Proactive Quality Control in Lorenz (1996) model

Tse-Chun Chen[1,*] **and Eugenia Kalnay**[1]

[1]University of Maryland, College Park, MD, USA
[*]tcchen@umd.edu

Introduction

Proactive Quality Control (PQC; Ota et al., 2013; Hotta et al., 2017) based on EFSO was proposed aiming to resolve the forecast skill dropout issues (Kumar et al., 2017) through identification and rejection of detrimental observations that may be harmful to the forecast. They showed using the Global Forecasting System (GFS) from the National Centers for Environmental Prediction (NCEP), that denying the detrimental observations identified by ensemble forecast sensitivity to observation (EFSO Kalnay et al., 2012) with 24-hr and 6-hr verification lead-time both reduced forecast errors in several forecast skill dropout cases. Hence, it was further proposed that PQC would be affordable in operational cycling to reduce or avoid skill dropouts in an online fashion. A major potential benefit in cycling-PQC is that the improved forecast may serve as a better background and lead to a cumulative improvement in the following analyses and forecasts. However, cycling-PQC has not been thoroughly tested yet. Idealized simulation experiments in a controlled environment can provide insights on how to optimally set up cycling PQC for realistic models.

Methods

The essential concept of PQC is to utilize the EFSO impact as observational QC for each DA cycle (e.g., 6 hours) for the identification of detrimental observations. The analysis is then modified to avoid the impact of those identified detrimental observations. It should be noted that EFSO cannot be computed until the next analysis becomes available for forecast error verification. The PQC algorithm can be summarized as inserting additional steps (verifying analysis for EFSO, EFSO computation, PQC analysis update, and the forecast from the updated analysis) into a standard DA cycle. The focus of this study is to compare the performance of five possible PQC analysis updates defined in Table 1.

Table 1. PQC update methods

Methods	Mechanism	Change in $\boldsymbol{K}$	Change in Spread	Repeat analysis	Computational cost
PQC_H	Recompute $\boldsymbol{K}$ without rejected observations	Large	Increased	Yes	High
PQC_R	Recompute $\boldsymbol{K}$ with up-weighted $\boldsymbol{R}$	Large	Increased	Yes	High
PQC_K	Reuse the original EFSO $\boldsymbol{K}$	None	None	No	Low
PQC_BmO	Assimilate background minus observation	Low	Reduced	(Serial update)	Medium
PQC_AmO	Assimilate analysis minus observation	Low	Reduced	(Serial update)	Medium

Results

Figure 1 compares the performance of all proposed PQC methods using EFSO (verified at 6-steps) and with varying percentages of rejected observations. Since the Kalman gain $\boldsymbol{K}$ of PQC_R approaches that of PQC_H asymptotically with increasing observational error, it is not surprising that PQC_H and PQC_R methods perform more or less the same in terms of both analysis and 30-step forecast error reduction. The errors are reduced the most when rejecting 10 % of the observations for the two methods. It is somewhat surprising that PQC_K, PQC_BmO, and PQC_AmO all outperform PQC_H and PQC_R, which are two most commonly used data denial methods. For the analysis quality improvement, the obvious choice of the threshold shifts towards 20%. PQC_K does not show any degradation of analysis until rejecting more than 60 % of the observations, whereas PQC_BmO and PQC_AmO stop showing improvement after 50% and even suffer from filter divergence beyond 60%. For the forecast quality improvement, the dependence of PQC_BmO and PQC_AmO on the thresholds are qualitatively similar to that in analysis performance. It is quite shocking to find that PQC_K has nearly no dependence on the thresholds between 10-th and 60-th percentile, especially when compared to the 10% optimal choice for PQC_H and PQC_R.

Discussion

Intuitively, the "flat bottom" of PQC_K (rather than the "check mark" shape of PQC_H and PQC_R) is more consistent with the estimated impact of the observations since the magnitude of the impacts between 10-th to 60-th percentile is really

