19A.6 THE DEVIL IS IN THE DETAILS: PREPARING RADAR INFORMATION FOR ITS PROPER ASSIMILATION

Frédéric Fabry and Alamelu Kilambi
McGill University, Montreal, Quebec, Canada

1. RECESS IS OVER

Dynamical and thermodynamical retrievals at the mesoscale have been demonstrated for 35 years and data assimilation for about 25 (see Sun and Wilson 2003 for details). For too long we have been taking shortcuts in the name of expediency to present proof-of-concepts via case studies. We should now clean up our act. To realize what must be done, we need to revisit the bases on which assimilation rests and reflect on challenges specific to mesoscale data assimilation.

1.1 Data Assimilation Is Information Optimization

The formalism of data assimilation is designed to optimize the use of the information provided by instruments now (observations), by instruments in the past (analysis), and by model and physical constraints (balance equations, etc). The success of data assimilation therefore relies on specifying as accurately as possible, among others, 1) the quantities being measured, and 2) the correlation structure of the error on those quantities. In other words, for each piece of information we are trying to combine together with others, we need to know as well as possible a) what it is that we are actually measuring or evaluating, b) with what error we are evaluating it, and c) what is the relationship or the correlation between this error and the errors of all of the other pieces of information we are trying to combine.

A simple example: If you have two measurements of the same phenomenon, one called *a* known with an uncertainty σ_a and one called *b* known with an uncertainty σ_b, their sum (*a*+*b*) will be known with an uncertainty σ_{a+b} such that $\sigma_{a+b}^2 = \sigma_a^2 + \sigma_b^2 + 2\,\mathrm{cov}(a,b)$. Here, cov(*a*,*b*) is the covariance between *a* and *b*, describing how are the error of measurement *a* and the error of measurement *b* linked quantitatively. To compute σ_{a+b} accurately, you need σ_a, σ_b, and cov(*a*,*b*).

I claim that a) we are not properly describing what quantities we are measuring by radar (or, in assimilation speak, that the observational operator *H* is too often incorrect); b) we have not done a good job of quantifying the measurement errors; and, c) we have essentially ignored to take into account how dependent the errors are among each other (or, in assimilation speak, that the covariance matrix **R** used is too simplistic). Getting *H* and **R** is difficult; but success relies on knowing them.

Corresponding author address: Frédéric Fabry, Dept. of Atmospheric and Oceanic Sciences, McGill University, Montreal, QC, H3A 2K6 Canada; frederic.fabry@mcgill.ca

1.2 Mesoscale-Specific Challenges

One of the reasons why we need to get *H* and **R** as well as possible is because sources of data at the mesoscale are limited. If we consider a 200-km by 200-km area, we may get a few soundings from balloons, commercial airplanes, and satellite-borne GPS receivers every hour providing a few tens of measurements each. There may be a few surface stations providing 5-min data and maybe a few tens providing hourly information. And a geostationary satellite may provide 50×50 measurements (4-km resolution) four times an hour per channel. In comparison, radar provides twelve times an hour reflectivity data over 360(°)×120(km)×20(angles) data points, most of them "no echoes" but it is valuable information; it also provides Doppler information at the same resolution as reflectivity but only where there is some echo. The table below summarizes these results.

Constraints per hour over 200×200 km	
Soundings (balloons, planes, GPS)	≈ 100
Surface observations	≈ 50–500
Geostationary satellite (per channel)	≈ 10,000
Radial velocity, assuming 10% coverage	≈ 1,000,000
Reflectivity (at 1-km range resolution)	≈ 10,000,000

TABLE 1: Number of measurements per hour of any atmospheric parameter that operational observing systems can provide over a 200×200 km area.

What Table 1 illustrates is how much weight radar data can have in an assimilation system. This is not a surprise: the availability of radar data in storms is, after all, the reason why we want to use them for data assimilation in the first place. What the table also illustrates is why radar data assimilation must be done rigorously right: there is very little that can invalidate radar data if they are wrong or wrongly assimilated.

