IMPLEMENTATION AND VALIDATION OF A GEO-STATISTICAL OBSERVATION OPERATOR FOR THE ASSIMILATION OF NEAR-SURFACE WINDS IN GSI

Project undertaken by Joël Bédard (DTC visitor) under the supervision of Joshua P. Hacker

Introduction

Biases and representativeness errors limit the global influence of near-surface wind observations. Although many near-surface wind observations over land are available from the global observing system, they had not been used in data assimilation systems until recently. Many are still unused. Winds from small islands, sub-grid scale headlands and tropical lands are still excluded from the UK Met Office data assimilation system (Ingleby, 2015), while other operational systems simply blacklist wind observations from land stations (e.g. Environment Canada). Similarly, the Rapid refresh (RAP) system use strict quality control checks to prevent degrading the near-surface wind analysis due to representativeness errors, similar to the Rapid Update Cycle (RUC) system, the RAP predecessor (Benjamin et al., 2007; Benjamin et al., 2010).

Model Output Statistics (MOS) methods are often used for forecast post-processing, and Bédard *et al.* (2015; 2016) evaluated MOS for use in the data assimilation. Doing so increases the consistency between observations, analyses and forecasts. Bédard *et al.* (2015) addressed representativeness and systematic error issues by developing a geo-statistical observation operator based on a multiple grid-point approach called GMOS (Geophysical Model Output Statistics: Bédard *et al.*, 2013). The idea behind this operator is that the nearest grid-points, or a simple interpolation of the surrounding grid-points, may not represent conditions at an observing station, especially if the station is located on complex terrain or coastal site. On the other hand, amongst the surrounding grid-points, there are generally one or several grid-points that are more representative of the observing site. Thus, GMOS uses a set of geo-statistical weights relating the closest NWP grid-points to the observation site. GMOS takes advantage of the correlation between resolved scales and unresolved scales to correct the stationary and isotropic components of the systematic and representativeness error associated with local geographical characteristics (e.g. surface roughness or coastal effects). As a result, GMOS attributes higher weights to the grid points that better represent the meteorological phenomena onsite.

The GMOS operator has been tested and compared with the results obtained from a conventional bilinear interpolation scheme used in most forward operators for in situ observations (Bédard *et al.*, 2015; Bédard *et al.*, 2016). By making background states and observations generally more consistent, GMOS produces relatively smaller innovations and analysis increments than bilinear interpolation, and it is less prone to generating strong perturbations in the resulting analysis. Results from observing system experiments also show that GMOS eliminates biases and significantly reduces representativeness errors as well as collocated observation error correlations, mainly over complex terrain (Bédard *et al.*, 2015). Due to the background-error covariances, near-surface wind observations impact the lower part of the atmosphere. Results also show that flow-dependent background error covariances from ensembles provide better vertical information propagation than static error statistics.

The evaluation of forecasts against observations show that Bilin significantly degrades upper-air fields when assimilating only wind data from ~5000 global SYNOP stations (Bédard *et al.*, 2016). GMOS on the other hand leads to better short-term near-surface wind forecasts and does not deteriorate the upper-air forecasts. However, the local impact decreases over time and is only significant for 6 h or shorter lead times when using hybrid error covariances. A detailed analysis based on initial model tendencies indicates that using the static component of the background error statistics leads to forecast impact that decays in time because it has no significant impact on the mass fields (the boundary-layer parameterization schemes diffuse the increments locally). On the other hand, when using the flow-dependent component of the background error statistics, the impact persists longer in the system because the analysis increments are modified for both wind and mass fields in a coherent way through multivariate covariances (pressure gradient forces are generated and counterbalance the vertical diffusion and orographic blocking schemes).

Near-surface wind observation from the global SYNOP stations were also assimilated along with the operational assimilation dataset in Environment Canada global deterministic prediction system (Bédard *et al.*, 2016). Forecast accuracy results are generally neutral. Nevertheless, forecasts and analyses from GMOS are more self-consistent than those from both Bilin and a control experiment (not assimilating near-surface winds over land) and the information from the observations persists up to 12 h lead time. Although these results are encouraging, they are not statistically significant as a large quantity of observations is already assimilated in the system (14 million observations per day).

