Challenges for the hybrid forecast models: representation of the systematic forecast error with machine learning

Sergey Frolov (CIRES/NOAA, Boulder, CO, USA)
William Crawford (NRL, Monterey, CA, USA)

Hybrid models and on-line bias correction

Emergence of machine learning (ML) tools is renewing interest (Pathak et.al. 2018, Bolton and Zana 2019; Bonavita et.al. 2020) in hybrid models of the form:

$$\frac{Dx}{Dt} = \underbrace{F(x)}_{\text{Deterministic physics-based tendency}} \odot \underbrace{Q(x)}_{\text{Stochastic tendency}} \odot \underbrace{b_w(x)}_{\text{Data-driven (ML) tendency}}$$

Function composition (e.g. addition or multiplication)

The simplest version of the Hybrid Model is to add a constant bias term as following:

$$\frac{Dx}{Dt} = F(x) + \underbrace{\overline{\Delta x}}_{\text{A constant bias correction}}$$

There is a long history of research that uses average analysis correction tendency as a bias correction term (Saha 1992, Bowler et.al. 2017, Bhargava et.al. 2019, Piccolo et.al. 2019, Crawford et.al. 2020).

We use our experience with this method (Crawford et.al. 2020) to highlight fundamental problems for the hybrid forecast models.

Next steps

1) Can bias correction be simplified to 1D?
   – Demonstrate that ML can replicate the statistics of archived analysis corrections (possibly conditioned on season (Julian day), lat/lon, time of the day, background state)

$$b_w(x(lat,lon), JD, lat, lon)$$

2) Develop on-line learning algorithms to deal with biased analysis.

$$b_2 = (1-\alpha)b_0 + \alpha\, b_1$$

3) Introduce the dependence of the correction on lead time of the forecast.

$$b_w(x(lat,lon), JD, \tau, lat, lon)$$

4) Develop stochastic version of the bias tendency estimate:

$$b_w(x) \sim N(mean(x^a), \mathrm{cov}(x^a))$$

References

- Crawford, W., Frolov, S., McLay, J., Reynolds, C. A., Barton, N., Ruston, B., & Bishop, C. H. (2020). Using analysis corrections to address model error in atmospheric forecasts. *Monthly Weather Review*, *148*(9), 3729–3745. https://doi.org/10.1175/MWR-D-20-0008.1
- Bhargava, K., E. Kalnay, J. A. Carton, and F. Yang, 2018: Estimation of systematic errors in the GFS using analysis increments. J. Geophys. Res. Atmos., 123, 1626–1637, https://doi.org/10.1002/2017jd027423.
- Bowler, N. E., and Coauthors, 2017: Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. Quart. J. Roy. Meteor. Soc., 143, 1280–1302, https://doi.org/10.1002/qj.3004.
- Saha, S., 1992: Response of the NMC MRF model to systematic-error correction within integration. Mon. Wea.Rev., 120, 345–360, https://doi.org/10.1175/1520-0493(1992)120,0345:ROTNMM.2.0.CO;2.
- Piccolo, C., M. J. P. Cullen, W. J. Tennant, and A. T. Semple, 2019: Comparison of different representations of model error in ensemble forecasts. Quart. J. Roy. Meteor. Soc., 145, 15–27, https://doi.org/10.1002/qj.3348.
- Bonavita, M., & Laloyaux, P. (2020). *Machine Learning for Model Error Inference and Correction.* https://doi.org/10.1002/ESSOAR.10503695.1

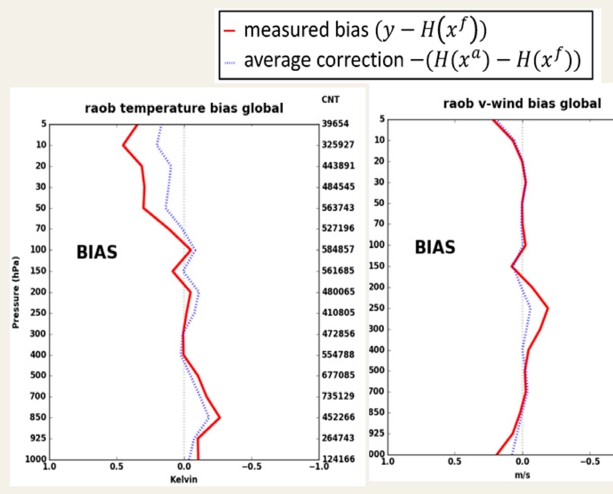
Acknowledgements

•Authors acknowledge support from the U.S. Office of Naval Research and the Office of Oceanic and Atmospheric Research of the U.S. National Oceanic and Atmospheric Administration.

