Appendix E: Some FDDA Definitions

Adjoint Method -

a computationally efficient method to obtain the gradient of a cost function (a function of a model output parameter) with respect to a model input parameter. See Variational Method and Optimal Control Theory below. Traditionally, this technique is used to fit a model to data distributed over a finite period of time. If model errors can be neglected over the time period spanning the observations, this technique can determine the optimal initial state such that the corresponding model evolution best fits the observations through the entire period. Other uses of this technique include parameter estimation (e.g., Optimal Nudging) and sensitivity analysis. The use of adjoint models is an application of optimization and optimal control theory.

Analysis (Objective analysis) -

Method which automatically produces a regularly-spaced gridded numerical representation (or its spectral equivalent) of a variable field from a set of observations distributed randomly in space. A good analysis should fit the data, but not so closely that data errors (measurement and representativeness) adversely affect the “smoothness” of the analysis. Types of objective analyses include “surface fitting” methods (e.g., polynomials and splines), empirical linear interpolation (e.g., Cressman successive correction method, Barnes analysis scheme), statistical methods (e.g., OI, Kalman Filter) and variational methods (e.g., 3-D variational analysis, adjoint or 4-D variational methods).

Assimilation -

The process by which observations are absorbed or ingested in a running prognostic model. Commonly, it also refers to the mere process of incorporation of observations in a prediction model.

Assimilated State -

In an idealized sense, this refers to a state in which observations and a numerical representation of the meteorological variables in a prediction model are interchangeable and obey prescribed physical constraints. Loosely, it also refers to the model state after insertion of observations.

Asynoptic Data -

Observations not all taken at the same time.

Background -

Same as first guess.

Continuous Assimilation (old definition) -

The process by which asynoptic data are inserted into a model at the closest model time step corresponding to the actual time of the observation. That is, each observation is put into the model once - at the exact time of the observation, with no temporal spreading. (This is the temporal equivalent of direct insertion in space). Note: The reason why this definition has disappeared from the literature is that continuous direct insertion causes excessive shock to a model and should never be used!

Continuous Assimilation (new definition) -

Continuous or repeated (every time step) dynamical assimilation where forcing functions which gradually “nudge” the model state toward the observations are added to the governing model equations. The adjoint technique is also referred to as a continuous assimilation method because it takes into account the model error throughout an assimilation period, but without necessarily adding any artificial forcing terms into the forward model.

Cost Function (Functional) -

An objective measure of success (performance index) when using a variational approach for objective analysis, data assimilation, sensitivity studies, etc. It typically includes the sum of the squares of differences between the analysis or model estimate and the observations (i,e., measure of the “fit” in a least-squares sense), and it may be extended or augmented to include some empirical or dynamical constraints (e.g., model constraint). This extended cost function, also known as the Lagrangian or Lagrange function, is used to compute the gradient of a model output parameter with respect to a model input parameter. See Adjoint Method and Variational Method.

Cycling Techniques -

Continuous or intermittent data assimilation techniques which require the model to be integrated both forward and backward in time about the observations.

Direct Insertion -

The process of putting an observation into a model at the closest grid point to its actual location, with no spatial spreading. (This is the spatial equivalent of the old definition for continuous assimilation in time). Note: Direct insertion causes excessive shock to a model and should never be used!

Dynamic Analysis -

An analysis method which uses a numerical model to continuously assimilate the data and thus provides coherence and dynamic balance among the different fields, and thus produces the best possible numerical representation of the atmosphere. The FDDA is usually applied continuously throughout the period without necessarily running any free forecast following the data assimilation period. Compare to Dynamic Initialization.

Dynamic Assimilation -

A continuing integration of the model is interrupted periodically for the current model representation to be updated by timely observations. This can refer to either continuous (every time step) or intermittent (every 1-12 hours) assimilation.