1.3 Task Required

It is hence crucial that radar data assimilation is done as carefully as possible to both provide the best information possible to the model and give a chance to other data sources to effectively complement radar data. This, in turn, sets three requirements for success:

1) The data must be clean and free of bias, much more so than has been the case until now;
2) The observation operators must be accurate;
3) The correlation structure of the error must be well defined and easy to use.

These three requirements drive this project. We present below where we are at in this task.

2. DATA AND ITS CLEANING FOR ASSIMILATION

2.1 Rethinking Data Cleaning

Radar meteorologists have spent considerable effort and gathered much experience in cleaning radar data. But although much progress has been done, there always remains a small fraction of data that cannot be cleaned properly, whether the difficulty comes in the velocity dealiasing, in the clutter removal, or in the target identification done with dual-polarization information. These data points must first be recognized, and then a decision has to be taken as to what to do with them.

Until now, that decision has been influenced by two facts: people in general, and forecasters in particular, can recognize and tolerate some bad data if they are trained to do so; however, data gaps are to be avoided as much as possible as they make humans very nervous. With data assimilation, the situation is reversed: data assimilation systems cannot tolerate unrecognized bad data; in parallel, large data gaps are accepted without many problems. A conclusion of this observation is that the way data cleaning must be done for data assimilation systems may have to be different than what has been done for forecasters so far.

In particular, data assimilation systems cannot handle biased data. Any source of data bias will cause bias in the resulting analysis, even if it is localized in space. For example, poorly removed clutter causes problems when assimilating radial velocity, as these cause errors in the obtained winds that persist in time and may ultimately falsely trigger instabilities (Fig. 1).

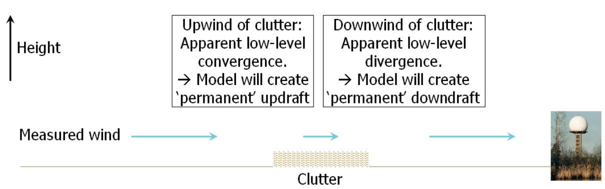


FIG. 1: Example of the effect of incompletely cleaned clutter on winds and on the data assimilation system. Weak clutter (below weather signal strength) can cause great difficulties to assimilation systems. Too aggressive clutter filters will also lead to convergence biases.

An extreme solution, though not necessarily a bad one, is to reject all data that has a chance of being corrupted. Indeed, with radar data, data corruption can come either from blunders, such as a badly dealiased velocity, or from sources of contamination with long correlations in time and space: ground clutter always reducing velocities at the same locations, second trip echoes enhancing reflectivities over large areas and bringing Doppler velocities from a higher altitude to lower ones, blockage, etc. Any of those sources of contamination will cause severe problems to data assimilation systems that expect data to be unbiased and whose errors have specific (and specified) behavior in space and time. Such rejected data must be classified as "no information" regions, not as "no echo" regions.

2.2 Focus on Clutter Filtering

All that being written, if we can find ways to clean the data in an unbiased way, then these should remain favored over throwing data away. In this spirit, we have worked on ways to improve the filtering of ground clutter in our system.

We continued the work of Fabry and Gadoury (2009) inspired by Moisseev and Chandrasekhar (2009): We first decompose the returns from each range gate into their spectral components using a Fourier transform on both horizontally and vertically polarized signals; then, we try to identify whether the returns from each Doppler component is dominated by clutter, weather, or noise; finally, we suppress the clutter components, fit a Gaussian curve through the remaining noise and weather components, and replace the suppressed clutter components with those from the curve fit. Work is in progress (the curve fitting is still being worked on), but results are promising (Fig. 2). They will be described more fully in a latter conference.

