Model Initialization

Learning Objectives

Following this lecture, students will be able to:

- Provide a basic description of how numerical models obtain their initial conditions.
- Distinguish between the strengths and weaknesses of in situ versus remotely sensed observations, including errors introduced by retrieval algorithms.
- Describe what is meant by model spin-up and how cold, warm, and hot starts differ from one another.
- Describe techniques by which observation targeting, or the process of identifying which observation(s) in which location(s) would result in the greatest forecast impact, can be performed.

Introduction to Model Initialization

Initial conditions are required for both idealized and real-data numerical simulations. As their name implies, initial conditions provide initial values for all model variables at all model grid points (or cells, 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* on the available observations.
- *Assimilating* the observations to update a first guess for the initial conditions.
- If applicable, ensuring *dynamical balance* in 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, with examples that include METAR, rawinsonde, buoy, and aircraft observations. By contrast, remotely sensed observations are those collected by sensors not located at the observation site, with examples that include Doppler radar, lidar, wind profiler, and satellite imagers.

Remote-sensing observation platforms measure electromagnetic energy at specific wavelengths, varying based on what they are trying to measure. Two classes of remote-sensing observation platforms exist: **active** and **passive**. Active remote-sensing platforms contain sensors that emit

radiation energy *and* measure the atmospheric response to that radiation, with Doppler radar being a representative 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 variables. For example, consider Global Positioning System (GPS) radio occultations, which make use of radio waves transmitted by GPS satellites. These radio waves are deflected as they travel through the Earth's atmosphere, the extent of which is known as the bending angle. Models do not predict bending angle, but they do predict the temperature and moisture fields that influence the bending angle. Thus, **retrieval algorithms**, representing physically or empirically derived relationships between observed and predicted variables, are needed to convert the observations to something usable by the model.

In situ and remotely sensed observations have strengths and weaknesses that must be accounted for when being used to initialize a model. 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 variables. As a result, errors associated with using a retrieval algorithm 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 atmospheric boundary layer eddies, as well as those taken by poorly sited instruments (e.g., bank thermometers). Apart from those collected from poorly sited instruments, temporally averaging the observations over some modest duration 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 Fig. 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 on flight route), are not available as frequently as are observations from other platforms. Since initial condition quality is partially 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 accurate only to 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.
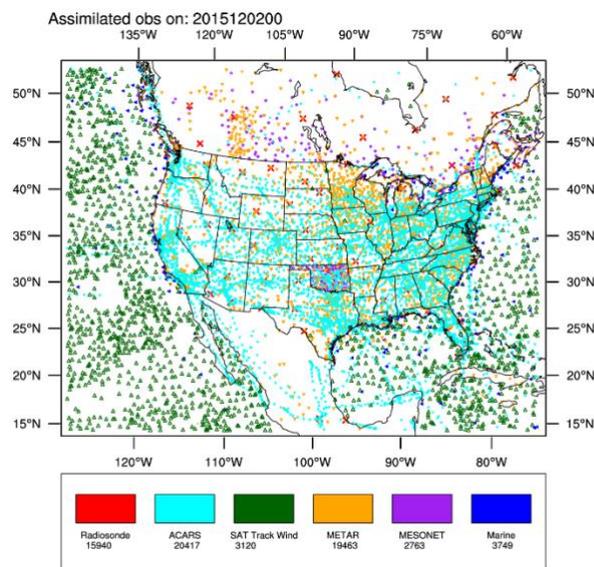

Figure 1. Observations, colored by type per the legend, assimilated at 0000 UTC 2 December 2015 by the ensemble adjustment Kalman filter used by the NCAR Ensemble. Satellite cloud-track winds (green) are the only remotely sensed observation type assimilated. Note the lack of observations over Mexico and Canada versus the United States, owing both to population density and observation infrastructure.