Dynamic Balance -

State of analyzed or model-based gridded fields in which the data values are consistent with some dynamic constraint(s) or a set of governing model equations. The result is intervariable consistency, where the mass and wind fields are dynamically and physically constrained (e.g., geostrophic balance, gradient balance, primitive-equation balance), and the vertical motion, divergence and precipitation patterns are also consistent. Dynamic consistency is critically important if these data are being used as input to other models, especially air- chemistry models.

Dynamic Initialization -

Observations are continuously assimilated into a model during a pre-forecast period (say, t = - 12 h to t = 0 h), after which the FDDA (usually a nudging approach) is turned off. The model is then integrated in a purely prognostic (free forecast) mode. This procedure produces an initial state at t = 0 h which is in dynamic balance with respect to the assimilating model and also contains model-generated mesoscale details (divergence and moisture patterns) which may not be resolved by the observations.

Estimation Theory -

A branch of probability and statistics concerned with deriving information about properties of random variables, stochastic processes and systems based on observed samples. Some of the important applications of estimation theory are found in control and communication systems, where it is used to estimate the unknown states and parameters of the system. Approaches used for estimation include least-squares, maximum-likelihood and Bayesian. Classes of problems include prediction, filtering and smoothing.

Filtering -

The filtering problem is that of determining the best estimate of the state vector of grid-point or spectral values at the end of the time interval over which data are provided, t = tf. The solution of this problem is provided for a linear system by the Kalman Filter. For a nonlinear system no solution which is both computable and truly optimal exists, but there are approximations (e.g., Extended Kalman Filter).

Prediction -

The prediction problem is that of determining the best estimate of the state vector after the last available observation, t > tf.

Smoothing -

The smoothing problem is that of estimating the state vector at interior points, t0 < t < tf. The Kalman smoother and the adjoint method exhibit certain analogies.

First Guess (Background) -

A state vector of the atmosphere which is subsequently corrected by observations via an objective analysis procedure. For example, the first guess may be climatology or a large- scale analysis.

Forced Adjustment -

Explicit balancing of the mass and wind fields in the model after insertion of mass or wind observations. A prior knowledge of the mass-wind relation is required to carry out the adjustment. This is done for the following reasons: a) to provide information to a variable that is not directly updated with observations, b) to minimize the insertion shock and c) to ultimately accelerate the assimilation process. (This is not possible at scales where the exact balance is not known).

Four-Dimensional Data Assimilation (FDDA) -

Data assimilation with three space dimensions and time. Combining current and past data in an explicit dynamical model such that the model's prognostic equations provide time continuity and dynamic coupling among the fields. McPherson (1975) points out that the traditional forward-marching data assimilation methods are not fully four-dimensional since the data insertion can only affect the model solution after it is inserted and not before. Cycling techniques and the adjoint method would fit his definition of fully four- dimensional data assimilation, while the other methods are perhaps only three-and-a-half dimensional at best.

Forward-Marching Techniques -

Continuous or intermittent data assimilation techniques in which the model assimilates data as the model is integrated only forward in time.

Identical/Fraternal Twin Experiments -

An OSSE in which the numerical model used for model simulation is also used to generate the observations is called an identical twin experiment. A fraternal twin experiment uses different models for simulation and production of the observations.

Indirect Insertion -

Insertion of an observation into a model by spreading the correction in space. For example, grid points within a certain radius of influence of the observation are influenced by the insertion of the observation. Note: Indirect insertion minimizes insertion shock.

Induction -

Modification of a model variable via insertion of another model variable. For example, when the wind field is assimilated, it influences or induces the temperature field to some extent. In a simple model, the mass field responds to the wind field, and vice-versa, by geostrophic adjustment considerations which involve computing the Rossby radius of deformation. Similarly, the wind (mass) field can induce the mass (mass) field by advection and the complex adjustment processes inherent in the governing model equations. Induction effects are very important since very few observing systems provide both mass and wind observations. Note: Even when both wind and mass data are available, they are most likely “unbalanced” and a geostrophic-type adjustment occurs within the model in an attempt to balance the inserted data.

