Challenges in Quality Control

ECMWF Annual Seminar on Observations 2021

Kristen Bathmann

Kristen.Bathmann@noaa.gov
I.M. Systems Group at NOAA/NCEP/EMC

September 2021

Thanks to Hui Shao of JCSDA/UCAR, Lidia Cucurull of NOAA/OAR/AOML, Jim Purser of IMSG of NOAA/NCEP/EMC, and Emily Liu and Andrew Collard of NOAA/NCEP/EMC

Outline

- **Introduction**

- **Satellite Radiance QC**: infrared cloud detection, interaction of bias correction and QC, and challenges over land

- **GNSS Radio Occultation QC**: types and challenges with forward modeling in the troposphere

- **Variational QC**: formulation, benefits and challenges

- **Summary**

Quality control is an important step of the data assimilation process.

$$2J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{Hx} - \mathbf{y})^T \mathbf{R}^{-1} (\mathbf{Hx} - \mathbf{y}) + J_c.$$

Quality control removes observations $\mathbf{y}$ that:

- have gross errors, or
- that cannot be properly simulated by the forward model H

Such observations are not assimilated.

If assimilated, problematic observations could potentially cause asymmetrical observation distributions, conjugate gradient minimization issues, a drift in the analysis state, issues with bias correction, or an erroneous analysis state.

Types of quality control include:

• **Background check (or gross check)**: rejects observations with large departures, $\mathbf{Hx} - \mathbf{y} > C$.

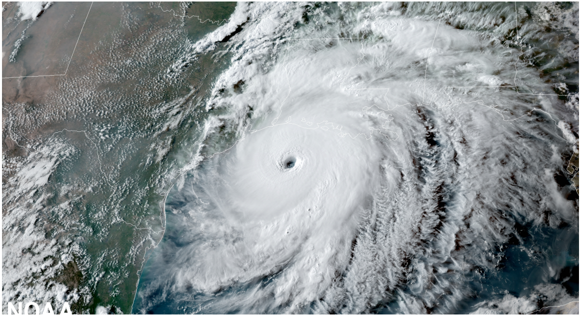
• **Variational quality control**: Accepts outliers, but adjusts observational weights.

• **Checks that are specific to observation type**: e.g. cloud detection for satellite radiances, or the super-refraction check for GNSS radio occultations.

In general, more sophisticated forward models require more sophisticated quality control.

Challenges with Quality Control

An observation can be perfectly good, and provide valuable information about the atmospheric state, but ultimately get rejected.

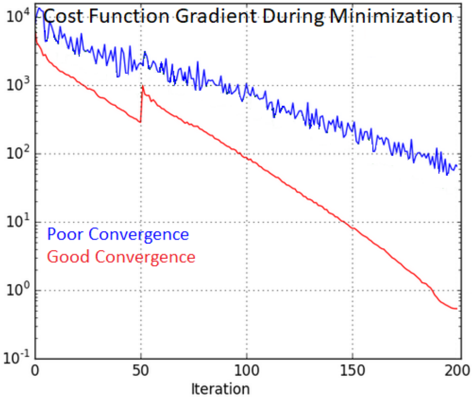


26 Aug 2020 23:01Z NOAA/NESDIS/STAR GOES-East ABI GEOCOLOR

Cloudy infrared radiances can potentially provide important meteorological information, but cannot be easily simulated by radiative transfer models.

Challenges with Quality Control

Conversely, problematic observations may pass quality control, and it can be difficult to trace problems in the analysis to these observations.



Minimization issues can sometimes be traced to tropospheric GNSS RO observations, or IR radiances sensitive to humidity.

Challenges with Infrared Radiances

The radiative transfer model:

$$I_v(\tau_t) = \epsilon S_v(\tau_s) e^{-(\tau_s - \tau_t)} + \int_{\tau_t}^{\tau_s} S_v(\tau) e^{-(\tau - \tau_t)} d\tau$$

$I_v(\tau_t)$: radiation at top of atmosphere
τ : optical thickness
v: wavelength
S : Planck function
ϵ : emissivity

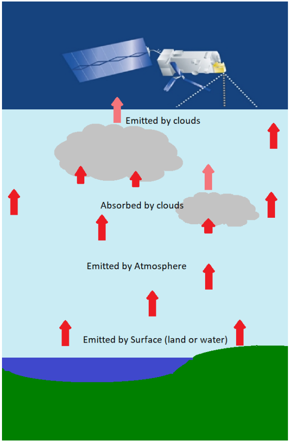
An instrument channel is a small interval around a wavelength v around which radiance is measured.