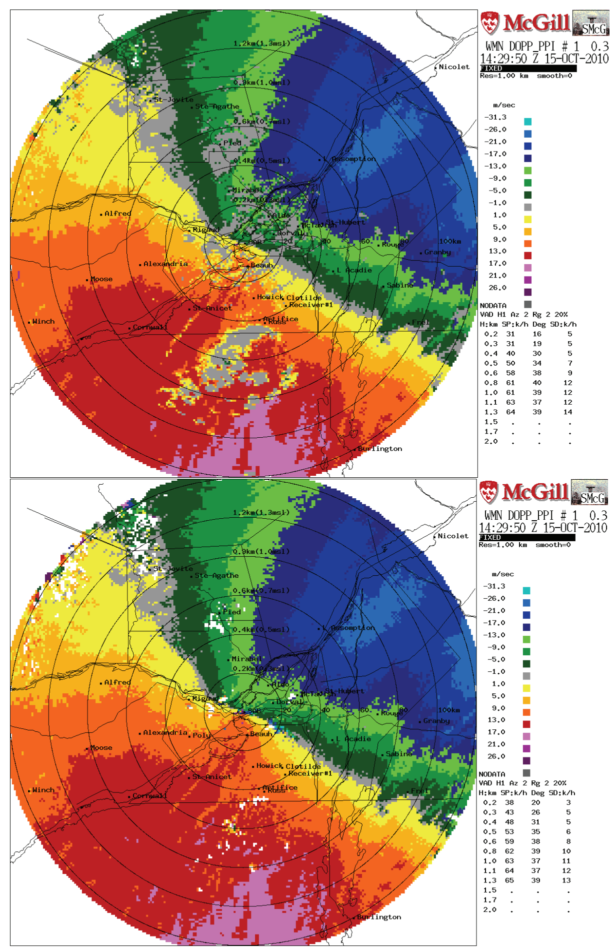


FIG. 2: 0.3° Doppler PPI (above) before and (below) after clutter filtering. Some contaminated (or believed to be contaminated) regions to the south and the north-north-west could not be recovered and are now displayed in white (no Doppler data).

2.3 What Radar Information Should Be Assimilated?

Data assimilation should work best if the data being measured are what is being provided. The idea is that since data assimilation combines information from all sources, it is best able to handle all the contradictions or the errors in the data and their interpretation. Hence, we asked ourselves: what quantities do radars really measure?

In parallel, the measurements from radars are somewhat unconventional, in particular that of reflectivity. A thermometer will report air temperature all the time unless the instrument fails. But reflectivity measurements can be interpreted one way if weather echoes are strong enough to be detected and if they dominate the return, another way if non-weather echoes dominate, and perhaps even another if echoes are too weak to be detected. This triggered another question: what does the information provided really mean?

We are still thinking about those two issues, and hence no definitive answers will be provided. But here are some of the avenues considered.

2.3.1 Quantities to be assimilated

If we limit ourselves to Doppler radars, an answer to the question "what does radar really measure" may well be: path- and radome-attenuated effective reflectivity factor ($\exp[-2\tau]Z_e$, where τ is the optical thickness of the medium between the radar and the target), and aliased Doppler velocity, the above possibly range-folded. Everything else (Z, true Doppler velocity) is derived based on assumptions or algorithms that can be at fault. In the long run, this may be the best approach: one could imagine using model parameters such as retrieved rainfall and temperature to compute rainfall and gaseous attenuation and correct for them, or using past wind information to figure out whether the current observations are aliased or not. But it is not clear at this time if assimilation systems have the required CPU to resolve all the ambiguities in the iterative fashion used in the minimization of the cost function J that evaluates the mismatch between current observations and the analysis coming from the current model run.

2.3.2 What reflectivity information should be provided?

The question of what information is provided from reflectivity is an interesting one. The best way to illustrate this point is to play the following scenario game: What information can we retrieve from each of the following scenario? Think before reading each entry.

