Introduction to Model Initialization

Whether conducting idealized or real-data numerical simulations, initial conditions are required. As their name implies, initial conditions provide initial values for all model variables at all model grid points (or volumes, etc.). The method by which observations are processed to define the initial conditions is known as initialization. The processing that occurs during model initialization includes:

- Performing quality control upon the available observations.
- Assimilating the observations so as to update a ‘first-guess’ for the initial conditions.
- If applicable, ensuring dynamical balance within the updated initial conditions.

These concepts are discussed further in this and subsequent lectures. We begin with the first concept, observation processing and quality control.

Observations

There exist two broad classes of observations: in situ and remotely-sensed. In situ observations are those collected by sensors located at the observation site. METAR, rawinsonde, buoy, and aircraft observations are examples of in situ observation platforms. By contrast, remotely-sensed observations are those collected by sensors not located at the observation site. Doppler radar, lidar, wind profiler, and satellite imagers are examples of remote-sensing observation platforms.

Remote-sensing observation platforms measure radiation energy. Two classes of remote-sensing observation platforms exist: active and passive. Active remote-sensing platforms contain sensors that both emit radiation energy and measure the atmospheric response to that radiation. Doppler radar is an example of an active remote-sensing platform. Passive remote-sensing platforms contain sensors that measure radiation emitted, scattered, or reflected by some feature; they do not emit radiation themselves. Satellite imagers are examples of passive remote-sensing platforms.

Most observations, particularly those that are remotely-sensed, are not of model dependent variables. For example, consider Global Positioning System (GPS) radio occultations, which make use of radio signals transmitted by GPS satellites. As these radio waves travel through the Earth’s atmosphere, they become deflected. The extent to which they are deflected, or the bending angle, is related in part to temperature and moisture content. Models do not predict bending angle, but they do predict variables that are related to bending angle. This necessitates a retrieval algorithm, a physically- and/or empirically-derived relationship between observed and predicted variables, to convert the observations to something usable by the model. Retrieval algorithms can operate both within and outside the context of a data assimilation routine.
In situ and remotely-sensed observations have strengths and weaknesses that must be accounted for when being used in the model initialization process. These include:

**In Situ Observations**

*Strengths*

- **Minimal use of retrieval algorithms.** In situ observation platforms, including METAR stations, buoys, ships, aircraft, and rawinsondes, typically provide observations of fields such as temperature, moisture content, wind speed and direction, and pressure that are closely related to model prognostic variables. As a result, errors associated with the use of a retrieval algorithm, described below, are minimized with in situ observations.

*Weaknesses*

- **Observation representativeness.** In situ observations are point observations. Thus, an observation may reflect local, sub-grid-scale variability that is not representative of the scales of motion resolved by the model. Representative examples include observations taken within mountain waves and planetary boundary layer eddies, as well as those taken by poorly-sited instruments (e.g., bank thermometers). With the exception of those collected from poorly-sited instruments, temporally averaging the observations over some modest interval may help to mitigate this weakness by damping the local variability.

- **Data density.** In situ observations tend to be tightly clustered around where people live; more precisely, they are unevenly distributed. A representative example of this is given in Figure 1 below. As a result, when only in situ observations are considered, the resulting initial conditions may be relatively uncertain in areas with lower data density.

- **Temporal availability.** Observations from selected platforms, particularly rawinsondes (once every 6, 12, or 24 h) and aircraft (dependent upon flight route), are not available as frequently as are observations from other platforms. As initial condition quality is in part related to the number of available observations, this can influence initial condition quality at times when rawinsonde or numerous aircraft observations are unavailable and remotely-sensed observations are not considered.

- **Observation uncertainty.** Sensors are calibrated such that the observations they collect are said to be accurate within a specified tolerance threshold. This defines the observation uncertainty inherent to a given observation platform. This is typically small in magnitude but must be accounted for in the data assimilation process.
Figure 1. Observations, colored by type per the legend, assimilated at 0000 UTC 2 December 2015 by the ensemble adjustment Kalman filter used by the NCAR Ensemble. Satellite cloud-track winds (green) are the only remotely-sensed observation type assimilated. Note the lack of observations over Mexico and Canada versus the United States, owing both to population density and observation infrastructure. Image obtained from the NCAR Ensemble page linked above.

