Model optimization with a genetic algorithm.

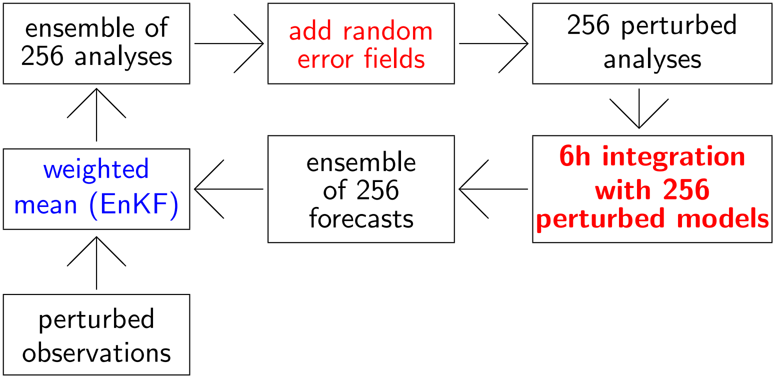
Pieter Houtekamer
October 7, 2020
Environment and Climate Change Canada

Collaborators: Xingxiu Deng, Bin He, Dominiques Jacques, Ron McTaggart-Cowan, Leo Separovic, Paul Vaillancourt, Ayrton Zadra

Central question

Can we use machine learning to improve a state-of-the-art Numerical Weather Prediction model?

Inspiration: Ensemble Prediction and Parameter Estimation System: Järvinen, Laine, Solonen, Haario, QJRMS, 2011.



All inputs and output are in the form of ensembles. Every 6h, the new observations are perturbed randomly. Isotropic random fields are added to the analyses and different configurations of the model physics are used.

Genetic algorithm for model configurations

1. Select 256 configurations,
2. Use the EnKF for one day to provide 4 sets of 256 background trajectories,
3. Use the about 3 000 0000 observations used by the EnKF to compute an ensemble score,
4. Over at most N=32 iterations:
    1. find the (bad) member that, when removed, improves the score the most,
    2. find the (good) member that, when removed, degrades the score the most,
    3. verify that replacement improves the score,
    4. replace the parameters of the bad member by those of the good member while adding a small perturbation to the model parameters,
5. continue at point 2.

Modified update rule

For continuous parameters use a perturbed and bounded contraction:

$$
\begin{aligned}
p_{try} &= \alpha p_{good} + (1-\alpha) p_{bad} + \epsilon_p \\
\epsilon_p &\sim N(0, (2\alpha - 2\alpha^2)\sigma_p^2) \\
p_{new} &= max(p_{min}, min(p_{max}, p_{try}))
\end{aligned}
$$

$\alpha = 0.9$: parameter for contraction towards p_{good}.
ϵ_p: noise term to avoid collapse in case of uninformed updates.
p_{min}, p_{max}: imposed minimum and maximum values.

For discrete parameters (switches in the model with various possible values):
Change to p_{good} with probability α.

CRPS

To mesure the quality of the ensemble $x_i, i = 1, \cdots, N_{ens}$, given an observation y_k, we can use the CRPS (Continuous Ranked Probability Score, Gneiting et Raftering 2007, Zamo et Naveau 2018):

$$CRPS(x, y_k) = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} \mid x_i - y_k \mid - \frac{1}{2N_{ens}^2} \sum_{i,j=1}^{N_{ens}} \mid x_i - x_j \mid$$

For more robust results, we use the observations $y_k, k = 1, \cdots, N_{obs}$ used by the EnKF during one day. The std dev of the observation σ_k is used for normalization:

$$CRPS(x) = \frac{1}{N_{obs}} \sum_{k=1}^{N_{obs}} CRPS(x, y_k)/\sigma_k$$

CRPS and member selection

With the CRPS, we associate a score $\mathcal{J}(1, 2, 3, \cdots, N_{ens})$ with a set of members $1, 2, 3, \cdots, N_{ens}$.
For one optimization step, we determine in a leave-one-out manner:

$$\begin{aligned} \mathcal{J}_1 &= \mathcal{J}(2, 3, \cdots, N_{ens}) \\ \mathcal{J}_2 &= \mathcal{J}(1, 3, 4, \cdots, N_{ens}) \\ \vdots &= \vdots \\ \mathcal{J}_{N_{ens}} &= \mathcal{J}(1, 2, \cdots, N_{ens} - 1) \end{aligned}$$

to obtain the worst performing member i_{bad} and the best performing member i_{good}. The procedure is continued with the set: $1, 2, i_{bad} - 1, i_{good}, i_{bad} + 1, \cdots, i_{good}, \cdots, N_{ens}$ where a small perturbation is added to the set of good parameters.

Cost function

A different choice for the evaluation function relies on the deterministic performance of each member individually and is more consistent with the Bayesian framework of particle filters in which members are penalized for poor individual performance.