With the objective of making a better use of near-surface wind observations and improving their impact on short-term tropospheric forecasts, this collaborative project aimed at assimilating near-surface wind observations over land in the NCEP RAP regional forecasting system. To address the statistical significance issue, near-surface wind observations from all available surface stations located over the North American continent were considered (19435 SYNOP, Metar and Mesonet stations). More specifically, the geo-statistical operator was implemented and tested in the hybrid version of the Grid-point Statistical Interpolation (GSI) data assimilation system (75% flow dependent / 25% static background error statistics), to demonstrate GMOS potential for possible use in operations. More specifically, the improvement of short-term tropospheric wind forecasts would benefit end users such as airports, wind turbine operators, farmers, outdoor recreation, security agencies, etc. Also the availability of the GMOS operator within the GSI code would provide operational weather prediction centers and researchers with an optional tool to assimilate observations significantly affected by representativeness errors.

Implementation

The GMOS observation operator was implemented in GSI version 1114 (*/glade/u/home/bedardj/Work/code/gsi_1114_v0*) following Bédard *et al.* (2015). A new module containing GMOS was added to the code (*geostatinterp_mod.f90*) and is initialized in the core section of GSI (*gsimain.f90*). The GMOS coefficients are set for each observing station in the 3D interpolation routine (*tintrp3.f90*) and stations ids are written in the observation space object in *setupw.f90* (the new w_ob_type is defined in *obsmod.F90*). The non-linear observation operator (*setupw.f90*) as well as the linearized and adjoint operators (*intw.f90*) call for the GMOS coefficients if the namelist variable *i_gsdsfc_gmos* is set to 1 in ~/Work/code/gsi_1114_v0/rapgsifix/gsiparm.anl.sh_gmos. This namelist option is added to *gsimod.F90* to enable the assimilation of near-surface winds over land using GMOS, and the default value is set to 0 in *rapidrefresh_cldsurf_mod.f90* to ensure backward compatibility. It should be mentioned that all modifications to the GSI code are preceeded by a commented statement starting with *JBedard* to make it easier to find changes in the code.

The GMOS statistical coefficients were trained using 2 months of historical (April 16 - June 19 2015) data after pre-processing the observations and the forecast fields from WRF model version 1120. The observation and forecast pre-processing codes can be found in the following directory: */glade/u/home/bedardj/Work/code/Training/* (*Read_BUFR_Training.py* and *Read_NETCDF_Training.py*, respectively). Note that it was necessary to thin the forecast data to enable transfer between NOAA and NCAR supercomputers (*UV_Thinning_2_first_levels.ksh*). The latest version of the training code (*Training_rotated_sigma_o.py*) includes coordinate transform output from the WRF model to account for the rotated forecast grid as well as the latitude and longitude of each model grid points. These parameters were obtained by print statements in *gridmod.F90* and are stored in *~/GMOS/Results/Work/* (*sin_cos_beta_ref_hi_rez* and *lat_lon_hi_rez* respectively). The training code uses site dependent observation error variances computed from innovation variances along with hybrid background error covariances (blending the homogeneous and flow dependent components) to perform the background check. The GSI code (*setupw.f90*) was also adapted to use the pre-computed site dependent observation error covariances from the file named *Whitelist_Sigma_O.dat* located in *~/GMOS/Results/GMOS_Weights*.

The strict quality control checks employed in the RAP system were disabled for two reasons: 1) to simplify the code (the quality control checks were previously located in different parts of the code), and 2) to allow assimilating greater amount of near-surface wind observations (the quality control checks are very restrictive: they allow using only a small amount of the available observations). The following quality controls were disabled:

- A "use list" was used in *sfcobsqc.f90* to identify stations with small representativeness errors: stations with large observation-forecast departure STD were blacklisted. The "use list" is not used anymore.
- Stations not presenting altitude or surface pressure or sea-level pressure measurements were rejected. The rejection flags in *read_prepbufr.f90* are now removed.
- Observations from stations below model orography were assimilated with inflated observation error variance to avoid representativeness errors. The inflation factors are now set to 1 in *setupw.f90.*
- Mesonet observations were not assimilated. The new convinfo file located in *~/Work/code/gsi_1114_v0/rapgsifix/nam_regional_convinfo_RAP.txt_gmos* now allows assimilating them.