Remotely Sensed Observations

Strengths

- **High spatial and temporal resolution**. Most remote-sensing platforms have high spatial and temporal resolution. For instance, National Weather Service Doppler radars provide observations every five minutes every 250 m in the radial direction (away from the radar) and every 0.5° above the horizon at elevation angles ranging from 0.5° to 19.5°. Further,

most satellite-sensed observations that are routinely used to initialize a model have a spatial resolution that is finer than that of the model ($\Delta x \sim$ 0.5-8 km).

Weaknesses

- **Observation uncertainty**. As with in situ platforms, sensors are calibrated such that the observations they collect are accurate to within a specified tolerance. 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.

- **Retrieval algorithm errors**. Relationships between remotely sensed and model variables are imperfect owing to our 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 large.

- **Limited ability to observe near the surface**. Satellite-based remote-sensing platforms are ideal for observing 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 atmospheric boundary layer.

Quality Control

Before an observation can be assimilated to update a model's initial conditions, one must ensure that the observation is of sufficiently high quality such that its assimilation does not degrade the initial conditions' quality. 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, to accurately distinguish between erroneous and robust observations, and to work with a minimum of human intervention.

There are multiple reasons why an individual observation may be in error. For example, an observation may be robust, but *corrupted values* of the observation's value, time, date, or location may occur upon its transmission to the agency responsible for the observation platform. Observations may also have errors due to an improper calibration of the sensor used to collect the observation. Such errors are known as *systematic errors*, as the extent to which the observation is erroneous is not random and thus theoretically can be corrected. By contrast, observations may have errors due to a sensor malfunction. Such errors are known as *random errors*, 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 are 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**: Is the observation out of the range of physically plausible 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. This includes both identifying erroneous observations and not excluding robust observations. In fact, incorrectly identifying a robust observation as erroneous can degrade the initial conditions and thus the subsequent forecast's 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 poor 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 the 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 (Fig. 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 valid at this time (Fig. 3). Consequently, the data assimilation systems used at the time to prepare initial conditions for most major modeling systems rejected this observation as being unrepresentative of either surrounding observations or the first guess for the initial conditions. However, when this observation is assimilated, the resulting initial conditions depart substantially from those provided by operational models (Fig. 4), and these departures influence the subsequent forecast – in this case, resulting in a better snowstorm prediction.
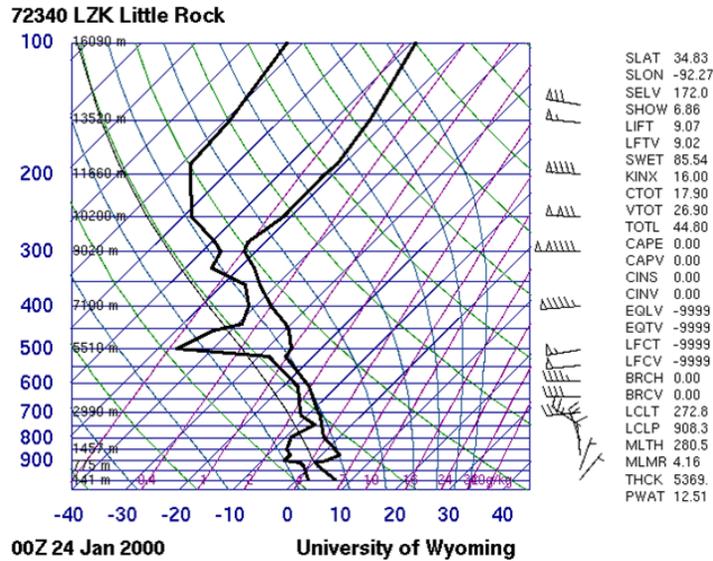
Figure 2. 0000 UTC 24 January 2000 Little Rock, AR sounding. Figure obtained from the University of Wyoming Atmospheric Sounding Archive.
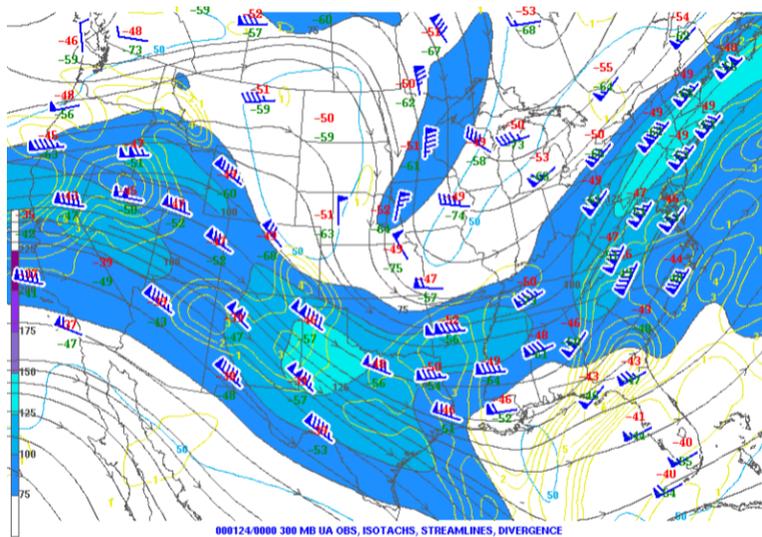