Most problems in assimilating satellite data come from:

- Instrument problems
- Processing errors
- Difficulty in simulating clouds and precipitation
- Difficulty in estimating surface emissivity

Challenges with Infrared Radiances: Cloud Detection

Clouds, while meteorologically important, are difficult to model in the infrared. The infrared cannot see through most clouds and many channels may be impacted by the presence of one.



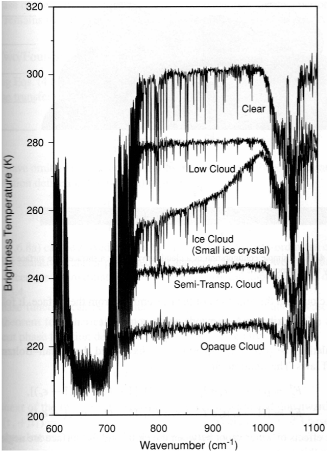


Figure: The interaction of clouds and infrared radiance (left) and the impact of cloud on brightness temperature observations (right, from K.N. Liou, 2004, page 115)

Challenges with Infrared Radiances: Cloud Detection

Cloud detection in the infrared focuses on **identifying the presence of a cloud, and removing channels that are impacted.** This can involve the estimation of a cloud top pressure and fractional coverage.

Often simplifying assumptions are made, such as overcast conditions and a single layer of cloud.

Additionally, algorithms may need empirical tuning, which may need to be adjusted from time to time.

Assimilating a cloud-contaminated observation without properly quantifying the error can have a detrimental effect on the analysis.

In the method of Eyre and Menzel, 1989, infrared cloud detection is accomplished by minimizing a cost function:

$$J = \sum_{i=1}^{\text{no. of chan}} \frac{1}{\sigma_i^2} \left(R_{\text{cloud},i} - R_{\text{observed},i}\right)^2 ,$$

where R is radiance and σ_i is the observation error. $R_{\text{cloud},i}$ is a function of cloud fraction, N, and differentiating J with respect to N allows for its estimation.

Then, a cloud is detected if

$$D \sum_{i=1}^{\text{no. of chans}} \frac{1}{\sigma_i^2} \left(R_{\text{clear},i} - R_{\text{observed},i}\right)^2 > \sum_{i=1}^{\text{no. of chan}} \frac{1}{\sigma_i^2} \left(R_{\text{cloud},i} - R_{\text{observed},i}\right)^2,$$

for some $0 < D \leq 1$.

As D increases, cloud detection becomes stricter. Generally as errors σ_j become smaller, cloud detection becomes stricter.

Challenges with Infrared Radiances: NCEP Cloud Detection Example

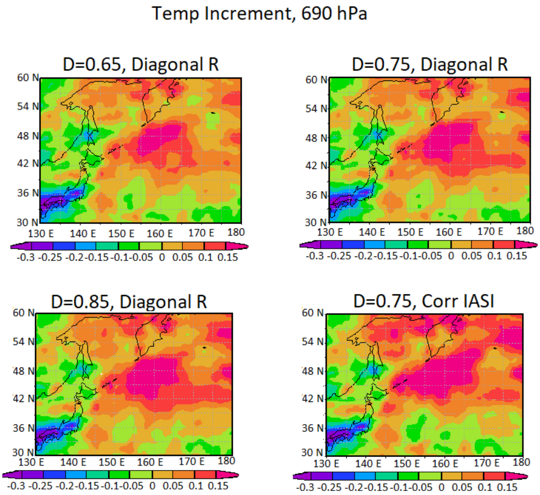


Figure: Temperature increments (analysis minus background) at 690 hPa, in a cloudy region. Cloud detection becomes stricter as D increases, and with correlated error, leading to larger increments.

