Application of ML/AI for Calibration of Parameters of the Multigrid Beta Filter (MGBF) and its Performance within the Three-Dimensional Real-Time Mesoscale Analysis (3D RTMA) Project

Miodrag Rancic, R. James Purser, Manuel De Pondeca, Edward Colon, Ting Lei
Lynker at NOAA/NWS/NCEP/EMC
Russ Treadon, NOAA/NWS/NCEP/EMC

UIFCW 2023
A UFS Collaboration Powered by EPIC

Presentation outline

- 1. AI projects at EMC
- 2. Brief introduction into MGBF
- 3. Calibration procedure
- 4. Future extensions

UIFCW 2023
A UFS Collaboration Powered by EPIC

2. AI projects at EMC (Environmental Modeling Center)

Application of AI/ML methodology has become a very hot item in both academic and RT2O circles. NOAA's EMC was for decades a pioneer in this discipline, thanks to efforts of Dr. Vladimir Krasnopolski. From recently we started a very ambitious plan of embracing AI technology in many ongoing projects. (However, only several of those from that list are presently already funded, or we were able to incorporate them in ongoing projects.)

Emulation (6)
UFS short-term forecast surrogate
ML-based soil moisture
ML estimated land surface emissiv
Increase ensemble size
ML calibration of Multigrid Beta Filt
ML-based temperature-salinty bala
Post processing (3)
RRFS downscaling
Improve RTMA background
ML derived aviation parameters from RTMA

Observations (7)
AMV superobs
3DRTMA observation QC
Machine Learing observatio
Aircraft observation QC
Observation anomaly detection
ML based radiance bias correction
MI/AI Approach to Converting SMAP SSS to Bulk Salinity for Data Assimilation

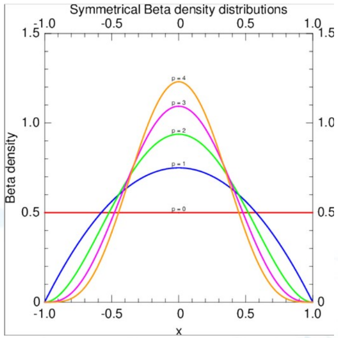
UIFCW 2023
A UFS Collaboration Powered by EPIC

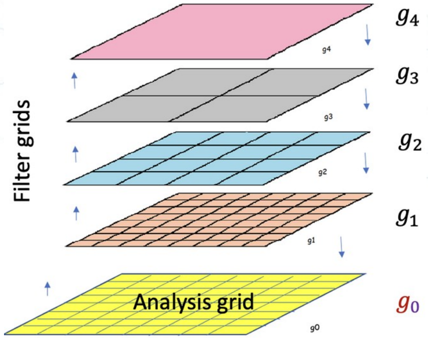
1. Brief introduction into MGBF technique

The Multigrid Beta Filter (MGBF) is a new technique for modeling of background error covariance (**B**), which is replacing the recursive filter (RF) as a quasi-Gaussian approximation of covariances in data assimilation at EMC.

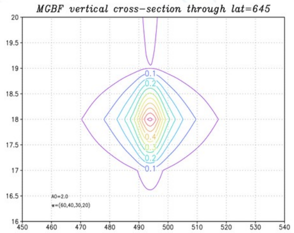
MGBF is based on the beta distribution with a finite support and is incorporated within a parallel multigrid structure to extend the spatial coverage. That all makes it more parallelizable, which results in much better scaling.

We plan to apply AI for calibration of parameters of MGBF





Parameters of the filter that control shape and size of the covariances are: intensities of horizontal and vertical aspect tensors and scale weights in the case of 4 grid generations



UIFCW 2023
A UFS Collaboration Powered by EPIC

3. Calibration procedure

We consider for now just the homogenous and isotropic case, in which we assume variability of covariances only in vertical direction with equal horizontal directions

Within "calibration", we are trying to define these 6 parameters of MGBF so that resulting covariance at each level has the shape and size equal to the background covariance, as suggested by the error statistics derived using the NMC method.

The NMC method is giving us for each of 3D variables, temperature (), stream function , velocity potential) and moisture the average values of a characteristic scale – e-folding distance – at each of of 65 model levels in both directions (and

UIFCW 2023
A UFS Collaboration Powered by EPIC

Similarly, error statistics gives us a horizontal e-folding distance () for each of 2D variables, surface pressure sea surface, ice surface, and land surface temperatures

Technically, based on input of only 2 variables (, , it would be very hard to conclude distribution of 6 parameters of MGBF ( ).

Therefore, we need to make some assumptions.