Remotely-Sensed Observations

Strengths

- **High spatial and temporal resolution.** Most remote-sensing observation platforms have high spatial and temporal resolution. For instance, National Weather Service Doppler radars are provide observations every five minutes with $\Delta r = 250$ m and $\Delta \lambda = 0.5^\circ$ at elevation angles ranging from 0.5° to 19.5° above the horizon. Of particular value is the high spatial resolution ($\Delta x \sim 1$-$16$ km) of satellite-based remotely-sensed observations that are more routinely used in the model initialization process.

Weaknesses

- **Observation uncertainty.** As with in situ platforms, sensors are calibrated such that the observations they collect are said to be accurate within a specified tolerance threshold. This defines observation uncertainty inherent to a given observation platform. This is typically small in magnitude but must be accounted for in the data assimilation process.

- **Retrieval algorithm errors.** Relationships between remotely-sensed and model variables are imperfect owing to limited understanding of the underlying physics. In some cases, the underlying relationships are ill-posed; e.g., an observed quantity is related to two or more model variables, and knowledge of the other variable(s) that we are also trying to update
is needed to obtain a given model variable. Consequently, retrieval algorithms introduce additional observation uncertainty that can compromise the quality of the initial conditions if this uncertainty is sufficiently large. Note that relatively limited assimilation of remotely-sensed observations is performed outside of operational forecast centers owing to the intrinsic complexity associated with retrieval algorithms.

- **Limited ability to observe near the surface.** Satellite-based remote-sensing platforms are ideal for observation collection within the upper atmosphere. However, some sensors are not capable of sensing, whether accurately or at all, below clouds. Consequently, there are generally fewer satellite-based remotely-sensed observations near the surface, including in the planetary boundary layer.

**Quality Control**

Before an observation can be assimilated so as to update the initial conditions for a numerical model forecast, one must ensure that the observation is of sufficiently high quality such that its assimilation does not degrade the quality of the initial conditions. The means by which this is accomplished are known as **quality control**. A quality control algorithm must be able to handle multiple observation types from multiple observation platforms in multiple locations, it must be able to accurately distinguish between erroneous and robust observations, and it must be able to work with a minimum of human intervention.

There exist multiple reasons why an individual observation may be in error. For example, an observation may be robust, but *corruption* of the observation value, time, date, or location may occur upon its transmission to the agency responsible for the observation platform. Observations may also be in error due to an improper calibration of the sensor used to collect the observation. Such an error is known as a *systematic error*, as the extent to which the observation is erroneous is not random and thus theoretically can be corrected. By contrast, observations may be in error due to a malfunction of the sensor used to collect the observation. Such an error is known as a *random error*, as the extent to which the observation is erroneous is altogether random. Finally, as described above, an observation may be representative of sub-grid- rather than resolved-scale variability. In this case, the observation is not erroneous but is not supported by observations taken at nearby times and/or locations. Such an error is known as a *representativeness error*.

There exist many tests that can be employed to determine whether an observation is in error. These include, but are not necessarily limited to, the following:

- **Sensor Limit Tests:** Is the observation outside of the range of values that the instrument used to collect the observation can reliably measure?
• **Climatological Limit Tests**: Is the observation well outside of the range of previously-observed values at the observation location?

• **Physical Limit Tests**: On the extreme of sensor limit tests, is the observation out of the true range of physical values, e.g., reporting negative relative humidity or wind speed?

• **Temporal Consistency Checks**: Is the observation inconsistent with observations taken at earlier (and, if available, later) times?

• **Spatial Consistency Checks**: Is the observation inconsistent with nearby observations? Does the observation depart significantly from the ‘first-guess’ for the initial conditions?

Quality control algorithms must reliably distinguish between robust and erroneous observations. It is straightforward to conceptualize how assimilating an erroneous observation can degrade initial condition and subsequent forecast quality. However, the incorrect identification of a robust observation as erroneous can also degrade initial condition and subsequent forecast quality. This most commonly occurs when an observation is erroneously determined to be non-representative.

To illustrate this point, we consider the example of the 24-25 January 2000 eastern United States blizzard. This event was associated with particularly large skill degradation of forecasts from regional and global numerical weather prediction models at lead times of 6-18 h. Zhang et al. (2002, *Mon. Wea. Rev.*) demonstrated that this forecast skill degradation primarily resulted from improper rejection of upper tropospheric wind observations from the 0000 UTC 24 January 2000 Little Rock, AR sounding.