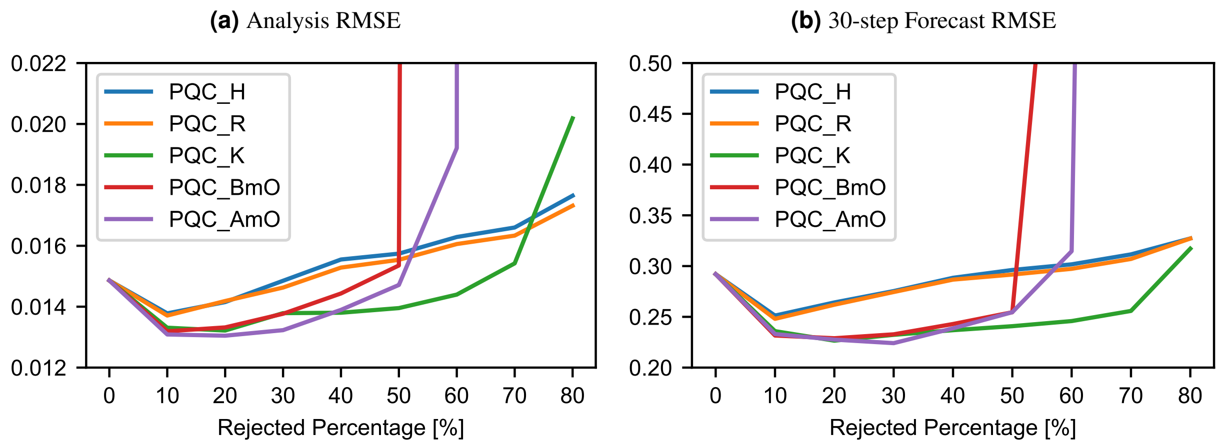

Figure 1. Performance of 6-step PQC with all 5 methods in terms of (a) analysis RMSE and (b) 30-step forecast RMSE as a function of rejection percentage.

small compared to that of those below 10-th percentile. And hence it should be insensitive ("flat bottom") to rejecting those observations between 10-th to 60-th percentile. This explains why the results are better for PQC_K than for PQC_H since PQC_K is more consistent with the nature of the computation of EFSO and the estimated impact. Note that EFSO provides the estimated impacts of each observation in the presence of all other assimilated observations, and hence the impacts remain valid as long as $\boldsymbol{K}$ does not change much. However, PQC_H and PQC_R significantly change $\boldsymbol{K}$ when rejecting some observations, thereby the accuracy of the estimated EFSO impacts becomes lower, and the PQC based on those impacts does not work as desired. The total AIs obtained at the end of the update consists of the AIs contributed from each individual observation and it is the AIs that determines the forecast error changes rather than the observation innovation. Hence, PQC should target the AIs corresponding to the detrimental observations rather than the observations themselves. And simple data denial by manipulating $\boldsymbol{H}$ and $\boldsymbol{R}$ does not necessarily reject the AIs that lead to forecast degradation especially when rejecting an excessive number of observations. PQC_K, by contrast, uses the exact same $\boldsymbol{K}$ to reject the exact detrimental AIs identified by EFSO and ends up with even larger improvements. In addition, the observations with largest impacts contribute to AIs among the most unstable modes, while the less impactful observations are associated with the neutral and stable modes which have little or no error growth. Hence, after rejecting the few very detrimental AIs, it does not matter much whether those less impactful AIs are rejected since the difference is very unlikely to grow in the future, thereby showing the "flat bottom" feature in the center of Fig. 1.

For PQC_BmO and PQC_AmO, they change $\boldsymbol{K}$ in a less radical fashion by "assimilating" new observations to the original analysis and yield improvements similar to PQC_K with a small number of rejected observations. But they suffer from filter divergence easily with a large number of rejected observations since the ensemble becomes overly confident due to the "additional" assimilation of opposite innovations. It is worth noting that the commonly observed difference in the impact estimated by EFSO and observing system experiments/ data denial experiments corresponds to the difference in PQC_K and PQC_H.

References

Hotta, D., T.-C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi, 2017: Proactive QC: A Fully Flow-Dependent Quality Control Scheme Based on EFSO. *Monthly Weather Review*, **145 (8)**, 3331–3354, doi:10.1175/MWR-D-16-0290.1, URL http://journals.ametsoc.org/doi/10.1175/MWR-D-16-0290.1.

Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A*, **64**, 1–9, doi:10.3402/tellusa.v64i0.18462, URL http://www.tellusa.net/index.php/tellusa/article/view/18462/xmlhttp://www.tellusa.net/index.php/tellusa/article/view/18462.

Kumar, V. K., A. Eichmann, J. Alpert, and S. Boukabara, 2017: Global Forecast Dropout Prediction Tool in Support of the NCEP Model Evaluation Group (MEG)—Collaboration Project Between JCSDA/NESDIS & NWS. *JCSDA Quarterly*, **(55)**, 5–13, doi:10.7289/V5V98648.

Ota, Y., J. C. Derber, E. Kalnay, and T. Miyoshi, 2013: Ensemble-based observation impact estimates using the NCEP GFS. *Tellus A*, **65**, 20 038, doi:10.3402/tellusa.v65i0.20038, URL http://www.tellusa.net/index.php/tellusa/article/view/20038/xml.