In present version of GSI, the background covariance is represented as a weighted sum of Gaussians with scaled correlation lengths defined as

$$\psi(x)=\sum_{i=1}^{n}\alpha_i G(x,\sigma_i L_x)$$

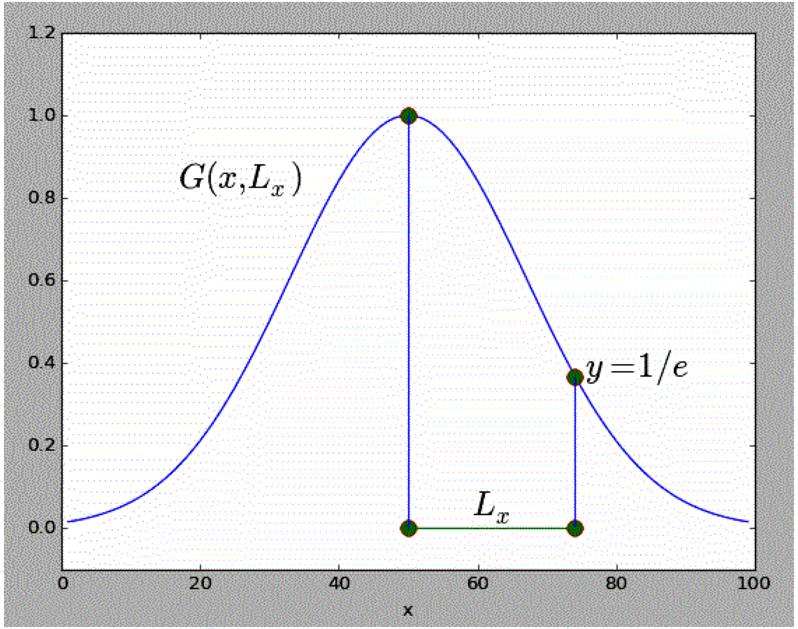
In version of GSI used in 3D RTMA,
in vertical direction:

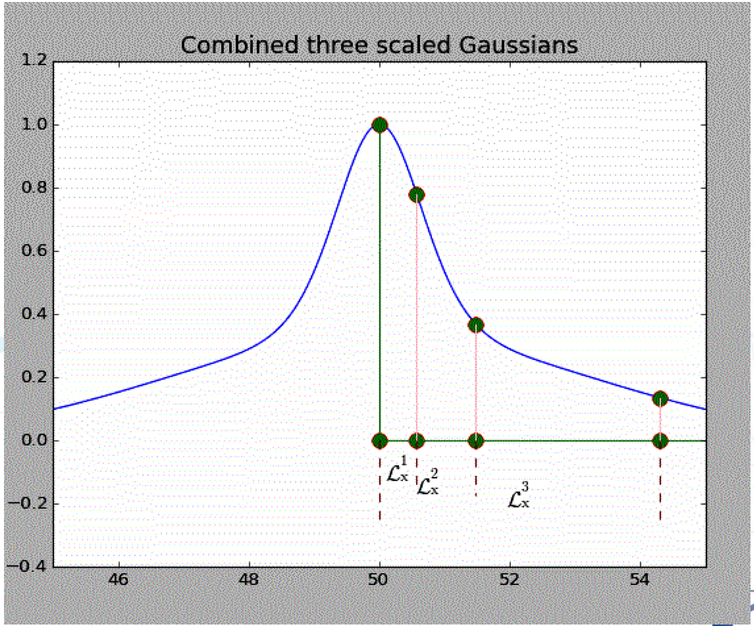
,

and in horizontal:

Thus, the covariances are taken to be much shorter than given correlation lengths.

UIFCW 2023
A UFS Collaboration Powered by EPIC





UIFCW 2023
A UFS Collaboration Powered by EPIC

In horizontal, the derived profile (not a Gaussian any longer), can be described with not one, but three (or more) characteristic distances,, defined maybe as a half of e-folding distance, a e-folding distance and a double of it.   The same procedure in the vertical direction give us three additional characteristic distances .

Thus, now we get 6 input variables to estimate 6 MGBF parameters:

( )

It seems that we now have enough information, but the question is how to do that!

We have found a solution in running a large series of tests with the standalone version of MGBF code