Figure 3. 300 hPa analysis valid at 0000 UTC 24 January 2000. Station plots depict observations of temperature (red, °C), dew point temperature (green, °C), and wind speed and direction (blue; half-barb: 5 kt, full barb: 10 kt, flag: 50 kt). Also depicted is the 0-h RUC analysis of streamlines (black), 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.

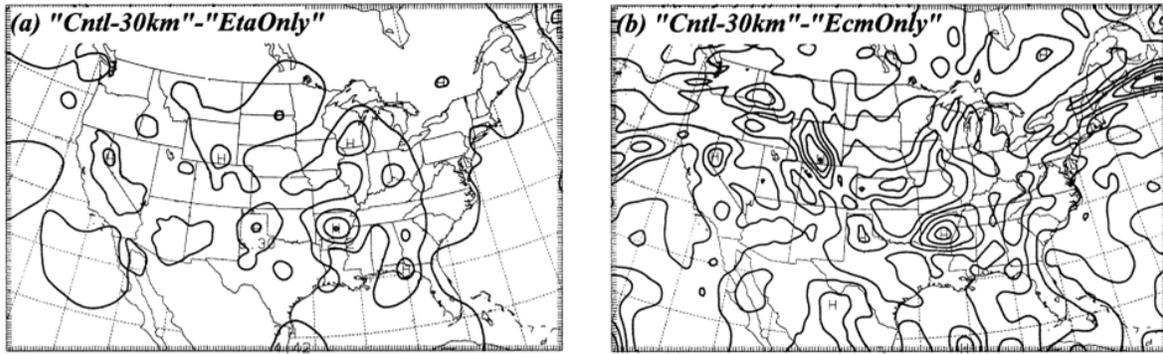**Figure 4**. 300 hPa wind difference magnitude (contoured every 3 m s^{-1}) between an initial analysis that assimilated the Little Rock, AR wind observation and initial analyses from the (a) Eta and (b) ECMWF models that did not assimilate the Little Rock, AR wind observation. Wind difference magnitudes 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 Fig. 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 Fig. 10.

Model Spinup

There are three spatial scales of interest with respect to model initialization: those resolved by the observation network (often the coarsest), the initial conditions, and the model simulation (often the finest). Some means of generating realistic atmospheric variability on scales smaller than those explicitly resolved by the observation network is necessary for the scales resolved by the initial conditions to be identical to those resolved by the model simulation. This is typically achieved by using the short-range forecast of an earlier simulation as the first guess for the initial conditions, which is then subsequently adjusted using the available observations.

