

Ensemble Kalman Filters Bring Weather Models Up to Date

By Dana Mackenzie

‘Twas the morning after Christmas, and in Santa’s wake an unforeseen visitor descended upon Europe. On December 26, 1999, a storm called Lothar barrelled out of the North Atlantic with hurricane-force winds. In Versailles, at the French royal palace, gusts of more than 100 miles per hour toppled ten thousand trees, including a Corsican pine that had been planted by Napoleon. Across Europe, three and a half million people lost electricity, some for as long as 20 days. More than 100 people and 400 million trees died as a result of the storm, which finally dissipated over Poland.

The storm caught most European weather services with their pants down. Britain, Germany, and Switzerland predicted Lothar only 18 hours before landfall. The French weather service, Météo-France, did better, issuing storm warnings 30 hours in advance—but even they reduced the forecasted wind speeds from 90 mph to 70 mph because their computer models were so far out of line with the British ones.

Why did European weather forecasters, for the most part, strike out on the storm of the century? Would better satellites or more powerful computers have helped? Surprisingly, the answer seems to be no. The most critical weakness of the weather models came in a little-known area called “data assimilation.” As meteorologist Per Uden, of the Swedish Meteorological and Hydrological Institute, wrote a month later, “The forecast problems [were] most likely due to data assimilation difficulties only.”

Data assimilation is the glue that binds raw data with the physics-based equations that go into computer weather models. These equations, like all differential equations, require “initial values” to be fed into them. If you tell them the temperature, velocity, and pressure of the air in every cubic inch of the Earth’s atmosphere, the equations can predict how that state will evolve. The problem is that *nobody knows* the correct initial conditions. Observations from balloons, buoys, and satellites provide some information—but only for specific places and times. And instruments can always break or malfunction. The weather service’s own previous forecasts also offer an abundance of information, but some of it will be outdated or incorrect.

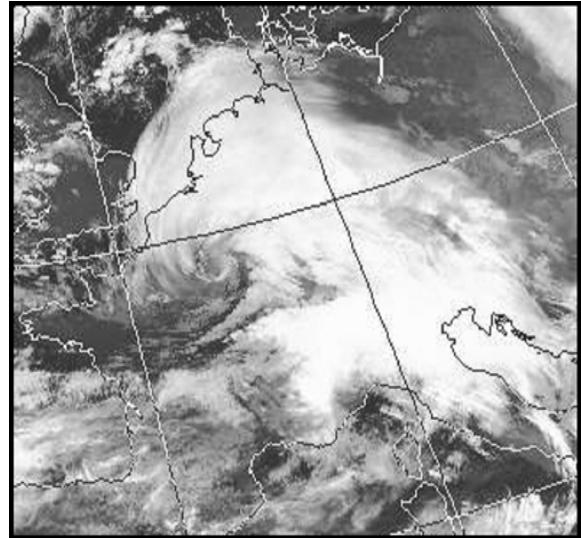
“If either the forecast or the measurement is terrible, we should ignore it,” says Dennis McLaughlin of the Massachusetts Institute of Technology. “But in most cases, each contains some information.” The trick is to blend the two sources, integrating new data into the model without tossing out the still-valuable information embodied in the old predictions. Atmospheric and oceanic scientists are constantly looking for better ways to do this. Some of the most promising schemes are actually new variations on an old idea: Kalman filters, a data assimilation method that has been used for years in inertial guidance systems for airplanes and spacecraft. Perhaps their most famous application to date was guiding the Apollo spacecraft to the Moon.

From Yesterday’s Rocket Science to Today’s Weather

Though one is notoriously precise and the other notoriously imprecise, rocket science and weather science actually have a lot in common. Rocket engineers have to reconcile a spacecraft’s current sensor readings with differential equations that tell them where it ought to be, based on their (probably imperfect) knowledge of its past. When the designers of Apollo needed a way to blend these two sources of information, they found it in Kalman filters.