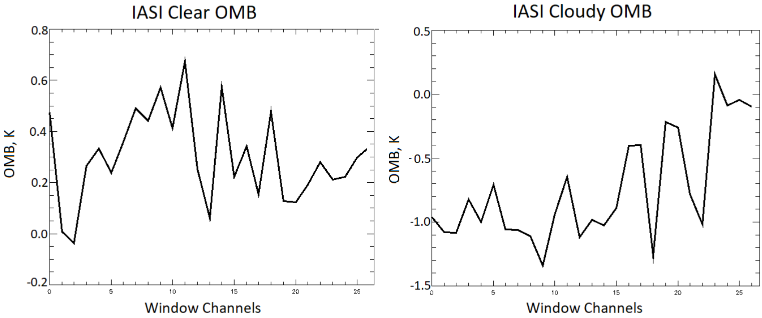


Figure: IASI OMB for window channels passing quality control; a suspected clear observation (left) and a suspected cloudy observation (right).

Challenges with Infrared Radiances: NCEP Cloud Detection Example

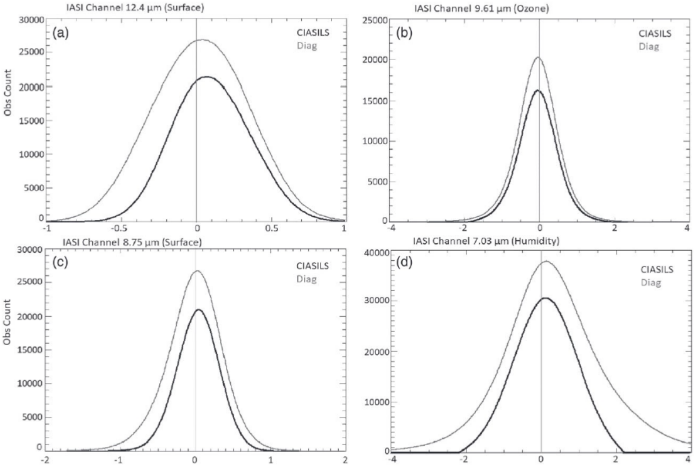


Figure: The observation minus background histograms for four IASI channels, over sea surfaces: Diag uses no correlated error, CIASILS uses correlated error with stricter cloud detection. (Bathmann and Collard 2021)

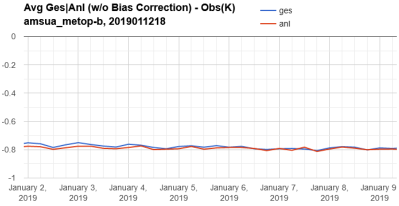
Challenges with Radiances: Interaction of QC and Bias Correction

Satellite observations are subject to significant biases.

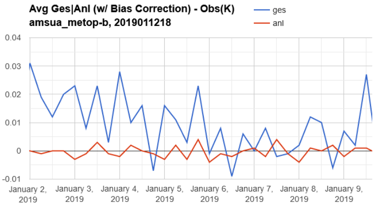
Except when the bias is due to the model background, we want to remove these biases, through bias correction.

Many centers use variational bias correction:

$$
\begin{aligned}
2J(\mathbf{x}, \beta) \quad &= \quad (\mathbf{x} - \mathbf{x}_b)^T \, \mathbf{B}^{-1} \, (\mathbf{x} - \mathbf{x}_b) \\
&+ \quad (\mathbf{Hx} - \mathbf{b}(\mathbf{x}, \beta) - \mathbf{y})^T \, \mathbf{R}^{-1} \, (\mathbf{Hx} - \mathbf{b}(\mathbf{x}, \beta) - \mathbf{y}) \\
&+ \quad (\beta - \beta_b) \, \mathbf{B}_\beta^{-1} \, (\beta - \beta_b) + J_c.
\end{aligned}
$$



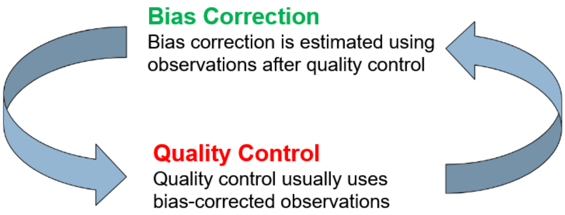
Without Bias Correction



With Bias Correction

Bias correction is applied before quality control. Bias correction is also re-estimated after quality control.