The 0000 UTC 24 January 2000 Little Rock, AR sounding (Figure 2) indicates 140 kt winds at 300 hPa level. However, surrounding observations at Springfield, MO, Shreveport, LA, and Jackson, MS indicate much weaker winds, as does the 0-h RUC model analysis (Figure 3). Consequently, the data assimilation systems used to provide initial conditions to most major modeling systems rejected this observation as being non-representative of either surrounding observations or the ‘first-guess’ for the initial conditions. When this observation is assimilated, the resulting initial conditions depart substantially from those provided by operational models (Figure 4), and these departures influence the subsequent forecast.

**Model Spin-Up**

There exist three spatial scales of interest with respect to model initialization: those resolved by each of the observation network, the initial conditions, and the model simulation itself. For high-resolution numerical simulations, these are listed in descending order of what each can resolve. In order for the scales resolved by the initial conditions to be identical to those resolved by the model simulation, some means of generating realistic atmospheric variability on scales smaller than those explicitly resolved by the observation network is necessary. This is typically achieved by using...
the forecast of an earlier numerical model simulation as the ‘first-guess’ for the initial conditions, which is then subsequently adjusted in light of the available observations.

**Figure 2.** 0000 UTC 24 January 2000 Little Rock, AR sounding. Figure obtained from the [University of Wyoming Atmospheric Sounding Archive](http://www.wyo.edu).

**Figure 3.** 300 hPa analysis valid at 0000 UTC 24 January 2000. Station plots depict observations of temperature (red, °C), dew point temperature (green, °C), and wind speed and direction (blue; half-barb: 5 kt, full barb: 10 kt, flag: 50 kt). Also depicted is the 0-h RUC analysis of streamlines (black), isotachs (shaded, kt), and divergence (yellow contours, x10⁻⁵ s⁻¹). Note the discrepancy between the 300 hPa wind speed observation at Little Rock, AR and the 0-h RUC isotach analysis. Figure obtained from the [Storm Prediction Center Surface and Upper Air Map Archive](http://www.spc.noaa.gov).
Figure 4. 300 hPa wind difference magnitude (contoured every 3 m s\(^{-1}\)) between an initial analysis that assimilated the Little Rock, AR wind observation and initial analyses from the (a) Eta and (b) ECMWF models that did not assimilate the Little Rock, AR wind observation. Wind difference magnitudes are approximately 12 m s\(^{-1}\) in (a) and 15 m s\(^{-1}\) in (b) near Little Rock, AR, roughly consistent with the departure of the observed wind speed from the 0-h RUC analysis in Figure 3. Note that not all differences in each panel result exclusively from the assimilation of the Little Rock, AR wind observation; other data differ, as do the data assimilation systems used. Figure reproduced from Zhang et al. (2002, *Mon. Wea. Rev.*), their Figure 10.

Also of interest is the treatment of model prognostic variables for which reliable observations do not exist, are extremely computationally-intensive to assimilate, or for which reliable relationships between observed quantities and the model prognostic variable do not exist. A prime example is given by microphysical quantities – mixing ratio and, if predicted, higher-order moments – for all species except water vapor. The ‘first-guess’ may be able to provide estimates of these variables, but it is difficult to impossible to accurately update them if only conventional observations are assimilated. Consequently, a numerical model simulation may elect to resolve the kinematic and mass fields associated with active clouds and precipitation but not to resolve the clouds and precipitation themselves.

With all of this in mind, there exist three classes of model initializations:

- **Cold Start**: Atmospheric variability is absent on the scales between those resolved by the numerical model and those resolved by the initial conditions. Initial conditions for prognostic microphysical variables except water vapor are typically missing.

- **Warm Start**: Atmospheric variability is present on all scales resolved by the numerical model. Initial conditions for prognostic microphysical variables except water vapor are still typically missing, however.

- **Hot Start**: Atmospheric variability is present on all scales resolved by the numerical model. Initial conditions for all prognostic microphysical variables are present, and as a
result the model initial conditions include explicit representation of precipitating features that are ideally in balance with the initial kinematic and mass fields.