How should we handle model variables for which reliable observations do not exist (and thus we cannot directly update the first guess for their values), are extremely computationally expensive to assimilation, and/or for which reliable relationships between the observations and model variables 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. A first guess can provide estimates of these variables, but we generally do not have direct observations of these quantities, and they are generally not well-related to variables for which we do have observations (e.g., the 500 hPa height is not directly related to the rain-water mixing ratio). Consequently, a numerical model simulation may elect to depict the kinematic and mass fields associated with active clouds and precipitation in their initial conditions but not resolve the clouds or precipitation.

With all of this in mind, there are 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 microphysical variables except water vapor are typically missing. 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.

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

- **Hot Start**: Atmospheric variability is present on all scales resolved by the numerical model. Initial conditions for microphysical species' mixing ratios 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. 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 microphysical variables except water vapor, the model must generate (or spin up) the needed fields. The time over which this occurs is known as the spinup period. The length of the spinup 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 leads to a general recommendation to begin a numerical model simulation 6-12 h before the forecast period of greatest interest.

Dynamical imbalances can result as the model spins up atmospheric variability on scales smaller than those resolved by the initial conditions. The model generates spurious inertia-gravity waves, which are associated with rapidly fluctuating surface pressure as they pass a location, to attempt to mitigate these imbalances. The domain-averaged local rate of change of the surface pressure tendency can thus be used as a metric to assess model spinup (Fig. 5). Note, however, that 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.
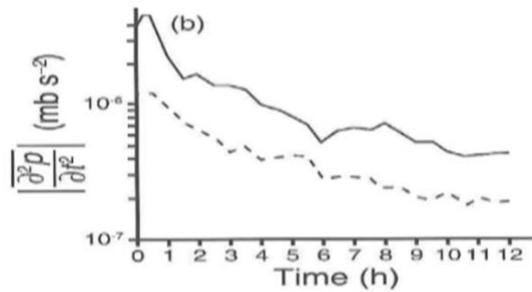
Figure 5. Domain-averaged local rate of change of the surface pressure tendency (hPa s^{-2}) as a function of time for two model simulations, one using well-balanced initial conditions (dashed) and one using poorly balanced initial conditions (solid). Note the logarithmic scale to the y-axis. As the model spin-up period ends, roughly between 6-12 h, the slopes of each curve asymptote to zero. Figure reproduced from Warner (2011), their Fig. 6.15b.

Observation Targeting

Fundamentally, forecast quality is directly proportional to both initial condition quality and model quality. Since first-guess initial conditions are typically provided by the short-term forecast from a previous simulation, the initial conditions' quality can be improved if model error is reduced. Initial-condition quality can also be improved through better using existing observations or deploying new observation platforms, whether they are new routine platforms or temporarily as part of a field program or targeted reconnaissance effort.

Observation targeting describes the processes by which where to optimally site additional observations to provide the largest forecast improvements can be determined. This is most often done in the context of a new observation platform or field campaign. To illustrate the concept of observation targeting, consider the example of numerical forecasts of the track of Hurricane Sandy. Fig. 6 depicts forecast tracks of Sandy from three successive cycles of the twenty-member GFS Ensemble. There is minimal spread in the track forecasts prior to approximately 120 h, after which time the 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 must identify 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 these observations will improve the forecast's skill. However, it should be noted that forecast improvement from assimilating targeted observations is miniscule when averaged over many cases (e.g., Fig. 7). This is likely due to several factors, including imperfect targeting methods and unaddressed model errors. 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.
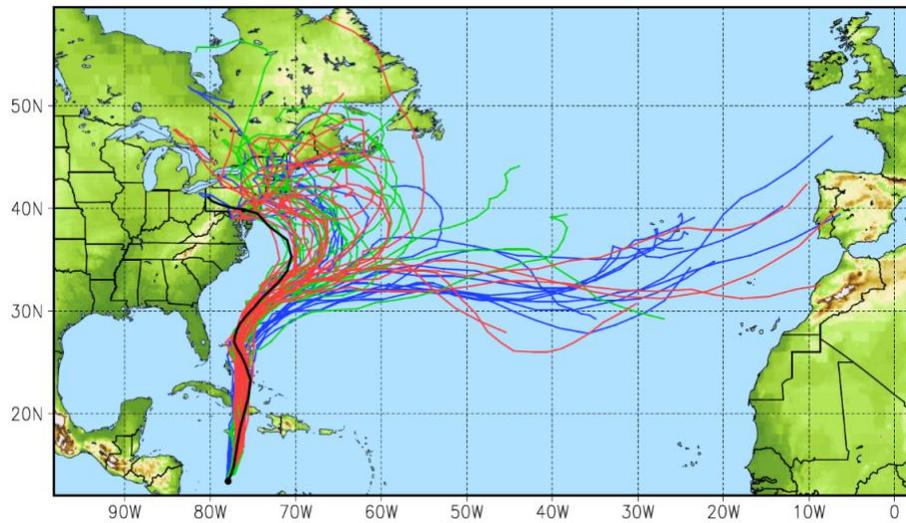