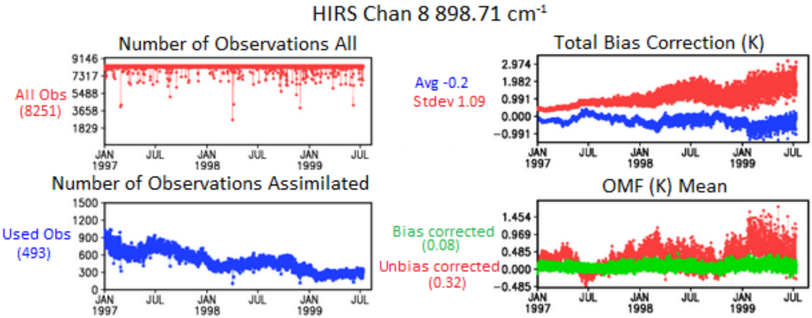
The radiance sample that is used in estimating bias correction must avoid OMB values with large model errors.



Challenges with Radiances: Interaction of QC and Bias Correction Example

In this case study, issues with surface albedo were present in the forecast model over land.

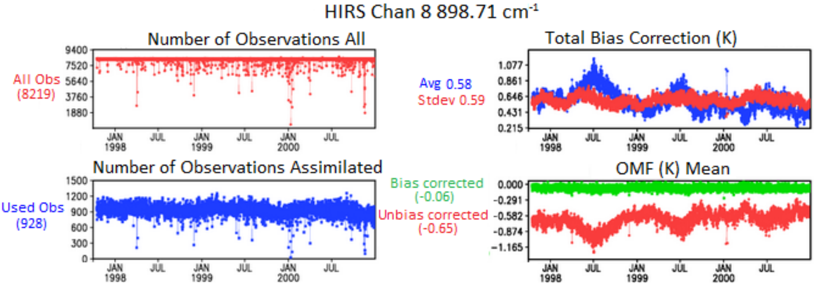
Bias estimates for infrared surface channels were misled by biases in the background over land.



Challenges with Radiances: Interaction of QC and Bias Correction Example

Two possible solutions at the time:

- Fix the surface albedo problem in the forecast model
- Exclude the surface sensitive channels over land



This experience motivated a change in the variational bias correction formulation:

$$h_{\text{old}}(\mathbf{x}, \beta) = \mathbf{H}\mathbf{x} + b^{\text{air}}(\mathbf{x}, \beta) + b^{\text{angle}}$$

$$h_{\text{new}}(\mathbf{x}, \beta) = \mathbf{H}\mathbf{x} + b^{\text{air}}(\mathbf{x}, \beta) + b^{\text{angle}}(\mathbf{x}, \beta)$$

Challenges with Infrared Radiances: Land

$$I_v(\tau_t) = \epsilon S_v(\tau_s) e^{-(\tau_s - \tau_t)} + \int_{\tau_t}^{\tau_s} S_v(\tau) e^{-(\tau - \tau_t)} d\tau$$

Assimilating surface-sensitive infrared observations over land can also be challenging. This is (mostly) because the radiative transfer model relies on an estimate of surface emissivity.

Over sea, sophisticated models exist to estimate surface emissivity. Over land, snow and ice, it is common to use a global database.

Cloud detection is also more difficult over these surfaces, because of emissivity uncertainty.

Bias correction can help, but may be heavily influenced by observations over sea.

Challenges with Infrared Radiances: Land

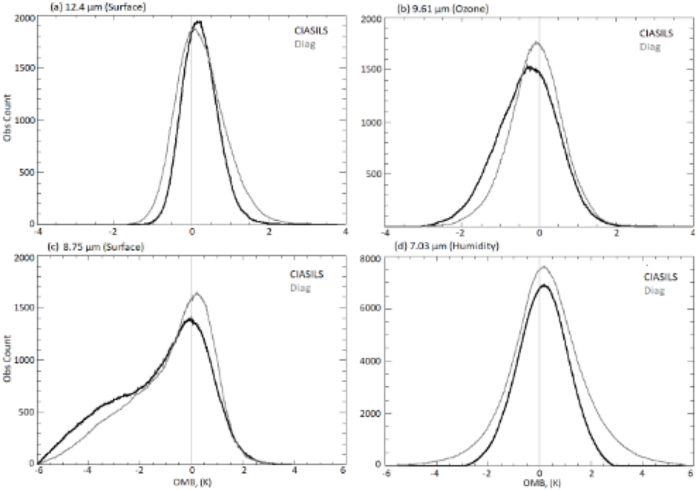


Figure: The observation minus background histograms for four IASI channels, over land surfaces: Diag uses no correlated error, CIASILS uses correlated error with stricter cloud detection. (Bathmann and Collard 2021)