- Measurement of echo from hydrometeors: This is the simplest: if the radar says that reflectivities of 15 dBZ are being measured, and if the echo comes from clouds or precipitation (not from turbulence), then this is what should be assimilated. The associated Doppler data is also to be taken as is barring any aliasing.
- Measurement of no echo: This is information, but it is not exactly "0 reflectivity". It is instead that the (attenuated) reflectivity is below the detectable signal of the given radar system at that particular elevation and range, possibly affected by beam blockage. So what should be ideally assimilated is "*no attenuated echo above x dB*", *x* being the minimum detectable signal at that specific location. Data assimilation systems should be modified to accept this more vague information. In Doppler, however, there will be no velocity data and this is truly "no information".
- Measurement of echo from clear air (turbulence or biological) returns: The reflectivity is not from cloud or precipitation echoes. So what does this tell us about the reflectivity from cloud or precipitation echoes? Not much, except that it has to be weaker than that from the clear air echo. In short, we are back to the "*no attenuated echo above x dB*", but this time *x* is the strength of the clear-air echo, or a few dBs below that, not the minimum detectable signal. The Doppler data may be more complicated: while turbulence echoes should move at the speed of the wind (at least that of the sub-region where the clear-air echo arises), it is less clear for biological returns: birds move, and so do many insects especially in late spring and early fall. The first author (Fabry) may have a contrary opinion to that of the consensus of the radar community, but it is not clear to me that insects are as good targets as advertised to measure winds.
- Measurement of ground or sea clutter: This case is actually similar to the above: "*no attenuated echo above x dB*", with *x* being the minimum weather echo one should be capable of retrieving in the presence of the clutter observed. Sometimes we may be stuck telling the data assimilation system something as stupid as "there is no echo above 65 dBZ", but this is what the radar is telling us, no more, no less. Doppler data should not be used and be given a value of "no information".

What this exercise illustrates is that we should both think about what information on cloud and precipitation do we get from each scenario, and do our best to both convey that information to the data assimilation system and make sure that the data assimilation system makes the best use of it.

3. WORKING ON OBSERVATION OPERATORS

Data assimilation gathers information from newly-collected data by minimizing the difference between current observations and simulated observations using the model information. Observation operators (often referred as H) are equations, or lines of code, that allow a data assimilation system to simulate observations from a specific instrument from model variables. A radar observation operator must hence use model variables such as mixing ratios of precipitation, 3-D winds, etc., from the model grid to simulate what reflectivity and Doppler observations should be expected. If a data assimilation system can accurately simulate what a radar would observe under given known atmospheric conditions, then it can use the true radar data to tune the model variables until simulated observations converge towards the true ones. If it cannot simulate well radar observations, especially if the simulated observations are biased, then providing correct radar observations will result in the wrong information being assimilated. Accurate and unbiased observation operators are hence necessary to the success of data assimilation systems.

Doppler velocity is easier to simulate, less affected by bias, and less ambiguously tied to model variables than reflectivity. This partly explains why it is more often assimilated than reflectivity. But this apparent simplicity is deceptive: I have yet to see a publication where the observation operator of Doppler velocity is correct. Some have used the simplistic $V_r(r)=(ux+vy)/r$; others have used more sophisticated observational operators. In most case, the result is that simulated Doppler velocities are biased compared to what should be simulated if the physics of the observation process were correctly taken into account.

We worked on a derivation of the best possible observation operator for Doppler velocity (Fabry 2010, 2012) and got the following beast:

Phase shift to velocity

$$V_{r-meas} = \frac{V_{nyq}}{\pi}\,\mathrm{atan2}(B,A) \quad \text{where } \mathrm{atan2}(B,A) = \mathrm{sgn}(B)[1-\mathrm{sgn}(A)]\frac{\pi}{2} + \tan^{-1}(B/A)\ (\text{if } A \neq 0),\ \text{or } \mathrm{sgn}(B)\frac{\pi}{2}\ (\text{if } A=0, B\neq 0);$$

$$A = \iiint_{beam_i} \cos\left[V_r(r,\theta,\phi)\frac{\pi}{V_{nyq}}\right]\frac{Z_e(r,\theta,\phi)\exp(-2\tau)}{r^2}\left\{\sum_{i=1}^{N} G^2[(\theta-\theta_i)\cos(\phi'),\phi-\phi_i]\right\}d\theta\cos(\phi')d\phi\sum_{l=0}^{L-1}\left[\sum_{j=1}^{M} W\left(r-r_j-\frac{lct_{prf}}{2n}\right)\right]dr;$$

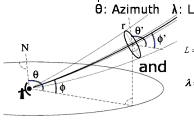
$$B = \iiint_{beam_i} \sin\left[V_r(r,\theta,\phi)\frac{\pi}{V_{nyq}}\right]\frac{Z_e(r,\theta,\phi)\exp(-2\tau)}{r^2}\left\{\sum_{i=1}^{N} G^2[(\theta-\theta_i)\cos(\phi'),\phi-\phi_i]\right\}d\theta\cos(\phi')d\phi\sum_{l=0}^{L-1}\left[\sum_{j=1}^{M} W\left(r-r_j-\frac{lct_{prf}}{2n}\right)\right]dr.$$