A representative example of a cold start model initialization is that in which the analysis from a larger-scale model such as the GFS is used to provide initial conditions for a higher-resolution numerical model simulation. A warm start model initialization typically results from the use of a cycled data assimilation system to generate realistic kinematic fields on the scales resolved by the numerical model simulation. Most operational forecast models use warm start initializations. The HRRR model, which assimilates radar reflectivity data to prescribe initial values for latent heating rate and microphysical prognostic variables, is an example of a hot start initialization.

In the absence of atmospheric variability on the scales between those resolved by the numerical model and those resolved by the initial conditions, or in the absence of initial conditions for all prognostic microphysical variables except water vapor, the model must generate (or spin-up) the needed fields. The time period over which this occurs is known as the spin-up period. The length of the spin-up period typically ranges from 1-12 h depending upon the extent to which the model needs to generate the necessary small-scale variability and microphysical data. This results in the general recommendation to begin a numerical model simulation 6-12 h prior to the forecast period of greatest interest.

As the model spins-up atmospheric variability on smaller scales than those resolved by the initial conditions, dynamical imbalance results. In response, the model generates spurious inertia-gravity waves to attempt to restore balance. In a local context, inertia-gravity waves are associated with a rapidly-fluctuating surface pressure. The domain-averaged local rate of change of the surface pressure tendency can thus be used as a metric to identify inertia-gravity wave activity such as may be associated with model spin-up, as illustrated in Figure 5. Inertia-gravity wave activity declines but does not altogether end as the model spin-up period ends; the model can produce physically-realistic inertia-gravity waves in response to imbalance throughout its duration.

**Figure 5.** Domain-averaged local rate of change of the surface pressure tendency (hPa s⁻²) as a function of time for two model simulations, one using well-balanced initial conditions (dashed) and one using poorly-balanced initial conditions (solid). Note the logarithmic scale to the y-axis. As the model spin-up period ends, roughly between 6-12 h, the slopes of each curve asymptote to zero. Figure reproduced from Warner (2011), their Figure 6.15b.
**Observation Targeting**

Fundamentally, forecast quality is directly proportional to both initial condition quality and model quality. As ‘first-guess’ initial conditions are typically provided by the short-term forecast from a previous numerical model simulation, initial condition quality can be improved if model error is reduced. Initial condition quality can also be improved through better use of existing observations or the deployment of new observation platforms. Absent new routine platforms, however, initial condition quality may also be improved if additional, non-routine observations are assimilated (e.g., as may be collected during a field program or as part of a targeted reconnaissance effort). This is particularly true when the additional observations come from data-sparse regions or the additional observations come from locations where large uncertainty in or forecast sensitivity to the initial conditions exists.

**Observation targeting** describes the processes by which the siting of additional observations may be determined to provide optimal benefit to forecast quality. This is most often done in the context of a new observation platform or field campaign. To illustrate the concept of observation targeting, let us consider the example of numerical forecasts of the track of Hurricane Sandy. Figure 6 depicts forecast tracks of Sandy from three successive cycles of the twenty-member GFS Ensemble. There exists minimal spread in the track forecasts prior to approximately 120 h, after which time forecast tracks significantly diverge, with some forecasts indicating a track out to sea and others toward the United States or Canadian Maritimes.

Presumably, this forecast divergence results in part from uncertainty in the initial conditions, but where? And of what model variables? An observation targeting method, applied to deterministic or ensemble forecasts, must indicate the observation types and locations that would have the greatest impact on the model forecast, whether operationally in the context of the next model cycle or in hindcasts for this model cycle. Ideally, collecting and assimilating targeted observations will improve model forecast accuracy. However, it should be noted that improved forecast quality due to assimilating targeted observations is miniscule when averaged over many cases. An example is provided in Figure 7. This is likely due to several factors, chief among them being imperfect observation targeting methods and model error exerting a dominant influence on forecast quality. That said, there are many different methods by which observation targeting may be accomplished, and we now wish to briefly describe several of these methods and their applications.
**Figure 6.** 240-h forecast tracks for Hurricane Sandy (2012) from the GFS Ensembles initialized at 1200 UTC 23 October 2012 (blue), 1800 UTC 23 October 2012 (green), and 0000 UTC 24 October 2012 (red). The analyzed position of Sandy at 1200 UTC 23 October 2012 is denoted by the black square, and the observed track of Sandy through dissipation is given by the black line.