Challenges with Infrared Radiances: Land

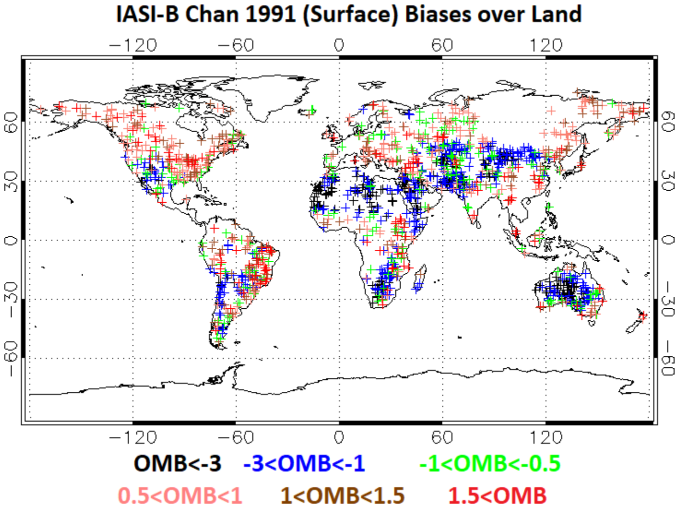


Figure: The locations of IASI observations of channel 1991 (8.75 μm) over land.

Challenges with Infrared Radiances: Summary

Infrared quality control is very important, but can be challenging.

QC could be improved by working towards better cloud detection and better surface emissivity modeling. Enhancements to bias correction may also help.

Channels with features that are not well modeled by the radiative transfer model should be blacklisted (for example IASI 1991 over land).

Bias correction should be carefully monitored, as there can be negative interactions with QC.

Recent work in all-sky assimilation permits the assimilation of microwave and midwave infrared radiances in the presence of cloud. This relies on an observation error that is a function of cloud liquid water.

In fact, situation dependent observation errors can provide great benefit, compared to using the same error (and error covariance) globally.

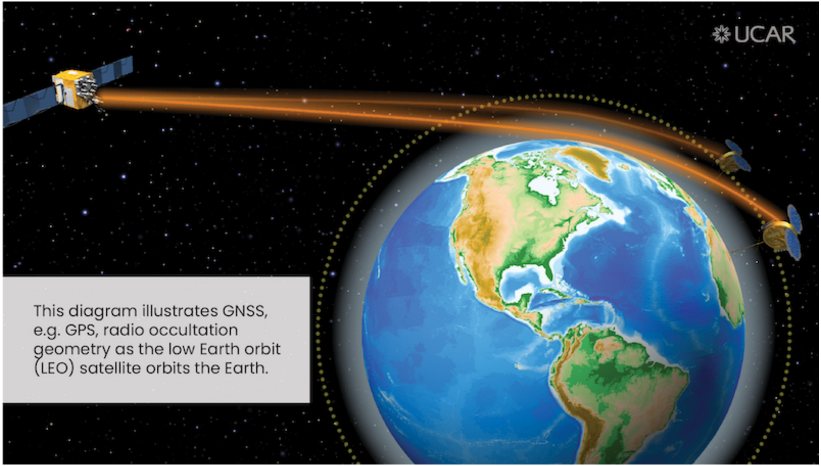


Figure: Courtesy UCAR. As the LEO satellite moves behind the earth, it obtains a profile of bending angle measurements, α, as a function of impact parameter a.

Forward Model for bending angle α (Cucurull et. al, 2007):

$$\alpha(a) = -2a \int_a^{\infty} \frac{d \ln(n)/dx}{\sqrt{x^2 - a^2}},$$

with refractivity n given by

$$10^6(n-1) = \frac{c_1 P}{T} + \frac{c_2 P_w}{T} + \frac{c_3 P_w}{T^2}.$$

c_1, c_2, c_3: constants
T: temperature
P: pressure
P_w: partial pressure of water vapor.