Contact information

Sergey Frolov: sergey.frolov@noaa.gov
William Crawford: william.crawford@nrlmry.navy.mil

Challenge 1: Source of truth



Measured and estimated bias compared to the radiosonde observations in NAVGEM. Average 3.5 month cycling experiment for 6 hour forecast.

ML methods require samples of truth to estimate the bias model $b_w(x)$:

$$b_w(x) = \arg\min_w \left\{ \mathrm{E}\left[\left(x^{true} - x \right) - b_w(x) \right] \right\}$$

Self-analysis provides an attractive choice to be used as truth $x^{true}=x^a$:

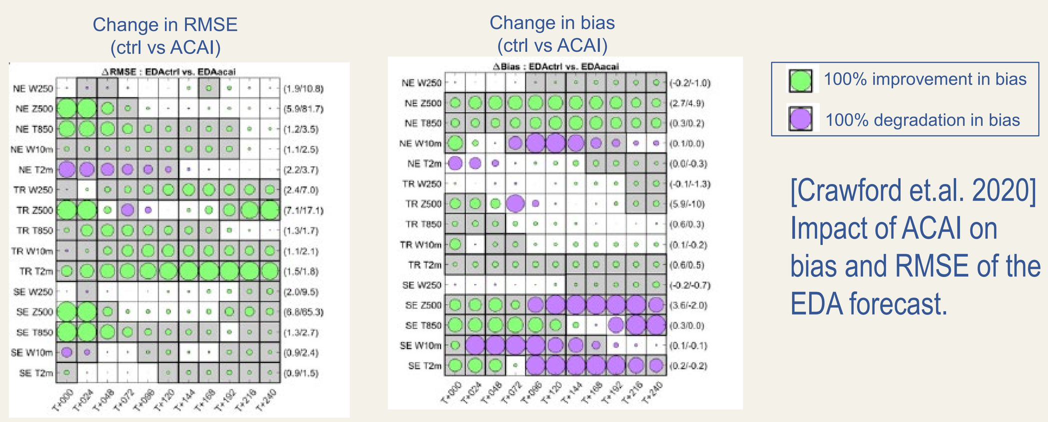
$$b_w(x) = \mathrm{E}\left[\frac{x^a - x^f}{\Delta t^a} \right] = \frac{\overline{\Delta x^a}}{\Delta t^a}$$

However, due to biases in the model and obs., x^a is biased (Dee and Da Silva 1998). E.g. Figure above.

Options for better sources of truth:

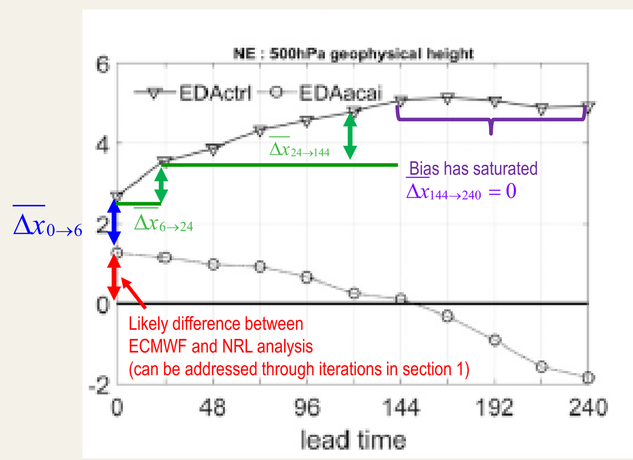
- Bias-aware DA (e.g. Dee 1998, Tremolet 2006, Laloyaux et.al. 2020)
- Re-play to a "less biased" analysis (ERA5), bias corrected-analysis, or a weighted mean of analysis ensemble.
- Iterative bias correction (batch or on-line estimates).