Figure 6. 240-h forecast tracks for Hurricane Sandy (2012) from the GFS Ensembles initialized at 1200 UTC 23 October 2012 (blue), 1800 UTC 23 October 2012 (green), and 0000 UTC 24 October 2012 (red). The analyzed position of Sandy at 1200 UTC 23 October 2012 is denoted by the black square, and the observed track of Sandy through dissipation is given by the black line.
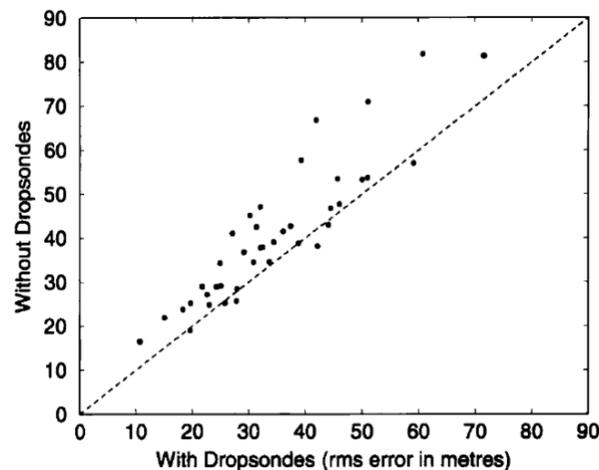

Figure 7. Root-mean squared error (m) of 1000 hPa and 500 hPa geopotential height forecasts at lead times of 30, 36, 42, and 48 h from the FASTEX field experiment. Root-mean squared error is presented for two sets of ECMWF forecasts, one in which no targeted dropsonde observations are assimilated and one in which all targeted dropsonde observations are assimilated. The singular-vector technique was used to target the dropsonde observations, and forecasts are verified only over the targeted regions. Dots to the left of the dashed line indicate a positive impact from assimilating targeted observations (e.g., larger error without targeted dropsondes). Figure reproduced from Montani et al. (1999, *Quart. J. Roy. Meteor. Soc.*), their Fig. 8.

Ensemble Variance/Spread-Based Methods

This method assumes that forecast error growth is largest where the initial condition uncertainty and thus the potential for initial-condition errors are largest. For an ensemble of model initial conditions, uncertainty can be quantified by the ensemble spread or variance. Initial-condition uncertainty is typically largest where there are few observations and where sharp gradients in model variables exist. 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 is large can theoretically help constrain model error growth and improve forecast quality. This is true in a general sense, but how can the best locations for these observations be determined? 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 are multiple quasi-objective ensemble-based methods by which the locations and types of targeted observations that may exert the greatest positive impact on the subsequent forecast can be identified.