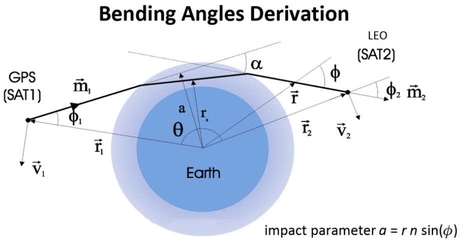


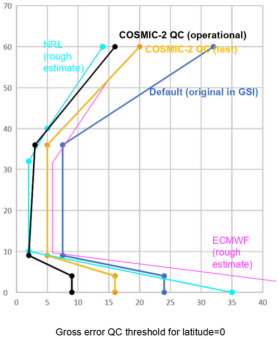
Figure: Courtesy Hui Shao, UCAR.

Quality control checks for GNSS RO:

- Reject observations taken above a certain height (typically > 45 km), or below a certain height.
- Reject observations affected by large gradients of atmospheric refractivity.
- Reject observations that are outside of the vertical boundary of the model levels.
- Reject observations that fail a background check, or a gross check.
- Reject observations that have problematic jacobians.

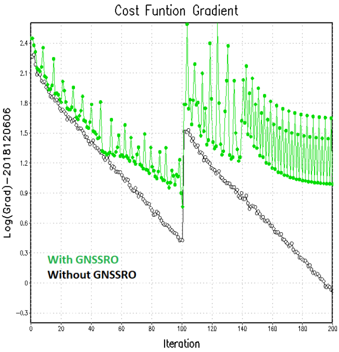
These checks are largely driven by challenges in forward modeling.

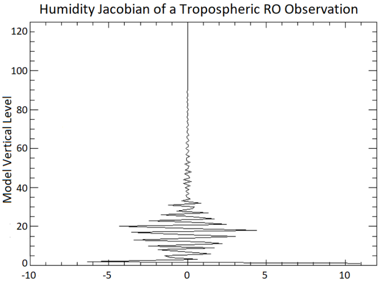
In the background check, observations are rejected if their incremental bending angles are greater than a cutoff value.



This height-dependent cutoff value is based on two seasonal experiments from 2009, (Cucurull et al, 2013). Failure to re-tune the background check for new instruments or as the forecast model evolves can result in RO observations having a negative impact on the forecast.

Challenges with GNSS Radio Occultations





Observation jacobians are used in minimization code. Assimilating an observation with a jacobian like this can introduce non-linearities in the cost function and ultimately result in minimization issues.

The immediate solution: tighter gross check below 5km and reject observations with larger jacobian values.

Ongoing work: improvements to the forward model.

Challenges with GNSS Radio Occultations

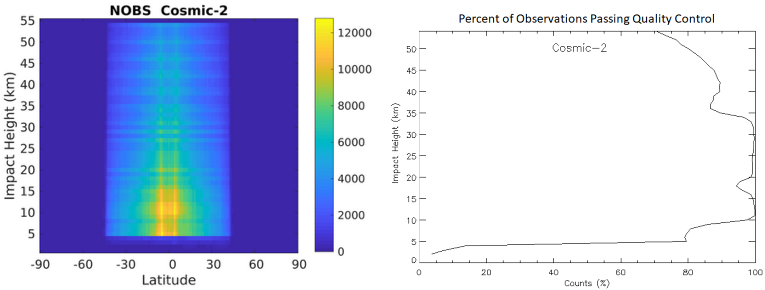


Figure: The total number of Cosmic-2 observations passing quality control over a three week period.

A deeper investigation into the GNSS RO forward operator revealed issues under super-refraction conditions when the number of vertical layers increased from 64 to 127 (credit Lidia Cucurull, NOAA/OAR/AOML).

When the atmospheric refractivity is greater than $-157 Nunits/km$ a GNSS RO ray will be bent so much that it does not leave the atmosphere. This is super-refraction (SR).

Under SR conditions, the assimilation of RO below the height of the SR layer is an ill-conditioned problem.

The GSI has a QC check that rejects observations near or below an SR layer.

Making adjustments to the SR check allowed a relaxation of the gross error check in the troposphere. Further adjustments are the subject of ongoing work.

GNSS RO assimilation presents unique challenges.

Designing quality control procedures for GNSS RO requires detailed knowledge of its forward model,

$$\alpha(a) = -2a \int_a^\infty \frac{d\ln(n)/dx}{\sqrt{x^2 - a^2}}$$

Many checks reject quality observations that cannot be properly simulated.