**Figure 7.** Root-mean squared error (m) of 1000 hPa and 500 hPa geopotential height forecasts at lead times of 30, 36, 42, and 48 h from the FASTEX field experiment. Root-mean squared error is presented for two sets of forecasts, each conducted with the ECMWF model: one where no targeted dropsonde observations are assimilated and one where all targeted dropsonde observations are assimilated. The singular vector technique was used to target observations, and forecasts are verified only over the forecast sensitivity regions. Dots lying to the left of the dashed line indicate a positive impact from assimilating targeted observations. Figure reproduced from Montani et al. (1999, *Quart. J. Roy. Meteor. Soc.*), their Figure 8.
Ensemble Variance/Spread-Based Methods

This observation targeting method assumes that forecast error growth is largest where the initial condition uncertainty, and thus potential for initial condition error, is largest. For an ensemble of model initial conditions, uncertainty can be quantified in light of the ensemble spread, or variance, in one or more fields. Initial condition uncertainty is typically largest in regions of few observations and near sharp gradients in model dependent variables. Representative examples of each include northern Mexico or much of Africa, where few observations are routinely available, and near shortwave troughs or frontal boundaries.

Collecting and subsequently assimilating observations from locations in which initial condition uncertainty has the potential to constrain model error growth and improve forecast quality. This is true in a general sense, but in which locations would targeting and collecting observations lead to the greatest positive impact upon the forecast for the feature and/or region being considered? One could trace the initial condition uncertainty forward in time and space to obtain a first-guess as to where targeted observations should be collected, but other methods described below provide more robust means by which this may be accomplished.

Ensemble Sensitivity Metrics

There exist multiple quasi-objective ensemble-based methods by which the locations and types of targeted observations that may exert the greatest positive impact upon the subsequent forecast can be identified.

One example, ensemble sensitivity analysis (e.g., Ancell and Hakim 2007, Mon. Wea. Rev.), can be used to relate some forecast metric of interest $J$ to a model variable $x$ at the same or earlier time. Specifically,

$$\frac{\partial J}{\partial x} = \frac{\text{cov}(J, x)}{\text{var}(x)}$$

This represents the change in the forecast metric as one changes the model variable and can be shown to represent a form of linear regression between $J$ and $x$. cov represents covariance and var represents variance. Typically, this expression is multiplied by the standard deviation of the model variable $x$, such that this expression gives the expected change in $J$ that results from a one standard deviation change in $x$. An example of this method is depicted in Figure 8.
Figure 8. Ensemble sensitivity metric (shaded; mm change per one standard deviation change in the model variable $x$) relating changes in the forecast metric $J$ (here, 22-25 h area-averaged forecast precipitation over central Oklahoma) to a model variable $x$ (here, 315 K isentropic potential vorticity) at (a) 0 h, (b) 6 h, (c) 12 h, and (d) 18 h. Positive (negative) values indicate that a one standard deviation increase in $x$ is associated with a positive (negative) change of the specified amount in $J$. Contours in each panel depict the ensemble-mean 315 K isentropic potential vorticity (PVU). Values of the ensemble sensitivity metric that are statistically-significant to greater than 95% confidence, indicating robust linear relationships between $J$ and $x$, are stippled. Figure reproduced from Torn and Romine (2015, Mon. Wea. Rev.), their Figure 5.

Applying this method to targeted observations, consider Figure 8c. The 0000 UTC model run indicates that the 22-25 h forecast precipitation over central Oklahoma is particularly sensitive to the intensity of a shortwave trough evident in the 315 K isentropic potential vorticity field over the Texas Panhandle at 1200 UTC. Thus, it stands to follow that targeting observations of fields related to isentropic potential vorticity near 1200 UTC over the Texas Panhandle, followed by assimilating them into the initial conditions for a model simulation starting at 1200 UTC, could result in improved, or at least more certain (less variable), precipitation forecasts at 10-13 h (equivalent to the 22-25 h forecast period from 0000 UTC) over central Oklahoma.