Point velocities to phase shift | Power at antenna | Antenna gain for each pulse | Multiple trips | Range function of each gate

with $V_r(r,\theta,\phi) = u(L,\lambda,z)\cos(\phi')\sin(\theta') + v(L,\lambda,z)\cos(\phi')\cos(\theta') + [w(L,\lambda,z) - V_T(L,\lambda,z)]\sin(\phi').$

E-W wind | N-S wind | Vertical wind + fall speed

ϕ: Elevation L: Latitude ϕ′: Elevation angle of beam w.r.t. ground at the target
θ: Azimuth λ: Longitude θ′: Angle w.r.t. local North at the target



and

$$L = 90 - \tan^{-1}\left(\frac{\sqrt{1-c_a^2}}{c_a}\right) \qquad c_a \approx \cos\left(\frac{r\cos\phi}{a}\right)\cos(90-L) + \cos\theta\sin\left(\frac{r\cos\phi}{a}\right)\sin(90-L)$$

$$\lambda = \lambda_{rad} + \tan^{-1}\left(\frac{c_b}{\sqrt{1-c_b^2}}\right) \qquad c_b \approx \sin\theta\sin\left(\frac{r\cos\phi}{a}\right)\Big/\sqrt{1-c_a^2} \qquad \theta' = \tan^{-1}\left[\left(\frac{d\lambda}{dL}\right)_r \cos L\right]$$

Needless to say, this cannot be easily used in data assimilation systems, but it can serve as a good benchmark to test which simplifications to the observation operator leads to major errors, and which do not. For example, even if we use the accurate equation for radial velocity at the center of the beam (the $V_r(r,\theta,\phi)$ in the middle of the beast above), and we test it against the fuller equation ($V_{r\text{-meas}}$), we get errors with long correlation distances that will also lead to systematic errors in the wind retrieved (Fig. 3).

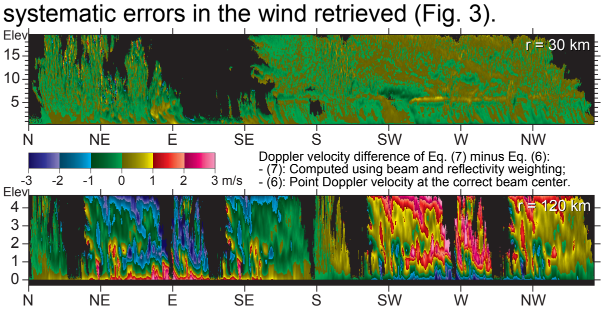


FIG. 3: Difference in the computed Doppler velocity between $V_{r\text{-meas}}$ and $V_r(r,\theta,\phi)$ at 30 km (top) and 120 km (bottom) range. Errors 1) are significant, 2) have long correlation distances, and 3) are anti-correlated 180° apart like in a VAD → Wind error.

Errors with long correlation distances are bad news: since errors are similar in sign and magnitude over large regions, combining multiple data points will not cause the error to diminish significantly. This considerably reduces the usefulness of the data. In this particular case, we find that it is crucial to take into account both the antenna beamwidth and the reflectivity weighting in the Doppler velocity calculations to get good results: mean reflectivity gradients displace the altitude and position from which the returns are coming from, and the Doppler information is coming from a different altitude than the one that is expected. And at near range, if one does not take correctly into account the observation geometry and the 3D velocity of targets, similarly large simulation errors will be observed (Table 2; Fabry 2010, 2012).