Initialization (Dynamic Balancing) -

The process of suitably adjusting the model initial conditions to control the high-frequency motions and achieve a greater balance of the initial state for a prognostic model. This can be done gradually using a continuous dynamic initialization via nudging, or at the end of each analysis-forecast cycle of an intermittent approach using normal mode or digital filtering techniques.

Insertion -

Process of replacing/modifying model-simulated values with “observed” values of the same variable. Same as assimilation.

Insertion Frequency -

How often data are inserted in the context of intermittent assimilation.

Insertion Shock -

Gravitational mode energy associated with insertion of data into a model. The inserted data are often incomplete (since only one variable may be available) and usually incompatible with the model fields (since both the observations and the model simulation contain errors). Following the data insertion an imbalance is created between the mass and wind fields. The model fights to restore balance by generating gravity waves. If this gravity-wave activity or insertion shock persists it is a sign of an imperfect assimilation. Insertion shock is usually not a problem for the continuous assimilation schemes (e.g., nudging) because the corrections are made gradually and are relatively small. On the other hand, insertion shock is almost always a problem when using intermittent assimilation schemes which perform the correction all at once and thus, normally require an explicit initialization step (e.g., nonlinear normal mode initialization, NNMI) following the data insertion and before the next model forecast.

Intermittent Assimilation -

Type of FDDA used at many of the world’s operational centers. This procedure inserts observations stratified into groups at regular intervals, and is also known as an “analysis-forecast cycle”. It involves initializing an explicit prediction model, using the subsequent forecast (e.g., 1 to 12 h) as a first guess (background) in a static, three-dimensional objective-analysis step (usually OI), and then repeating the process every 1 to 12 h.

Kalman Filter -

From Estimation Theory, a sequential estimation process involving both a linear dynamic model and a linear prognostic model of the dynamic model's statistical errors. It can be thought of an extension to the OI approach where the error covariance statistics are no longer prescribed but predicted to vary in both space and time.

Measurement Error -

Same as instrument error

Memory -

The ability of a model to retain the information inserted into the model via data assimilation.

Multivariate -

A type of statistical objective analysis (e.g., OI) where the weights used to spread observational corrections to the grid for a particular variable use information for more than that one variable. For example, the objective analysis of wind or height often makes use of the geostropic wind law and the observed and model statistics for both wind and height.

NNMI -

Nonlinear Normal Mode Initialization. A technique which assures that there is no growth in inertia-gravitational wave energy at the beginning of a model integration. This is in contrast to removing all inertia- gravitational modes as in Linear Normal Mode Initialization (also referred to as Normal Mode Initialization). It has been shown that some inertia-gravity wave energy is needed in the initial conditions to control gravitational-mode growth. The nonlinear terms in the model equations were successfully incorporated in the normal mode appraoch (NNMI) independently by both Machenhauer and Baer in 1977). See Initialization. However, these normal mode techniques have had limited success on smaller scales and in the tropics, especially for limited-area models. Even diabatic NNMI technques have difficulty dealing with the classical spin-up problem. Note that spin-up and balance are not interchangeable.

Nudging (Newtonian relaxation) -

A continuous assimilation method that relaxes the model state toward the observed state by adding to one or more of the prognostic equations artificial tendency terms based on the ng to one or more of the prognostic equations artificial tendency terms based on the difference between the two states. When the differences are computed at the model grid points using gridded analyses of data, it is called "Analysis Nudging", and when the differences are computed directly using the observations which are distributed nonuniformly in space and time, it is called “Obs Nudging”. Combinations of these two approaches can be used to assimilate different types and scales of data simultaneously within the model’s nested-grid framework. We call this approach, which is now commonly used in MM4 and MM5, “multiscale FDDA”. Finally, the proportionality factor which multiplies the difference between the observed and model values to produce the nudging tendency term is called the nudging coefficient, and when it is optimally determined using an adjoint approach, we refer to this type of nudging as “Optimal Nudging”.Optimal Control Theory dynamical systems with one independent variable, usually time, in which control (input) variables are determined to maximize (or minimize) some measure of the performance (output) of a system while satisfying specified constraints. See Adjoint Method.

