors are those caused by scale aliasing or sampling bias. These errors are often handled by using statistical approaches. In ARM, they are additional dealt with by using known physical constraints.

1.4 Forcing data from field experiments prior to ARM

In the objective analysis of field experimental data of a sounding array, since the large-scale vertical velocity is always calculated from the mass continuity equation by vertically integrating the horizontal divergence of winds, both the horizontal advective tendencies and the vertical velocity in the large-scale forcing in Equations (1)-(2) can be obtained by using finite difference approximation of the horizontal derivatives. To do the finite differencing, the data need to be regularly spaced. Since the balloon sounding arrays are never regularly spaced, interpolation or extrapolation is needed to preprocess the data of temperature, water vapor and winds into a regular set of grids. This method is referred to as “regular grid method” in Zhang et al. (2000) (Figure 1a). In this method, the forcing data is calculated at each grid. Area averages are performed to obtain forcing for the domain that covers overlaps with the sounding array.

An alternate method is to write the advective tendencies in Equations (1)-(2) into flux form. The horizontal flux divergence terms, when averaged over a domain, is calculated from the line-integral method (Figure 1b). The approach, referred to as the “line-integral method” in Zhang et al. (2000), therefore gives the domain averaged forcing values.

A key element in the regular grid method is the fitting of atmospheric state variables to the grids by interpolation and extrapolation. The fitting method depends on the choice of the functional form, which can be quite subjective. Commonly used methods are linear fitting, quadratic or spline fitting (Davis-Jones 1993; Thompson et al. 1979). The more convenient algorithms are the Barnes (1964) and the Cressman (1959) schemes (Lin and Johnson 1996). In these schemes, a background field (or initial guess) is used at the observational locations, the difference between the observation and the initial guess field is then interpolated to the regular grids to adjust the background fields at these grids. The calculation can be performed iteratively to reach the desired corrections. Both the interpolation method and the number of integrations can affect the final analysis. More sophisticated uses statistical interpolation scheme such as Ooyama (1987).

The line integral method depends on the quality and number of atmospheric measurements at the boundary of the study domain. It therefore contains fewer subjective assumptions than the regular grid method. An important requirement is that there needs to be sufficient amount of measurement stations at the domain boundary. The line integral method typically does not use measurements inside the study domain in the calculation of the lateral boundary fluxes.

The regular grid method is more suited to analyze data with many scattered measurement stations, while the line integral method is more suited for a sounding array with few measurement stations. Zhang et al. (2000) presented a hybrid approach in which the regular grid method is used to improve the lateral boundary fluxes in the line integral method, therefore making use of sounding data besides those at the boundary stations.

Both methods have been used in the past to derive SCM large-scale forcing data in field experiments. One of the most widely used legacy data was from the Global Atmosphere Research Experiment (GARP) Atlantic Tropical Experiment (GATE) in 1974. Ooyama (1987) described his five year concentrated research to derive the objective analysis by designing a statistical regular grid method to which a penalty function is imposed to ensure smoothness of the fields. While the GATE data by Ooyama (1987) data have been widely used, a standard analysis algorithm is not available because many subjective procedures and judgment were through trial and error tests for each data point.

Another widely used SCM forcing data were from the Tropical Ocean and Global Atmosphere Coupled Ocean–Atmospheric Response Experiment (TOGA COARE) from November 1992 to February 1993. Lin and Johnson (1996) used the Barns analysis and the regular grid method to derive the forcing data. They Johnson also tried a cubic-spline mechanical fitting algorithm described in Ooyama (1987) and found that the results are similar to those from the Barnes analysis. Frank et al. (1996) also analyzed the TOGA COARE data but used the line-integral method. The difference of the moisture budgets from these two analyses over the Intensive Flux Array (IFA) gives a rough estimate of uncertainties: The time averaged diagnosed precipitation over the experiment period is 5.7–6.1 mm/day in Lin and Johnson (1996) and 10.5–11.8 mm/day in Frank et al. (1996). Uncertainties in the transient data are even larger, and these are likely the typical me. Therefore, although the analyzed data can be used to study the qualitative temporal variation of the large-scale atmospheric phenomena such as the Maddan-Julian Oscillation (MJO), their use to simulate the observed cloud fields needs special caution.

Large-scale forcing data have been also calculated for other shorter field experiments. Many of these are in regions of Asian or Australian monsoons. They have been summarized in Zhang et al. (2000). Uncertainties of the analyzed data are likely similar to those in TOGA-COARE. These uncertainties represent fundamental limits of data from the balloon sounding arrays as described in Section 1.3. In concluding his analysis of the GATE data, Ooyama (1987) expressed his view on these uncertainties: “To make gold, one must start with gold.”

1.5 The ARM variational analysis of forcing data

Recognizing the limit of accuracy in the large-scale forcing data derived from balloon sounding measurements, especially transient forcing data for SCMs to compare with cloud and radiation measurements from the ARM program, we developed a constrained variational algorithm to incorporate more measurements in the derivation of the forcing data. We enforce known constraints to the forcing data. These constraints include column-integrated conservations of atmospheric masses of moist air and water vapor as well as heat and momentum. They are written as

 (1)

 (2)

 (3)

 (4)

In the above, (u, v, *s, q*) are the atmospheric state variables of winds, dry static energy, and water vapor; *ps* is the surface pressure; *ql* is the cloud liquid water content; *φ* denotes the geopotential height. The bracket represents vertical integration. *Es* is the surface evaporation. *Prec* is the surface precipitation. *R* is the net downward radiative flux; the subscript TOA and SRF represent the top-of-the-atmosphere and the surface. *L* is the latent heat. *SH* is the surface sensible heat flux. denotes the wind stress at surface . Other variables are as commonly used.