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

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

This represents the change in the forecast metric as one changes the model field 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 Fig. 8.
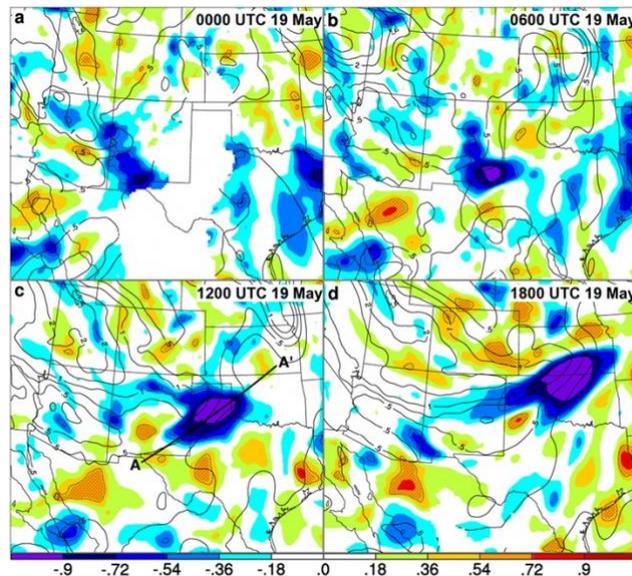
Figure 8. Ensemble sensitivity metric (shaded; mm change per one standard deviation change in the model variable **x**) relating changes in the forecast metric **J** to a model variable **x**. Here, **J** is the area-averaged forecast precipitation over central Oklahoma between 22-25 h into the forecast and **x** is the 315-K isentropic potential vorticity at (a) 0 h, (b) 6 h, (c) 12 h, and (d) 18 h. Positive values indicate that a one standard deviation increase in **x** is associated with a positive change of the specified amount in **J**, whereas negative values indicate that a one standard deviation increase in **x** is associated with a 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 Fig. 5.

Applying this method to targeted observations, consider Fig. 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 in the 12-h forecast. This suggests that targeting observations of variables that are related to isentropic potential vorticity in the 12-h forecast over the Texas Panhandle, followed by assimilating them into the initial conditions for a subsequent simulation starting at that time, could result in improved (or at least more certain/less variable) precipitation forecasts at 10-13 h (equivalent to the earlier 22-25 h forecast period) over central Oklahoma.