The background check performed in *setupw.f90* is kept, but has been modified to use site dependent observation error variances and the non-linear version of the observation operator. A gross error check (wind speed must be between 1m/s and 30 m/s) has also been added in *setupw.f90*. The flow dependent quality controls now stand in a single line of code and the static ones are applied offline while computing the GMOS statistical components. Stations where GMOS statistical coefficients are not available are not assimilated (this feature replaces the need of a uselist). This procedure allows performing station quality control using historic data, simplifying the assimilation code, and assimilating a larger amount of wind observations from 19435 SYNOP, Metar and Mesonet stations (type 281, 284, 287, 288, 292, 293, 295).

Experimental design

A full observing system experiment (OSE) was performed using GMOS along with site dependent observation error variances and the new quality control checks to assimilate near-surface winds over land along with all observations assimilated operationally. It is compared against the unmodified operational version of RAP/GSI over the January 14 - January 31 2015 period.

It is important to note that the limited area version of the GSI code has been ill adapted from its global version. The background field is initially degraded in the code to fit the minimization grid, rather than computing the innovation at full resolution. Thus, the innovations only include low resolution information from the background if the minimization is performed on a coarser grid. Mathematical inconsistencies result from the fact that the analysis increment is optimized using the low resolution background field, and is then applied to the full resolution background fields. For these reasons, the two OSE are performed using the same grid for both the minimization and forecast model by setting the namelist option *grid_ratio_wrfmass* to 1 in *gsiparm.anl.oneobs_sfcwnd_hy.sh*. Although Ricardo Todling at GMAO-NASA is currently working to solve the issue in the GSI code, the operational system is affected by this deficiency as it minimizes on a coarser grid (*grid_ratio_wrfmass=2*).

Conclusion

As of now, the GMOS operator was implemented in the GSI code and the statistical coefficients were obtained using 2 months of historical data. The code was tested and validated using controlled test cases. The 2 weeks of evaluation runs are complete. Results from NOAA-GSD's standard verification package show that the assimilation of near-surface winds has a neutral impact on numerical weather predictions in general (no statistically significant differences). As mentioned in Bédard *et al.* (2016), more detailed analyses on specific cases (e.g. high impact weather) are necessary to evaluate the impact of near-surface wind observations, especially over highly observed areas like the North American continent. Such analyses will be performed by Ming Hu (NOAA), because the DTC visitor (Joël Bédard) no longer have access to the NOAA supercomputer facility, nor the experimental results.

References

Bédard J, Yu W, Gagnon Y, Masson C. 2013. Development of a geophysic model output statistic module for improving short-term numerical wind predictions over complex sites. Wind Energy, **16** : 1131–1147.

Bédard J, Laroche S, Gauthier P. 2015. A geo-statistical observation operator for the assimilation of near-surface wind data. Quart. J.R. Meteor. Soc., **141**, : 2857–2868.

Bédard J, Laroche S, Gauthier P. 2016. Near-Surface wind data assimilation: temporal propagation of the analysis increment and multivariate impact on forecasts. Submitted for publication in Mon. Wea. Rev.

Benjamin SG, Moninger WR, Sahm SR, Smith TL. 2007. Mesonet wind quality monitoring allowing assimilation in the RUC and other NCEP models. Preprints, 22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction, Park City, UT, Amer. Meteor. Soc., P1.33.

Benjamin SG, Jamison BD, Moninger WR, Sahm SR, Schwartz B, Schlatter TW. 2010. Relative short-range forecast impact from aircraft, profiler, radiosonde, VAD, GPS-PW, METAR, and mesonet observations via the RUC hourly assimilation cycle. Mon. Wea. Rev., **138** : 1319–1343.

Ingleby B. 2015. Global assimilation of air temperature, humidity, wind and pressure from surface stations. Quart. J.R. Meteor. Soc., **141** : 504–517.