The idea behind Kalman filters can be traced back to a simple calculus problem. If you have two measurements, x_1 and x_2 , of an unknown variable x , what combination of x_1 and x_2 gives you the best estimate of x ? The answer depends on how much uncertainty you expect in each of the measurements. Statisticians usually measure uncertainty with the variances, σ_1^2 and σ_2^2 . It is not too hard to show, then, that the combination of x_1 and x_2 that gives the least variance is

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2.$$

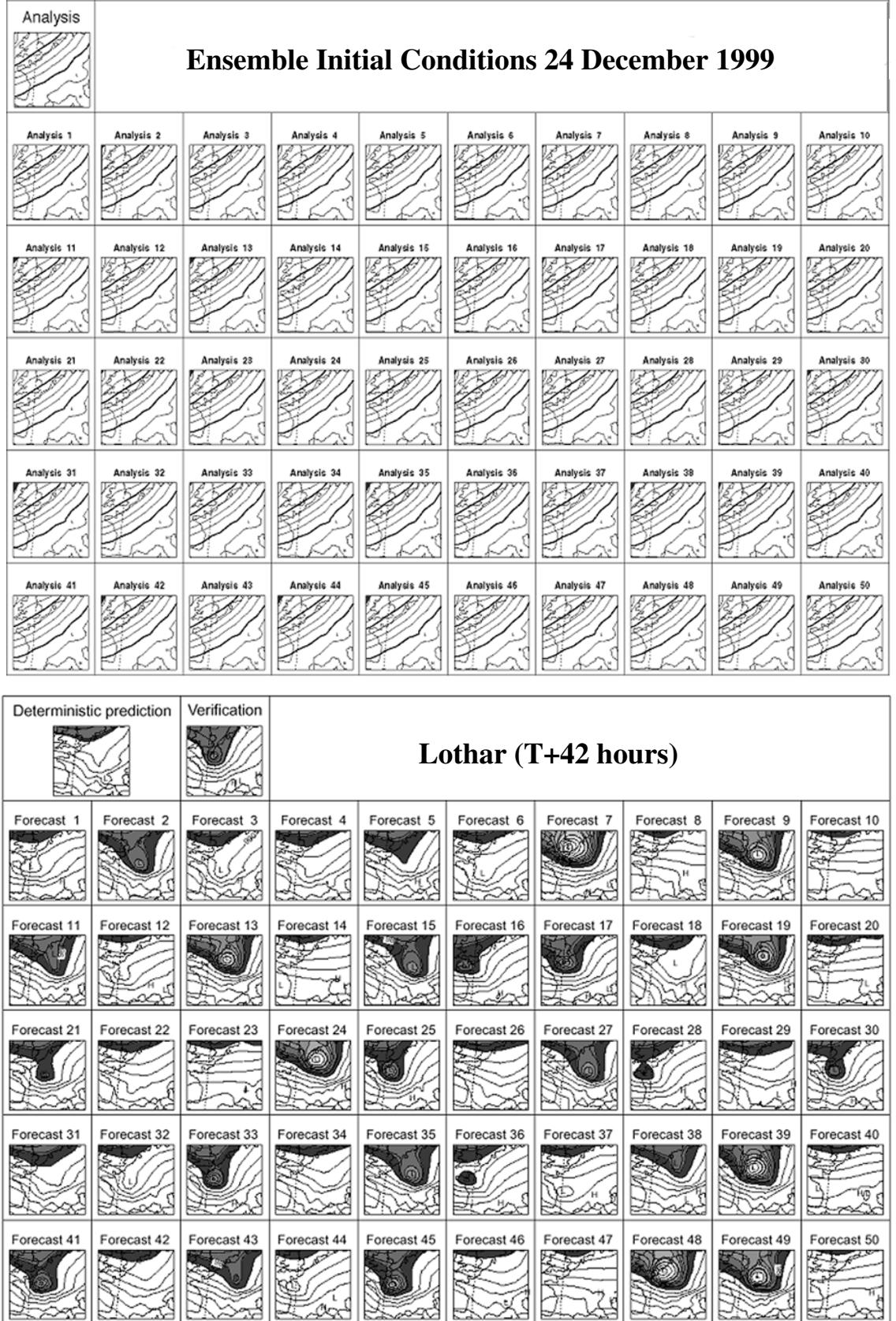


Radar image obtained during the storm Lothar, which, with no warning to speak of, wreaked havoc on Europe in December 1999, causing more than a hundred deaths, destroying some 400 million trees, and leaving millions without water or electricity.