Challenge 2: Short-range vs long-range errors



[Crawford et.al. 2020] Impact of ACAI on bias and RMSE of the EDA forecast.

- On-line bias correction (ACAI) has overall positive impact on RMSE in both ensemble (figure above) and deterministic systems (see paper).
- However, ACAI can degrade bias for certain variables.
- Bias degradation often occurs at later lead times.
- Crawford et.al. identified several reasons for bias degradation (e.g. see figure below).



Error (including the bias growth often saturate as forecast progresses. Hence, the fast error (bias) growth during the first 6 hours is not representative of bias at later lead times [Crawford et.al. 2020].

Proposed solution:
Develop forecast time-lead dependent bias correction

$$b_w(x, \tau_1 \to \tau_2) = \arg\min_w \left\{ \mathrm{E}\left[\Delta x_{\tau_1 \to \tau_2} - b(x, \tau_1 \to \tau_2) \right] \right\}$$

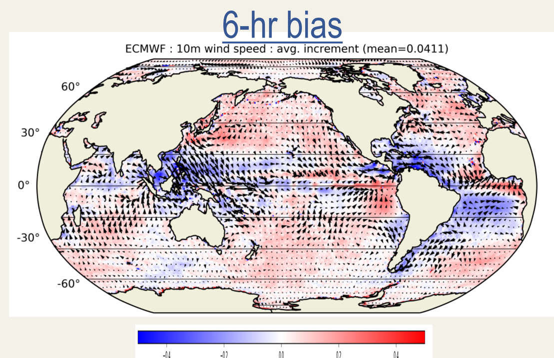
Where the training data is an incremental bias accrued by the forecast from forecast lead τ_1 through τ_2:

$$\Delta x_{\tau_1 \to \tau_2}(t) = [x^a(t+\tau_1) - M_{\tau_1}(x^f(t))] - [x^a(t+\tau_2) - M_{\tau_2}(x^f(t))]$$

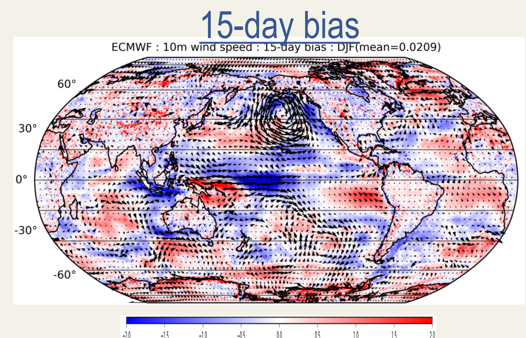
It is likely that time dependent corrections would need to be trained iteratively.

Challenge 3: Multiscale errors

6-hr bias



15-day bias



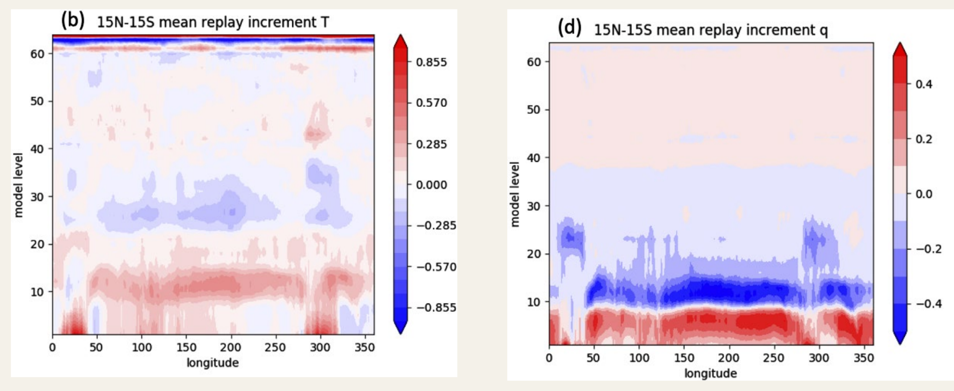
Tropics (1D-like):

- Magnitude of the bias is about the same for 6h and 15d.
- Likely can be approximated with a 1D bias model that depends on lat-lon.