A simplification of the observational operator that holds promise is the following:

- Compute an effective angular beam displacement $\Delta\phi$ using the reflectivity at the top and the bottom of the beam of width ϕ_{rad}:

$$\Delta\phi = \frac{\phi_{rad}}{48}\left[dBZ_e\left(r,\theta,\phi+\frac{\phi_{rad}}{2}\right) - dBZ_e\left(r,\theta,\phi-\frac{\phi_{rad}}{2}\right)\right];$$

- Use it to select the model location from where Doppler velocity will be computed:

$$z' \approx z + \frac{\Delta\phi}{r}\cos(\phi'),$$

$$V_{r-meas} \approx V_r(r,\theta,\phi+\Delta\phi)$$

$$V_r(r,\theta,\phi+\Delta\phi) = u(L,\lambda,z')\cos(\phi'+\Delta\phi)\sin(\theta') + v(L,\lambda,z')\cos(\phi'+\Delta\phi)\cos(\theta') + [w(L,\lambda,z') - V_T(L,\lambda,z')]\sin(\phi'+\Delta\phi)$$

If the radial velocity from this location off the beam center is computed properly, the resulting errors and their correlation distance are considerably reduced (Fig. 4).

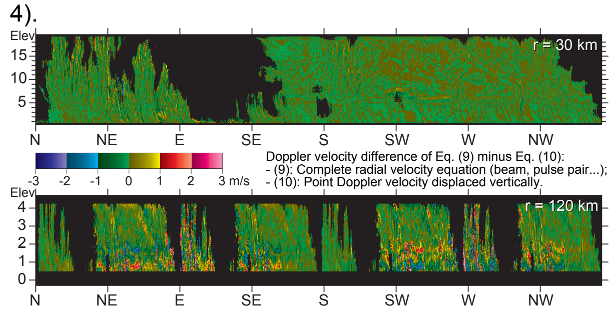


FIG. 4: Difference in the computed Doppler velocity between $V_{r\text{-meas}}$ and $V_r(r,\theta,\phi+\Delta\phi)$ at 30 km (top) and 120 km (bottom) range. Errors are much smaller and have shorter correlation distances than without taking into account an effective beam displacement (Fig. 3).

We still have to work on the reflectivity observation operator. We expect to see that simulations of reflectivity measurements using model fields will have many of the same errors as for Doppler velocity ones. But there will be additional issues related to the fact that most models *infer* reflectivity from mass or mixing ratios and we will have to deal with *Z-M* relationship errors.

Effect not considered	Mean wind error (m/s) (azimuth-dependent V_r error)		Divergence error (s^{-1}) (additive error on V_r)		Effective height error (m)	
	r = 30 km	r = 120 km	r = 30 km	r = 120 km	r = 30 km	r = 120 km
Horizontal projection: missing cos(ϕ')	**Up to 6% overestimate**	*Up to 0.5% overestimate*				
Horizontal projection (Earth curvature)	Tiny	Tiny				
Beam bending (assuming 4/3 Earth)	Tiny	Tiny	Tiny	Tiny	< 25	*< 100 (ϕ>1°)* **> 100 (ϕ<1°)**
Elevation angle error (0.05°)	Tiny	Tiny	$< 5 \cdot 10^{-7}$	$< 10^{-7}$	25	*100*
Vertical projection: not including V_T			**10^{-5} to 10^{-4}**	*$< 4 \cdot 10^{-6}$*		
Vertical projection: wrong DSD → bad V_T			*$2 \cdot 10^{-6}$ to $2 \cdot 10^{-5}$*	$< 10^{-6}$		
Vertical projection: using ϕ, not ϕ'	Tiny	< .05%	*$\sim 2 \cdot 10^{-6}$*	*$\sim 2 \cdot 10^{-6}$*		
No beam weighting					*Up to 60*	**Up to 1000**
No reflectivity weighting on beam					*Up to 60*	**Up to 1000**
No account for gas attenuation					< 1	< 15
No account for rain attenuation					Up to 5	*Up to 100*
No account for range folding	**Could be big**	**Could be big**	**Could be big**	**Could be big**		
No account for signal processing	*Small if no aliasing*	*Small if no aliasing*	*Small if no aliasing*	*Small if no aliasing*		
Beam blockage					*Up to ~250 at very low ϕ*	**May be large at very low ϕ**

TABLE. 2: Summary of the type and magnitude of errors introduced in radial velocity calculations by the poor handling of many effects at near range (30 km) and at far range (120 km). Errors in radial velocity calculations manifest themselves in up to three "symptoms": errors in mean wind speed, errors in divergence, and errors in the (effective) height of the wind. The font used on the effect name describes the correlation distances (normal font for errors with short (< 10 km) correlation distances, italic for errors with 10-100 km correlation distances, bold for errors with long correlation distances, and bold-italic for systematic biases) while the one used for the magnitude describes the importance of the error (normal font for insignificant errors, italic for significant errors, bold for critical errors that will likely result in assimilation problems). See Fabry (2010, 2012) for details and context.