This equation is perfectly consistent with McLaughlin's comments: If one of the variances (σ_1^2 , say) is infinite, then you should use the other variable (x_1) as your estimate. But if both variances are finite, the formula tells you the correct weights to assign them.

In a Kalman filter, the first measurement, x_1 , comes from sensor data, and the second "measurement," x_2 , is really not a measurement at all: It is your last forecast of the spacecraft's trajectory, or the state of the atmosphere. In real applications, these measurements are not numbers, but vectors. In the case of the Apollo spacecraft, the vectors had a few tens of components. The state of the atmosphere, on the other hand, is represented by a vector with tens of *millions* of components. In the top-of-the-line model of the European Center for Medium-Range Weather Forecasts (ECMWF), the atmosphere over Europe and the North Atlantic is described by six pieces of data at each point in a $500 \times 500 \times 50$ grid, or 75 million numbers in all.

For forecasts of a vector quantity, the scalar variances, σ_1^2 and σ_2^2 , are replaced by two covariance matrices, which



Slight changes in initial data result in dramatically different forecasts. Top, weather forecasters' best estimate of atmospheric conditions (contours represent isobars of pressure) on December 24 ("Analysis"), along with 50 slight perturbations. Bottom, the ECMWF model's best-guess forecast ("Deterministic prediction"), along with the 50 ensemble members and a map of the conditions actually observed ("Verification"), 42 hours later; shading indicates low pressure. Ensemble methods, as results like these show, can detect in advance the possibility of weather that differs markedly from the best-guess forecast. Courtesy of the European Center for Medium-Range Weather Forecasts.

represent the uncertainty in the measurements and the uncertainty in the forecasting model, respectively. (They also take into account any correlations between uncertainties. If, for example, the temperature in my town is higher than expected, then the temperature 10 miles down the road is likely to be higher than forecast as well—the two errors are positively correlated.) As the pioneering control theorist Rudolph Kalman discovered in 1960, when the weights attached to the data and the previous forecast are chosen to minimize the variance of the new forecast, they automatically optimize the weights attached to *all previous data* as well. Mathematically, Kalman's procedure is a standard least-squares estimate, but computed in a unique, sequential way. When new data arrives, the rocket controller (or weather forecaster) does not need to haul out all the old sensor data in order to update the rocket's trajectory (or the weather forecast). All that's needed is the most recent forecast, which has the previous sensor data built into it.

Kalman filters are about as sweet a method as you can find in applied mathematics. They are provably optimal if the differential equations describing the system are linear, and their sequential structure makes them ideal for real-time computation. But they did not catch on right away in meteorology. The first meteorologist to advocate them was Michael Ghil of the University of California at Los Angeles, in a lecture at an ECMWF workshop in 1980. "I was basically laughed off the stage," Ghil says, perhaps with a touch of exaggeration. "But now you have 200 or 300 people working on them."

When adapting Kalman filters to weather forecasting, meteorologists faced two immense problems, according to Ghil: "Size and nonlinearity." The size problem needs little explanation. If the state vector has 75 million components, then the covariance matrices are 75-million-by-75-million arrays—too much data for even the most powerful supercomputer to cope with in real time.

Nonlinearity is more subtle but equally deadly. To make the Kalman filter work, the user needs a statistical model for the observational errors and for the errors in the computer model (which "propagate" the observational errors through time). It is easy to believe that measurement errors will have a Gaussian distribution—that is, they will follow a bell-shaped curve, with most of the errors close to zero and only a smattering of larger errors (the tails of the curve). But the errors propagated by the model will follow a bell-shaped curve only if the differential equations are *linear*—if, in other words, small changes to the initial conditions produce only small effects. But this is emphatically not the case in the real atmosphere or ocean, where propagated errors can follow a very different distribution, with, for example, two humps instead of one.

Nothing could illustrate this point better than the storm Lothar. When 50 different versions of the December 24 weather map—virtually indistinguishable to the naked eye—are plugged into the ECMWF weather model, about half of them forecast no storm on December 26 (see illustration on this page). (Among those forecasting no storm was ECMWF's best estimate of the atmosphere's state on December 24.) About 40% of them forecast a whopper of a storm. Between the two groups there is almost no middle ground. Clearly, the atmosphere on December 24 was near a tipping point—it could have gone either of two ways, and it tipped toward a history-making storm. Perhaps something as small as the turbulence of Santa's sleigh made the difference.