The evolutionary algorithm optimizes the 1-member score $\mathcal{J}_1(x)$:

$$\mathcal{J}_1(x_i) = \frac{1}{N_{obs}} \sum_{k=1}^{N_{obs}} \frac{(\mathcal{H}_k x_i - y_k)^2}{\sigma_k^2}$$

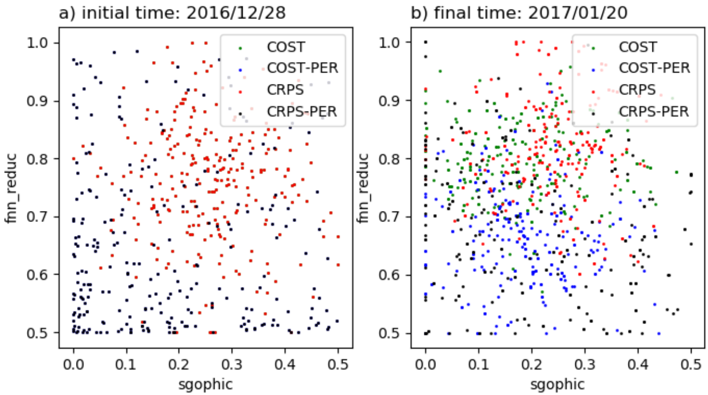
Note that only the distance between the state of x_i and the observations is used to assign a penalty to parameter set i.

Experiments

1. STABLE: parameter estimates do not evolve,
2. COST: deterministic evaluation function,
3. COST-PER: det. eval. function and perturbed initial parameter sets,
4. CRPS: probabilistic evaluation function,
5. CRPS-PER: prob. eval. function and perturbed initial parameter sets.

The perturbed initial parameter sets have a different initial mean and std dev, but respect the same minimum and maximum values.

Example of initial and final ensembles



List of model options
Paul Vaillancourt and Ayrton Zadra (2019)

1. recenter around EnKF or EnVar [0,1],
2. algorithm for deep convection: {kfc2,kfc3},
3. trigger for kfc2/3: $[0.03, 0.08]$m/s,
4. closure for shallow convection: {equilibrium,cape},
5. evaporate detrained condensate: {false,true},
6. updraft radius kfc2/3: [1300,1700]m,
7. updraft radius kfc2/3 over water: [800,1300]m,
8. critical phase blocking height: [0.0,0.5],
9. Mixing length: {black62, turboujo, boujo},
10. reduction factor turbulent flux: [0.5,1.0],
11. minimum Obukhov length: [5,20]m,
12. stability function: {beljaars91,delage97},
13. radius for ice in radiation scheme: $[15,35]\mu$m.

The recentering

The ensemble of EnKF analyses $x^a_{i,\mathrm{EnKF}}$ is translated:

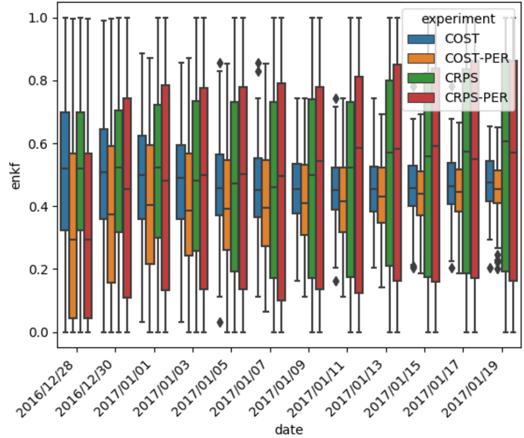
$$x^a_{i,hyb} = x^a_{i,\mathrm{EnKF}} + \gamma(x^a_{\mathrm{EnVar}} - \overline{x^a_{i,\mathrm{EnKF}}}) \quad (1)$$

Three interesting values:

- $\gamma = 0$: the EnKF analysis is not changed due to the EnVar,
- $\gamma = 1$: recenter around the EnVar analysis,
- $\gamma = 0.5$: as proposed by Penny (2014) and tested at ECMWF by Bonavita and Hamrud (2015), equal weight is given to the EnKF and the EnVar.

To obtain the best value of γ, we start the experiment with an ensemble of different values $\gamma_i, i = 1, \cdots, 256$ in the range [0,1]. The evolutionary algorithm will let the best value emerge.