Mid-latitudes at longer lead times (3D-like):

- Bias grows with the forecast lead.
- Strong imprint of synoptic flow (e.g. Aleutian low).
- A 3D model of bias might be more appropriate.

Challenge 4: Flow-dependent errors



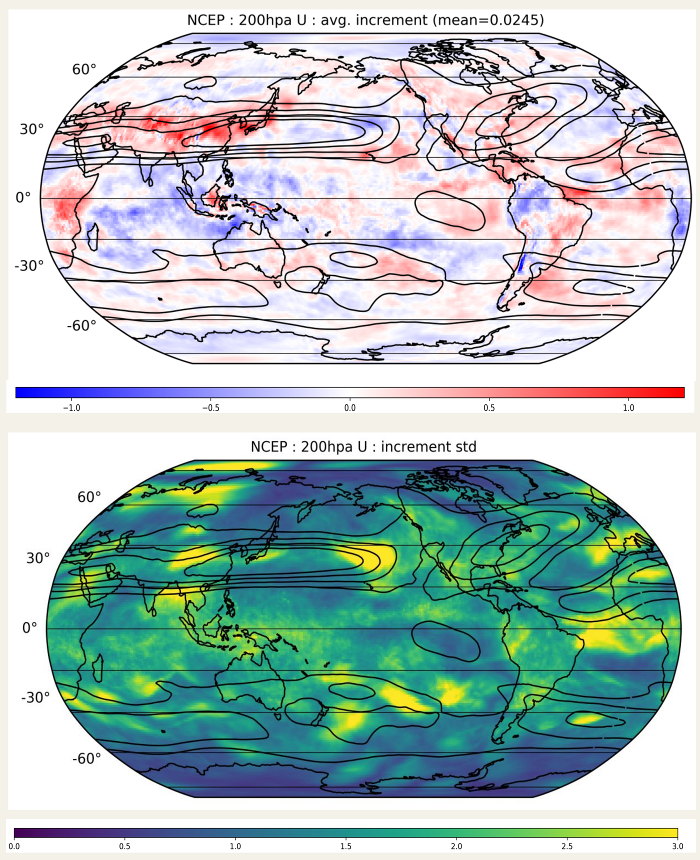
[Bengtsson et.al. 2019 MWR]: Time and latitudinally average replay increments over the tropical belt. Red areas indicate where the IFS analyses are moister or warmer than the short-range forecasts.

(above) Replay increments show systematic bias in the IFS physics tropical physics that has strong dependence on the height of the boundary layer.

Proposed solution:
Train a 1D ML model that uses either 1D state as an input or the diagnosed height of the boundary layer.

$$b_w(x, lat, lon) = \arg\min_w \left\{ \mathrm{E}\left[\Delta x(lat,lon) - b(x(lat,lon), lat, lon) \right] \right\}$$

Challenge 5: Stochastic representation of error



Error (including the bias growth often saturate as forecast progresses. Hence, the fast error (bias) growth during the first 6 hours is not representative of bias at later lead times [Crawford et.al. 2020].

Errors of the day (e.g. the analysis corrections) have a strong variability compared to the mean error.

It would be beneficial to able generate a random draw from the distribution of analysis increments

$$b_w(x) \sim N(mean(x^a), \mathrm{cov}(x^a))$$

Proposed solution:

- Use variational autoencoder to compress archive of analysis increments as a gaussian probability distribution in some latent space L.
- Use the decoder from the VAE to generate samples from the latent space that look like sample of the analysis increments

Challenge 6: Simultaneous development of the forecast model, bias correction, and data assimilation

•Currently model development and data assimilation are developed and tuned sequentially (tune physics first then tune DA).
•Common criticism of bias estimation is that it will complicate physics tuning (biases will be hidden by the DA).
•Suggestions:
  – Still do your best to train unbiased model.
  – When bias estimation is used, the goal is to minimize the magnitude of the bias correction.
  – New methods provide a formal way to diagnose magnitude of the bias.
  – Possible extensions to diagnose source of the bias (attribution of bias to specific tendency terms) based on the analysis increments.
  – Only use bias correction inside of the DA window.