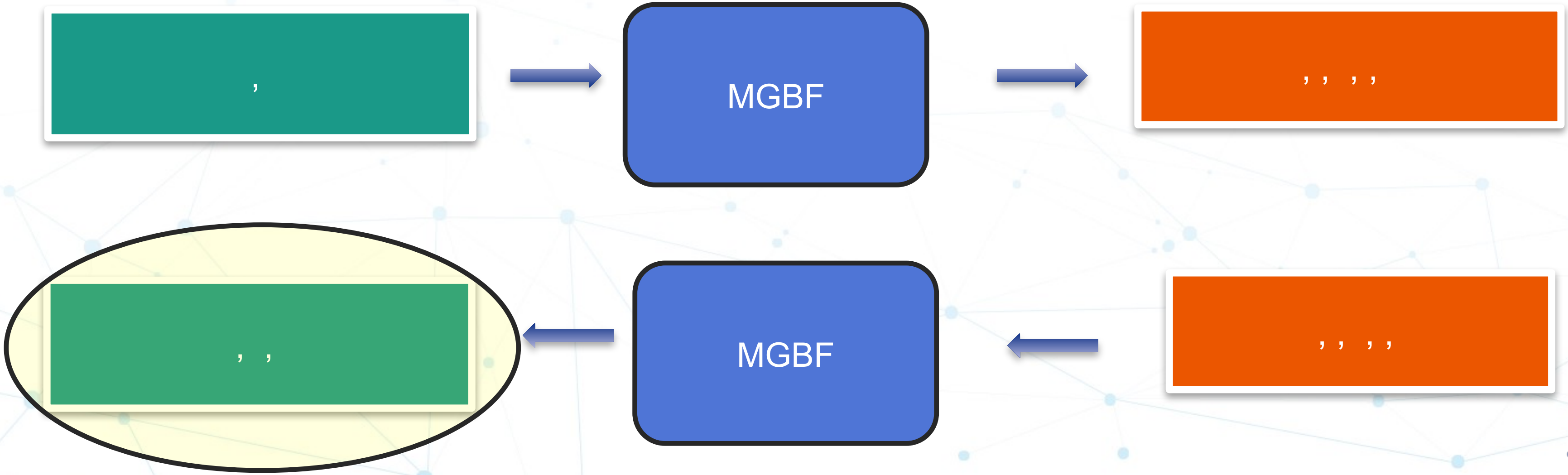
- At the same resolution and over the domain of same size at which GSI will be applied
- Initialize each the run with a unit impulse at a certain level
- Use slightly different inputs parameters ( ) in each run
- Measure the same characteristic lengths of resulting covariance in each run, that is e-folding distance from the maximum, one half and one quarter of it, both in horizontal and vertical directions:

( )

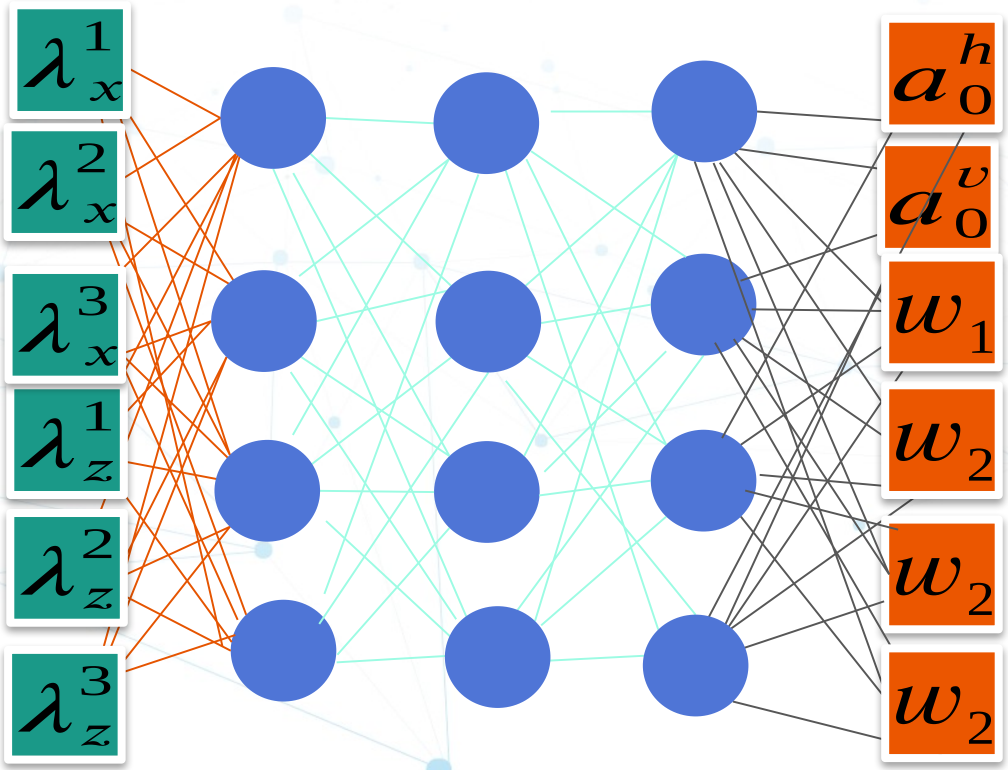
- In addition, it is important to measure the maximum value of the covariance