OI -

Optimal Interpolation, a statistically-based (statistical linear estimation) objective-analysis technique where the set of weights assigned to the observations used to correct the first guess (background) is computed so as to minimize the analysis error variance. This technique is dependent on prescribing error covariances for the observed and background (model) fields.

OSE -

Observing Systems Experiment, where the effects of real-data observing systems on numerical simulations are investigated.

OSSE - Observing Systems Simulation Experiment, similar to an OSE except model-simulated data are used as observations instead of real data.

Rebound Effect -

For example, when using nudging for dynamic initialization, the shock felt by the model if the nudging term is suddenly removed without gradually decreasing its magnitude. It is best to ramp down the magnitude of the nudging term before turning the FDDA off so that the primitive-equation balance can be maintained immediately following the assimilation period. This type of shock is always present when using an intermittent FDDA approach, and thus requires an explicit initialization (e.g., NNMI) step.

Reference Level -

A level of known altitude where temperature, pressure or perhaps wind are specified. It is mainly used as a starting height for hydrostatic integrations.

Relaxation -

Same as nudging.Representativeness Error - Uncertainty introduced by the multi-scale contributions of an observation. That is, a strong wind on top of a mountain may be correct, but not representative of the scale (grid resolution) of the model or objective analysis.

Spin-up -

The process of establishing harmony between the dynamical and physical processes (e.g., convection, boundary-layer fluxes, etc.) in a model. Its symptoms include underprediction of divergence fields and precipitation during the early stages of a model integration. Because of the underdeterminacy problem and incomplete and contaminated observations, this harmony rarely exists at the beginning of a model forecast. The effects of model spin-up can be greatly reduced by performing a dynamic initialization. Note: Initialization (Dynamic Balancing) techniques such as NNMI do not necessarily reduce spin-up problems.

State Vector -

Represents the state variables of a system or model (e.g., winds, temperature, moisture, etc.)

Static Initialization -

Use of an objective-analysis technique to define the initial state of a model, but an explicit balancing (initialization) step is still required. For example, use of a Successive-Correction or an OI technique as part of an intermittent assimilation scheme is generally considered a static initialization, even though a model forecast is used as a first guess, because an explicit initialization step (e.g., nonlinear normal mode initialization) is still required. A true dynamic initialization (e.g., using nudging during a preforecast period) does NOT require a separate balancing step.

Underdeterminacy -

The common condition which arises when the size of the model, defined as the total number of grid points multiplied by the number of dependent variables, exceeds the number of observations.

Updating -

Same as insertion.

Rejection -

When the model state, following data insertion, resembles that which would have resulted without any data insertion. When the impact of the inserted data is minimal, the model has basically rejected the injected information.

Tangent Linear Model (TLM) -

A linear prognostic model for perturbations of the state vector often used within the adjoint approach. In other words, the TLM can predict how small perturbations in the model initial state will grow with time.

Univariate -

A type of statistical objective analysis (e.g., OI) where the weights used to spread observations of a particular variable to the grid use statistical information for only that one variable. For example, moisture is usually analyzed using a univariate approach.

Variational Method -

Use of the calculus of variations to impose dynamic or empirical relationships as mathematical constraints by minimizing a performance index (cost function or functional) while performing the objective analysis. This functional includes objective constraints which commonly measure the fit of the analysis to the observations as well as its smoothness and consistency to dynamic relationships (geostrophy, thermal wind, etc.) These constraints can be satisfied approximately (weak constraint) or exactly (strong constraint). When this method is used in four dimensions to include time, the dynamic constraint can be the advection or diffusion equations, conservation of vorticity or even a full primitive-equation model. The adjoint method is merely a more efficient way to solve this four-dimensional variational problem (a constrained minimization problem) when a numerical model is used as a strong (exact) constraint.

3-D VAR -

A three-dimensional (the space dimensions) application of the variational method.

4-D VAR -

A four-dimensional (with time as the fourth dimension) application of the variational method.

(contributed by Dave Stauffer of Penn State)