The background check is system dependent, and should evolve with time.

Some quality control checks exist as risk mitigation strategies.

Efforts to improve the use of GNSS RO observations should focus on empirical tuning and on improvements to forward modeling.

Variational Quality Control (VarQC) is not a criteria for rejection. Rather it is is a re-weighting of observations.

With VarQC it is possible to significantly loosen or do away with the gross check, and accept outliers into the assimilation.

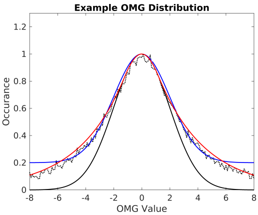
Every observation can contribute to the analysis, but those with large OMB values will be strongly down-weighted.

The observation cost function is reformulated such that the probability distribution function conforms to a predetermined shape, for example a Huber norm distribution.

The basic idea of VarQC: when $\frac{x}{\sigma} = \frac{O-G}{\sigma}$ is close to zero, we assume a normal, Gaussian distribution.

When the observations deviate from the guess by a large amount, putting the $O - G$ value in the tails of the probability distribution, this is an indication that non-Gaussian effects of measurement error (measurement error + representation error) dominate.

In this case, we try to control the shape of the tails (through parametrization), while preserving the Gaussian shape of the central part of the distribution.



VarQC changes the formulation of this distribution in the observation cost function $2J = -\ln(F(x))$.

A purely Gaussian distribution is defined by

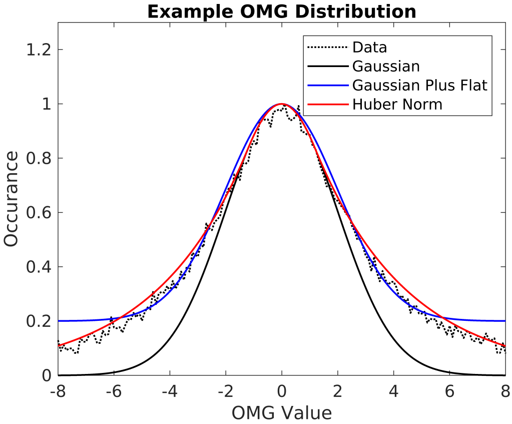
$$F_G(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\rho_G(x)}{2}\right), \rho_G(x) = \frac{x^2}{\sigma^2}.$$

The Gaussian plus flat distribution is

$$F_F(x) = \frac{\gamma + \exp\left(-\frac{x^2}{2\sigma^2}\right)}{\gamma + 1}.$$

The Huber norm distribution is

$$F_H(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\rho_H(x)}{2}\right), \quad \rho_H(x) = \begin{cases} \frac{x^2}{\sigma^2} & |x| \leq c \\ \frac{2c|x| - c^2}{\sigma^2} & |x| > c \end{cases}$$


Example OMG Distribution
Data
Gaussian
Gaussian Plus Flat
Huber Norm
Occurance
OMG Value
1.2
1
0.8
0.6
0.4
0.2
0
-8
-6
-4
-2
0
2
4
6
8

The weight applied to an observation after VarQC is ratio of the new observation cost function to the purely Gaussian observation cost function:

$$W = \frac{J_o^{QC}}{J_o}.$$

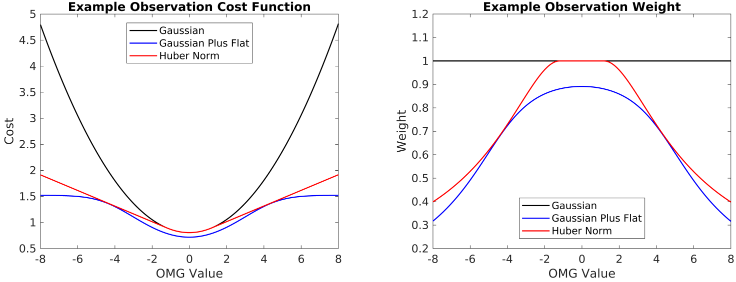


Figure: Observation cost function (right) and the weight given to the observation after applying VarQC (left).