4. THE COVARIANCE MATRIX OF THE OBSERVATIONAL TERM

The observational term of the data assimilation equation refers to the difference between the simulation of observations $H(\mathbf{x})$ and the observations $\mathbf{y}$. The data assimilation system tries, among other constraints, to minimize the mismatch between $H(\mathbf{x})$ and $\mathbf{y}$ using a matrix $\mathbf{R}^{-1}$ as weight. The weight depends on the confidence we have that $H(\mathbf{x})$ and $\mathbf{y}$ should match, and that confidence is an inverse function of the magnitude of errors in observations as well as errors in the observation simulations.

The covariance matrix of the observational term ($\mathbf{R}$) hence describes both the magnitude of the error on each {simulated minus true measurement} pair as well as the extent with which each of these errors is correlated with every other {simulated minus true measurement} error. If we are to assimilate n observations, the matrix $\mathbf{R}$ will have a size of $n \times n$. In radar data assimilation, n can reach millions. $\mathbf{R}$ is hence huge and can be complicated. The task of properly quantifying $\mathbf{R}$, or even simply using it, is daunting. As a result, $\mathbf{R}$ is drastically simplified in all radar data assimilation systems: first, errors in measurement simulations are ignored; then, measurement errors are often assumed to be independent of each other or, in rare cases, the correlation of errors is assumed to be Gaussian in space.

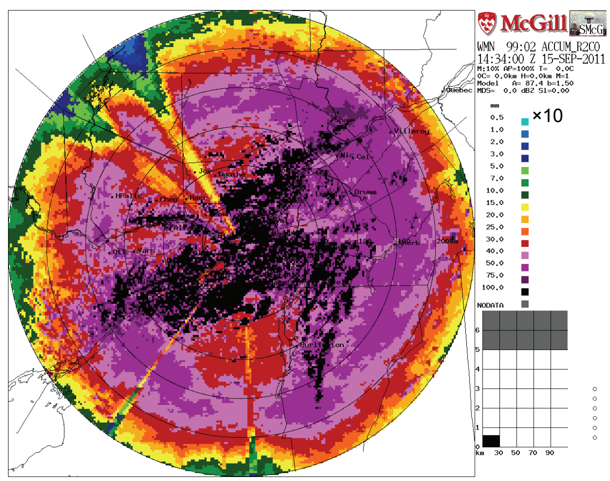


FIG. 5: 3-month of rainfall accumulation from the uncorrected reflectivities measured on the 0.5° PPI. If one ignores black areas that are clutter-contaminated pixels, we see that errors in the accumulation problem have some radial patterns (caused by beam blockage) and range-dependent ones (primarily due to the vertical profile of reflectivity).

Yet, any casual look at radar data reveals that errors typically have radial or azimuthal pattern. Indeed, radar meteorologists quickly learn to suspect any unusual pattern whose axis is in a radial direction or along a range ring. Data errors are also probably correlated in range and/or in azimuth, and not following a Gaussian pattern (Fig. 5). The other term contributing to $\mathbf{R}$ comes from errors in observation simulations (Fig. 3), and their correlation structure is far from Gaussian either.