A related method utilizes empirical orthogonal functions (EOFs) to identify modes of variability at a given forecast time from an ensemble numerical simulation. Once the leading mode(s) of variability have been identified, linear correlation is used to connect them to variability within a given model variable at the same or earlier forecast times. Figures 9 and 10 illustrate this method, and Zheng et al. (2013, Wea. Forecasting) provides more details about the method. The application of this method to targeted observations is similar to that for ensemble sensitivity.
Figure 9. (upper left) 120-h GFS Ensemble-mean forecast mean sea level pressure (contoured, hPa) and standard deviation (shaded, hPa) from the 0000 UTC 15 September 2017 run of the GFS Ensemble. (upper right) The leading EOF of the ensemble sea level pressure 120-h forecasts is characterized by uncertainty primarily in the along-coast position of tropical cyclone Jose. (lower left) The second EOF of the ensemble sea level pressure 120-h forecasts is characterized by uncertainty primarily in Jose’s intensity. (lower right) The third EOF of the ensemble sea level pressure 120-h forecasts is characterized by uncertainty primarily in Jose’s proximity to the United States east coast. Nearly half of the variance explained by the EOF decomposition is explained by the leading EOF; together, the second and third EOFs explain 30% of the total variance. Figure from http://breezy.somas.stonybrook.edu/CSTAR/Ensemble_Sensitivity/EnSense_Main.html.

Note that the ensemble sensitivity, EOF, and similar methods discussed below and elsewhere all use linear correlation. Correlation does not necessarily imply physical causation, which must be kept in mind when interpreting the results from either method. Even at their best, these methods are able to capture only a fraction of the true variance in the system, in part because the real atmosphere is highly non-linear. Nevertheless, they provide a means of identifying locations where targeted observations might improve a specific aspect of a numerical model forecast, namely within the context of ensemble numerical simulations.
Figure 10. Ensemble sensitivity metric, here representing the linear correlation between the leading EOF pattern in the upper right panel of Figure 9 and the GFS Ensemble-forecast 500 hPa geopotential height field at forecast lead times of 2.5 to 5 days (or 0 to 2.5 days prior to the valid time of the EOF analysis depicted in Figure 9). The upper leftmost panel indicates that decreasing the 500 hPa geopotential height in eastern North America (cool colors) in the five-day forecast is associated with positive EOF1, indicating a further northeast position along the coast for tropical cyclone Jose at the 120-h forecast time. Tracing this back to earlier forecast times, the lowermost right panel indicates that a stronger, faster-moving shortwave near the Canada-United States border in the 2.5-day forecast is associated with positive EOF1, also indicating a further northeast position along the coast for Jose at the 120-h forecast time. Figure from http://breezy.somas.stonybrook.edu/CSTAR/Ensemble_Sensitivity/EnSense_Main.html.

Adjoint-Based Methods

Similar to the ensemble-based methods described above, the adjoint method seeks to quantify the sensitivity of some forecast measure to the initial conditions. An adjoint operator, or the inverse of the linear version of the numerical model, identifies the quantitative impact of small, arbitrary perturbations of the initial conditions upon the chosen forecast measure. The perturbations may be random or may be obtained by some other means (e.g., from the difference in previous non-linear and linear model forecasts valid at the new forecast initialization time). An idealized schematic of the conceptual underpinnings and operation of adjoint-based methods is provided in Figure 3 of the “Lateral Boundary Conditions” lecture notes. Applied to targeted observations, the adjoint method identifies locations where initial condition uncertainty and, by extension, forecast error growth are large by identifying where small perturbations to the initial conditions have the largest impact upon the subsequent forecast.
Singular Vector Technique

The singular vector method provides a method to estimate how initial condition uncertainty or error propagates forward in time. It identifies the structures whose amplitudes grow most rapidly over a short but physically-relevant time frame (e.g., 24 h). A linear version of the numerical model is used to identify these structures. The singular vector method is applied to deterministic model forecasts, and the resulting singular vectors provide a basis to optimally generate an ensemble of initial conditions. This last point operates under the assumption that initial condition perturbations should be made where error growth is most rapid, such that an increased likelihood of bounding the true solution within the range of solutions offered by the ensemble is obtained. Applied to targeted observations, singular vectors provide insight as to the locations where added observations would have the greatest theoretical impact upon reducing forecast error growth if assimilated into the initial conditions. An illustrative example of the benefit obtained by using singular vectors to target observations is provided in Figure 7.