Variational Quality Control with the Huber norm in Practice (Tavalato and Isaksen, 2015)

The shape of a Huber norm distribution is controlled by three parameters: σ, c_L and c_R.

$$F_H(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\rho_H(x)}{2}\right), \quad \rho_H(x) = \begin{cases} \frac{x^2}{\sigma^2} & c_L \leq x \leq c_R \\ \frac{2c_R x - c_R^2}{\sigma^2} & x > c_R \\ \frac{-2c_L x - c_L^2}{\sigma^2} & x < c_L \end{cases}$$

The observation error σ must be tuned such that σ^2 equals the variance of the Gaussian part of the distribution.

The background or gross error check should be relaxed.

The optimal transition points c_L and c_R can be found by using least-square fitting. The objective is to minimize

$$\sum_{i=1}^{n} \left(p(x_i)\ln(p(x_i)) - F_H(x_i)\ln(F_H(x_i))\right)^2,$$

$p(x_i)$: actual population in bin i
$F_H(x_i)$: population expected from the Huber distribution in bin i

Variational Quality Control: Benefits

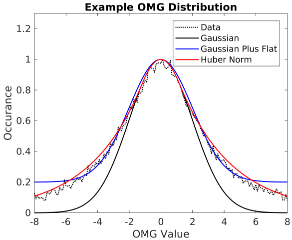
VarQC can benefit an analysis system. It permits the relaxation of the gross error check and allows outliers to influence the analysis.

Studies have shown that both the Gaussian plus flat and Huber norm distributions can positively impact a data assimilation system, but that the Huber norm yields better results.

VarQC is expected to have the biggest impact in intense weather events, such as tropical cyclones and extratropical storms. The analysis may not be able to resolve events on these scales well, and background departures may be larger than the background QC threshold. Such events may be poorly observed.

Variational Quality Control: Challenges

VarQC requires careful tuning. The assumed distribution must be a good fit to the data.



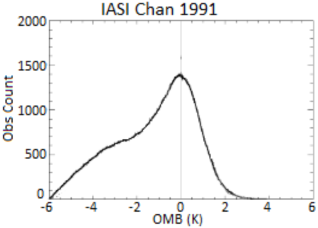


Figure: Simulated, well-behaved data (left) and IASI surface-sensitive brightness temperature over land (right) distributions.

The background should be of good quality. Otherwise, an observation may receive too small of a weight in the analysis.

Changing the observation cost function can introduce nonlinearity, and this necessitates an appropriate minimization technique. Introducing nonlinearity also has the potential to introduce multiple local minima in the cost function.

Summary

Quality control procedures are highly dependent on the assimilation system, and on observation type.

Assimilating observations of poor quality or that cannot be properly simulated can lead to problems in the analysis, which may be difficult to trace. Rejecting too many observations may also negatively impact the analysis.

It is important to understand the strengths and deficiencies of an observational forward model when designing a quality control procedure.

It is also important to carefully tune and re-tune quality control procedures as needed.

When issues arise, trying to determine the root cause is best (examine histograms, maps, jacobians, etc).

Solutions may not be easy, or feasible. Observation error inflation or situation dependent quality control may be good mitigation actions.

References

Kristen Bathmann and Andrew Collard.
Surface-dependent correlated infrared observation errors and quality control in the fv3 framework.
Quarterly Journal of the Royal Meteorological Society, 147(734):408–424, 2021.

L Cucurull, JC Derber, and RJ Purser.
A bending angle forward operator for global positioning system radio occultation measurements.
Journal of Geophysical Research: Atmospheres, 118(1):14–28, 2013.

Lidia Cucurull, JC Derber, R Treadon, and RJ Purser.
Assimilation of global positioning system radio occultation observations into ncep's global data assimilation system.
Monthly weather review, 135(9):3174–3193, 2007.

Jonathan R Eyre and W Paul Menzel.
Retrieval of cloud parameters from satellite sounder data: A simulation study.
Journal of Applied Meteorology and Climatology, 28(4):267–275, 1989.

Kuo-Nan Liou.
An introduction to atmospheric radiation.
Elsevier, 2002.

Robert James Purser.
Mathematical principles of the construction and characterization of a parameterized family of gaussian mixture distributions suitable to serve as models for the probability distributions of measurement errors in nonlinear quality control.
EMC Office Notes, 2011.

Christina Tavolato and Lars Isaksen.
On the use of a huber norm for observation quality control in the ecmwf 4d-var.
Quarterly Journal of the Royal Meteorological Society, 141(690):1514–1527, 2015.