A simple, though somewhat time-consuming, solution is to run 50 versions of every forecast, watching for possible outcomes that are quite different from the favored prediction. This approach is called "ensemble forecasting," and most weather services are experimenting with it. But it is not really the most principled approach, because it doesn't change the data assimilation system itself. More radical is a solution called an *ensemble Kalman filter*—which may also solve the size problem that has kept Kalman filters from becoming standard.

The Power of the Ensemble

Ten years ago, Barry Cipra reported in *SIAM News* (August 1993) that, according to control theorists, the Kalman filter "doesn't handle any serious nonlinearity" and the proposed remedies were "just too complicated." Cipra then added, prophetically: "But if history teaches any lesson, it's that everything could change with the appearance of one or two new papers."

The paper he foresaw in his crystal ball may have appeared the very next year, published in the *Journal of Geophysical Research* by a Norwegian oceanographer, Geir Evensen. (A Canadian meteorologist named Peter Houtekamer was not far behind him.) Evensen's idea was to use an ensemble to update not only the forecast, but also the covariance matrices. The advantage, Evensen wrote in a more recent paper, was this: "Instead of storing a full covariance matrix, we can represent the same error statistics using an appropriate ensemble of model states." That is, if the ensemble has 50 forecasts, then all the essential information in the 75-million-by-75-million covariance matrix can be obtained from a 50-by-50 matrix, the covariances between the ensemble members.

If this sounds a little too good to be true, bear in mind that the ensemble only approximates the error statistics. Evensen's "appropriate" glosses over the possibility that no small ensemble provides a good enough approximation. However, Evensen did prove rigorously that his filter gets closer to the standard Kalman filter as the ensemble gets larger. The practical question is, how big is big enough? Does a 50-member ensemble contain enough information from the full covariance matrix to provide a good forecast?

The answer will come only through operating experience. So far, according to Houtekamer, tests of simpler model systems at the Meteorological Service of Canada (MSC) have been encouraging. Next year, the MSC will become the first national weather service to use ensemble Kalman filters in actual forecasts. "At first it will be invisible," says Houtekamer, "except that forecasts from day 3 to day 10 may become a bit better." But the real value of the technique should lie in more realistic estimates of forecasting errors, and better early warning of low-probability events like Lothar. As for other countries, Houtekamer says, "As soon as we have a convincing demonstration here, the (U.S.) National Weather Service will probably do the same."

Meanwhile, meteorologists and oceanographers continue to explore other ways of simplifying and improving Kalman filters. Some are looking for ways to cut the number of variables that describe the state of the atmosphere down from millions to just a few that really matter. "Most people believe the ocean and the atmosphere are governed by a much smaller subset of variables," says Bob Miller, an oceanographer at Oregon State University. "There might be a curved surface in a million-dimensional space

that carries the essential behavior. If we could lay our hands on that, we would be ahead of the game.”

As an example, Miller cites the Kuroshio, a Gulf Stream-like current that flows off the coast of Japan. The Kuroshio switches back and forth between two distinct modes, each of which is stable for years at a time. In one mode, the current hugs the coast; in the other, it veers off the coast and comes back, creating a large offshore eddy. “This is typical nonlinear behavior,” Miller says. “There are two distinct equilibrium points. Once you draw that picture qualitatively, then mathematical theorems tell you that this is a fundamentally two-dimensional system.”

In this much simpler system, oceanographers could easily apply a standard Kalman filter. Miller argues, though, that a better choice would be a “nonlinear filter” based on maximizing the likelihood of a correct prediction rather than on minimizing the variance. “If you go out in the evening with a shotgun and see two flocks of ducks, and you aim for the least-squares location, you will go home hungry,” Miller says. The maximum-likelihood approach is to aim for one of the flocks instead.

For now, scientists are still searching for those two dimensions in a million-dimensional haystack. It is probably too early to predict whether ensemble filters or dimension-reducing methods will win in the end; perhaps each one will find its own niche. But either way, to (mis)quote an old *Star Trek* episode: “Resistance is futile. Your data will be assimilated.”

Dana Mackenzie writes from Santa Cruz, California.