What are the sources of data errors in reflectivity or in Doppler velocity? Fluctuations of the estimates due to the reshuffling of targets and the influence of noise are the better known ones; they have been well described (Keeler and Ellis 2000), their magnitude is known, and the covariance of errors with that of neighboring points is zero (errors at one location are not correlated with the errors at another). But they are not the only source of errors. Consider beam blockage. At ranges smaller than the source of blockage, data are not affected. At ranges beyond, reflectivities will be reduced, and Doppler velocities will come from a higher altitude than expected and be biased. Note again that data assimilation systems can only use unbiased data, and if someone did his homework, the effect of blockage should have been corrected for. But if the elevation angle scanned by the radar is a bit lower or higher than it should, that correction will be in error. The error will be very similar for all ranges beyond the blockage: if the beam is more blocked than anticipated at range r, it will also be more blocked at ranges beyond r. This error will hence be highly correlated in range, but only in the downrange direction. Some correlation may also occur in azimuth: if the beam was too low and there was blockage at azimuth ϕ, it will probably also be too low and there will likely be blockage at other azimuths near ϕ.

There exists a set of measurement errors that will have long correlation distance in range and/or azimuth: antenna pointing errors, propagation effects, transient transmitter or receiver problems, contamination by other transmitters or the sun… Quantifying them and their correlation structure requires an intimate knowledge of the hardware and its performance capabilities as well as the imagination needed to evaluate how each error source can affect both velocity and reflectivity measurements. Finally, once all measurement errors have been considered, we must also include the observation simulation errors and their correlation structure, and that will likely be the dominant term. We are ourselves just at the early stages of just figuring out what needs to be done to quantify the correlation of measurement errors. What is clear is that we will be busy for a while.

5. IN SHORT

To properly assimilate radar data is a complex task. Many of the difficulties have been ignored until now, and we are trying to attack many of them. What is already clear is that there are a few tasks one should undertake to improve radar data assimilation:

On the radar data side:

- [MOST IMPORTANT] Revisit data cleaning; on existing processed data, mark as "no information" (<u>not no echo</u>!) what may be even slightly contaminated; on raw data to be collected or processed, create a new data processing stream with data cleaning algorithms designed to eliminate all possibilities of bias and not be afraid to reject large data areas if they are not recoverable without correlated errors being introduced;
- On reflectivity, create a new type of information, "no weather echo above *x* dB", to be used for no echo and pixels contaminated by ground clutter or biological targets;

On the data assimilation side:

- Use better observation operators for Doppler velocity and reflectivity. It is very easy to ignore important terms in the observation simulation;
- [MOST IMPORTANT] Whether you chose to do the above or not, the correlation structure of the errors in the observation operator $H(\mathbf{x})$ needs to be better described, because its contribution will likely dominate that of the measurement errors in the covariance matrix of the observational term $\mathbf{R}$.
- Tweak the cost function code to accept and use information such as "no weather echo above *x* dB";
- Start thinking how to use a better $\mathbf{R}$. We plan to come up with a process for computing one.

6. ACKNOWLEDGEMENTS

This project was undertaken with the financial support of the Government of Canada provided through the Department of the Environment. But the views expressed herein are solely those of McGill University. Sincere thanks to Isztar Zawadzki for many discussions that have helped shape some of these ideas.

REFERENCES

Fabry, F., 2010: Radial velocity measurement simulations: common errors, approximations, or omissions and their impact on estimation accuracy. *Preprints*, 6th European Radar Conference, Sibiu, Romania, paper 17.2.

Fabry, F., 2012: How wrong are radial velocity measurement simulations in data assimilation systems, and how can they be improved? To be submitted to *Monthly Weather Review*.

Fabry, F., and J. Gadoury, 2009: Another method for ground target identification and filtering using spectral processing of dual-polarization returns. *Preprints*, 34th Conf. Radar Meteorology, Williamsburg, VA, Amer. Meteor. Soc., paper P5.19.

Keeler, R.J., and S.M. Ellis, 2000: Observational error covariance matrices for radar data assimilation. *Phys. Chem. Earth*, **25B**, 1277–1280.

Moisseev, D.N., and V. Chandrasekhar, 2009: Polarimetric spectral filter for adaptive clutter and noise suppression. *J. Atmos. Oceanic Tech.*, **26**, 215–228.

Sun, J., and J.W. Wilson, 2003: The assimilation of radar data for weather prediction. In *Radar and Atmospheric Science: A Collection of Essays in Honor of David Atlas*, Amer. Meteor. Soc. Monograph, 175–198.