UIFCW 2023
A UFS Collaboration Powered by EPIC

The essence of the problem can be described with this diagram



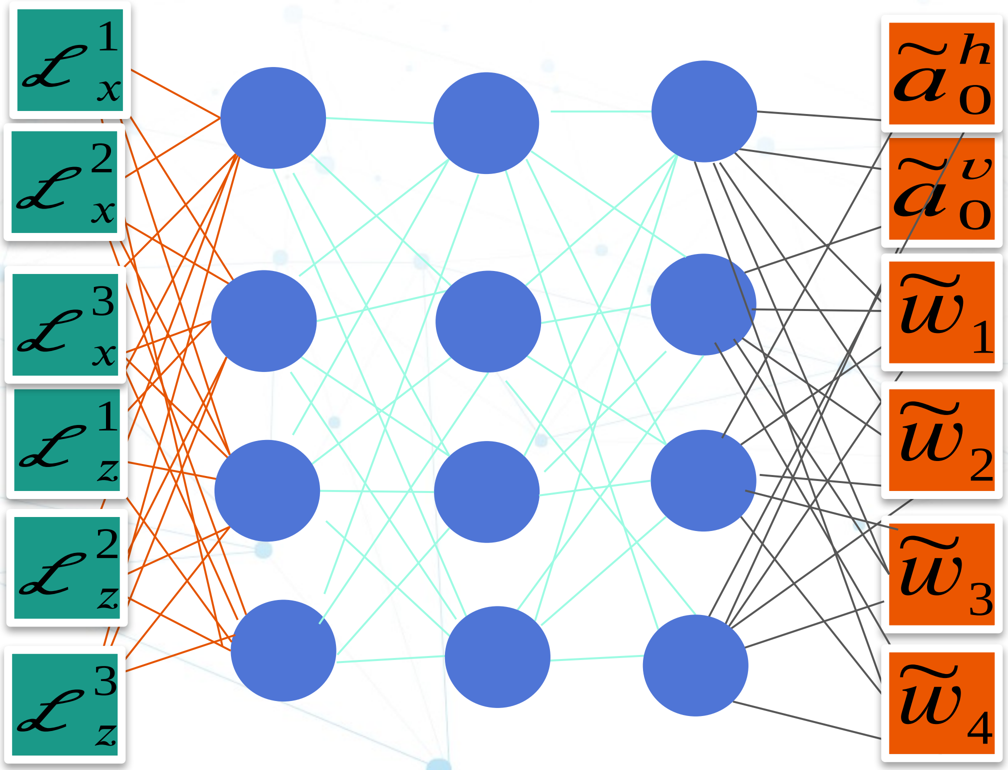
The problem could be solved by application of a deep neural net (NN) in which we reorder unput and output parameters



NN in a way emulates an inverse of MGBF by a nonlinear regression function.

We use half of data derived by running standalone version of MGBF for training (finding weighs and biases of NN) and half for testing

Once the NN is trained, we now use as an input correlation lengths derived from the error statistics and pick up on the output calibrated parameters of MGBF

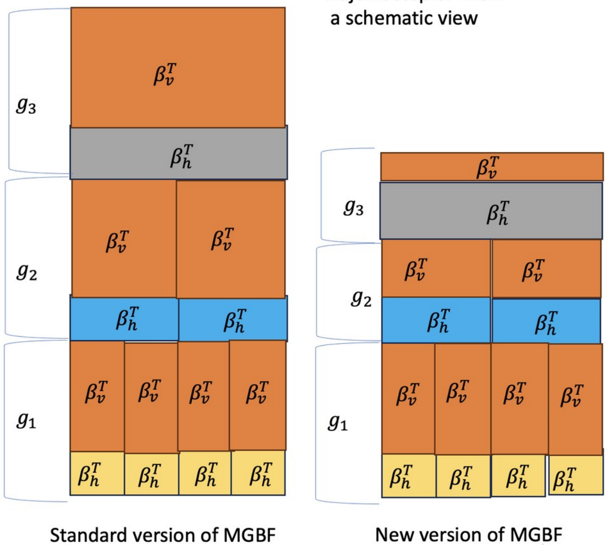


In the final step, we rerun the standalone version of MGBF with derived parameters, find maximum values of covariances, and divide by it derived scale weights, which will ensure that in GSI we get normalized covariances

4. Future Extensions

- In order to take full advantage of the described method, a new version of MGBF was developed to improve efficiency of filtering in vertical direction where we encountered very long covariances
- This was done by reducing vertical resolution of higher grid generation, which in a way mimics horizontal application of multigrid
- We are developing an upgrade of this version in which horizontal and vertical filter will be better separated, which will improve versatility of the described calibration procedure

Adjoint step of MGBF – a schematic view



- One limitation of the described method is assumption about the shape of the covariance that we made based on correlation length derived by the NMC method
- We are looking into various ways how to get more useful information directly from error statistics of the background field without making such assumptions

Thank you for your attention!
Stay safe and healthy!

UIFCW 2023
A UFS Collaboration Powered by EPIC