A related method uses empirical orthogonal functions (EOFs) to identify modes of variability at a given forecast time from an ensemble of numerical simulations. Once the leading mode(s) of variability have been identified, linear correlation is used to connect them to the variability of a given model variable at the same or earlier forecast times. Figs. 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.
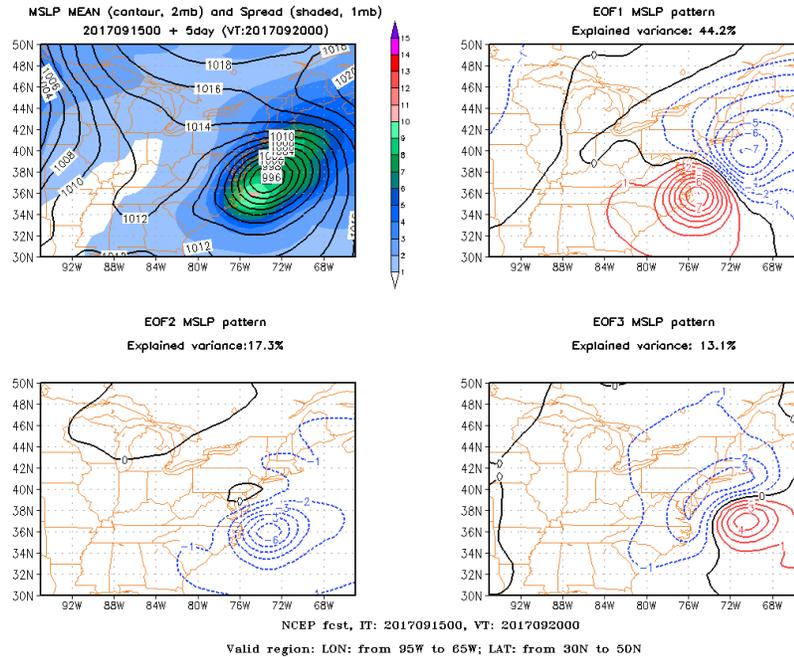
Figure 9. (upper left) 120-h GFS Ensemble-mean forecast mean sea level pressure (contoured, hPa) and standard deviation (shaded, hPa) from the 0000 UTC 15 September 2017 run of the GFS Ensemble. (upper right) The leading EOF of the ensemble's 120-h sea level pressure forecasts, explaining nearly half of the total variance, is primarily characterized by uncertainty in the along-coast position of tropical cyclone Jose. (lower left) The second EOF of the ensemble's 120-h sea level pressure forecasts is primarily characterized by uncertainty in Jose's intensity. (lower right) The third EOF of the ensemble's 120-h sea level pressure forecasts is characterized by uncertainty primarily in Jose's proximity to the United States coast. Figure obtained 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, particularly in the context of ensemble numerical simulations.
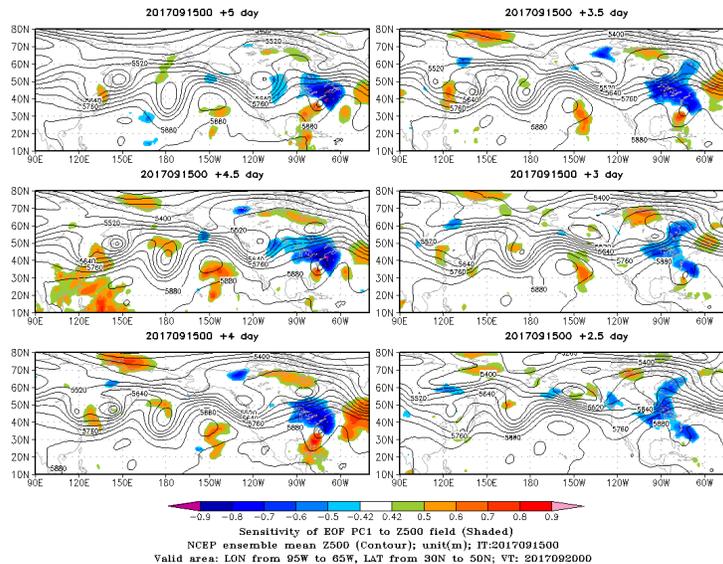
Figure 10. Ensemble sensitivity metric, here representing the linear correlation between the leading EOF pattern in the upper-right panel of Fig. 9 and the GFS-Ensemble–forecast 500 hPa geopotential height field at forecast lead times of 2.5 to 5 days (or 0 to 2.5 days prior to the valid time of the EOF analysis depicted in Fig. 9). The upper-left panel indicates that decreasing the 500 hPa geopotential height in eastern North America in the five-day forecast (cool colors) is associated with positive EOF1, indicating a further northeast position for tropical cyclone Jose at the 120-h forecast time. Tracing this back to earlier forecast times, the lower-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 for Jose at 120-h. Figure obtained 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, representing the inverse of the linearized version of the numerical model, identifies the quantitative impact of small, arbitrary perturbations to the initial conditions on the chosen forecast measure. These arbitrary perturbations may be random or may be obtained by some other means (e.g., from the difference in earlier non-linear and linear model forecasts valid at the new forecast's initial time). An idealized schematic of the conceptual underpinnings and operation of adjoint-based methods is provided in Fig. 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 on the subsequent forecast.

<u>Singular-Vector Technique</u>

The singular-vector technique provides a method to quantify how initial condition uncertainty or error propagates forward in time. It uses a linear version of the numerical model to identify the structures whose amplitudes grow most rapidly over a short but physically relevant time frame (e.g., 24 h). 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 Fig. 7.