The final analysis is obtained by minimizing the cost function of

 (5)

where variables with the subscript “o” represent first guess from pre-processed balloon sounding and wind profiler measurements; *B* is the error covariance matrix of each state variable.

Terms on the left-hand sides of Equations (X)-(X) are obtained from ARM measurements at the surface and satellite measurements at TOA. Area averaged precipitation is from radar measurements. Other surface variables are from the suite of ARM instruments deployed within the sounding arrange.

At the ARM SGP (Southern Great Plain) site, the following ARM measurement platforms were used: (i) the surface meteorological observational stations (SMOSs) that measure surface pressure, winds, temperature, and relative humidity; (ii) the energy budget Bowen ratio (EBBR) stations that measure surface broadband net radiative flux and surface sensible and latent heat fluxes; (iii) the Oklahoma and Kansas mesonet stations measuring surface precipitation, pressure, winds, and temperature; (iv) the Eddy Correlation (ECOR) stations that measure surface sensible and latent heat fluxes, and (v) the microwave radiometer stations stations measuring total cloud liquid water. Figure XX shows the locations of these stations. The satellite measurements are from NASA Langley at half hour temporary resolution (Minnis et al. 1995). At other ARM sites, the surface measurement stations are not as comprehensive as over SGP. The surface flux constraints therefore contain larger uncertainties due to possible sampling biases. In some cases, fluxes are derived from statistical interpolation between the limited number of station measurements and the background field from the reanalysis fields. Since each field experiment has different measurement configurations, the pre-processing of surface and atmospheric measurements may be different for different experiments. What we learned was that visual inspections of all input data are necessary to eliminate surprises.

The minimization of the cost function in Equation (X) requires the specification of the error covariance matrices. They errors are taken as the sum of those due to instruments and measurement biases and sampling biases. The sampling and measurement biases are assumed as 0.5 m/s for winds, 0.2 K for temperature, and 3% of the specific humidity for water vapor. These were estimated by instrument mentors. The sampling biases are estimated to be twenty percent of the temporal variances of the fields. In the past analysis, the errors are assumed to be independent among different locations and variables. This assumption is being revised to allow error covariances. The minimization algorithm used an iterative linearization method described in Zhang and Lin (1997).

The final analysis therefore is the closest to pre-processed balloon sounding and wind profiler data. It also satisfies the required constraints of Equations (X)-(XX). The divergence terms in these equation terms are calculated by using the line-integral method. The atmospheric state variables at the boundary stations are pre-processed by using the regular grid method so that data from all profiling stations are used.

The constraining requirements ensure that what enters an atmospheric column in terms of mass, water vapor, and energy is equal to what exits from the lateral boundary, at the TOA and surface after adjusting for column integrated change. The forcing data can be considered as a better fitting of the atmospheric analysis to more observational measurements.

We current treat the left hand side terms in Equations (X)-(XX) as known fields. Sensitivities of the analyzed fields to their uncertainties are used to characterize the errors in the forcing data (Zhang et al. 2000). In theory, these constraining variables can be also included in the cost function subject to variational adjustments based on their uncertainties. Furthermore, the imposed constraints can be expanded to include other known physical relationships and measurements, such as clear-sky water vapor and thermodynamic equations and radiance measurements at various wavelengths. The atmospheric state variables should ideally also include cloud hydrometeors, in which case radar reflectivity and cloudy-sky radiance measurements can be used as constraints. Additionally, the error covariance matrices in the cost function should be better calculated. Research is ongoing to make improvements to the current variational algorithm.

In addition to the analysis of the Intensive Observational Period (IOP) that has balloon measurements over a sounding array with high temporal frequency, ARM also developed the continuous forcing data in which the initial guess fields in the cost function of Equation (X) were entirely from hourly operational analysis but the surface and TOA constraining fields were from observations (Xie et al. 20??). The advantage of the continuous forcing is that it can be carried out for long periods and over most regions of the globe as long as surface and TOA measurements are available. The disadvantage is the compromise of the vertical structure of the analyzed fields that will be heavily dependent on the operational analysis.

It is currently

REFERENCES

Barnes, S. L., 1964: A technique for maximizing details in numerical map analysis. J. Appl. Meteor., 3, 396-409.

Cressman, G. P., 1959: An operational objective analysis scheme. Mon. Wea. Rev., 87, 367-374.

Xie, S., R. T. Cederwall, and M. Zhang, 2004: Developing long-term single-column model/cloud system - resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations. J. Geophys. Res., 109, D01 104.

Zhang M. H., J. L. Lin, R. Cederwall, J. Yio, and S.C. Xie, 2001: Objective Analysis of ARM IOP Data: Method, Features, and Sensitivity. Mon. Wea. Rev., 129, 295 – 311.