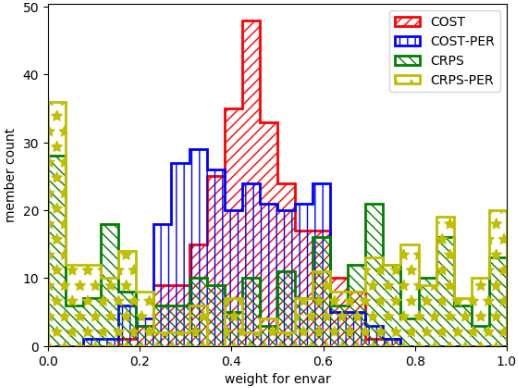
Convergence for the assimilation method (γ)



Box-whisker plots for γ are given for the 2-week optimization period. The value zero gives all weight to the EnKF and one gives all weight to the EnVar.

The CRPS-based optimization algorithm increases the interquartile range of γ. This suggests that differences between the EnKF and EnVAR sample data assimilation uncertainty in a realistic manner.

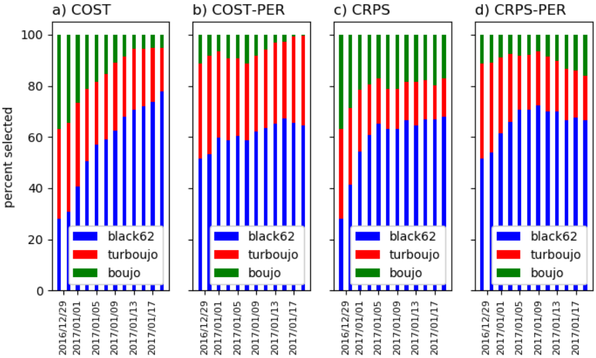
Histograms for the assimilation method (γ)



Histogram for γ at the end of the optimization.

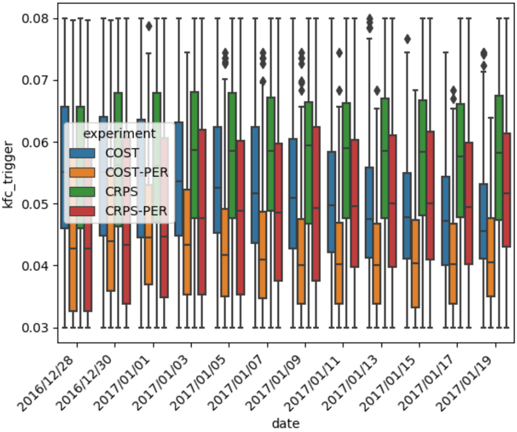
The evolution tends towards the distribution of the CMC-hybrid (Houtekamer, Buehner, De La Chevrotière, QJRMS, 2018) where half of the members is recentered on the EnKF and the other half on the EnVAR. (The noise in the update equation works against the formation of bimodal distributions).

Algorithm for the mixing length computation black62, turboujo or boujo



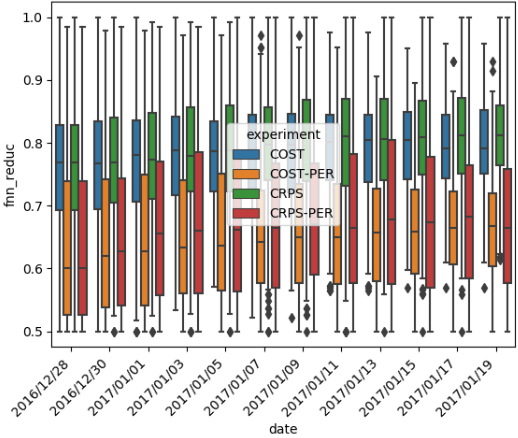
The possible values are in [black62, turboujo, boujo]. The optimization favors the black62 scheme. Subsequent deterministic experimentation also showed slightly better results with black62 at 39 km resolution. Model development continued with black62.

Trigger parameter for Kain-Fritsch



The acceptable domain is [0.03, 0.08]m/s. Prior deterministic experimentation suggested a value of 0.05. The optimisation has little impact on the interquartile range.

Parameter fnn_reduc: Reduction of boundary layer cloud factor turbulence



The acceptable domain is [0.5,1.0]. Very little evolution is observed.

Correlations between model parameters

Correlations between parameter values can be estimated from the parameter ensemble. An earlier version of the experiments showed a 0.62 correlation between the hybrid gain parameter γ and the choice of either the kfc2 or kfc3 algorithm for deep convection. This was traced back to the unnecessary use of a humidity adjustment procedure in the EnKF system.

Current results show no correlations bigger than 0.45.

The lack of correlations may be a side-effect of the formulation of the noise term. It is perhaps relatively large and is independent for each parameter.

Medium-range verifications

Table: Percentage improvement in five-day GEPS CRPS verifications with respect to the tuned CMC-hybrid. For the u, v wind components, temperature T, and geopotential height, values are the average improvement at the levels 10, 50, 100, 250, 500, 850 and 925 hPa. For dewpoint depression, the average is only over the levels 250, 500, 850 and 925 hPa. The verification is against radiosondes.

	STABLE	COST	COST -PER	CRPS	CRPS -PER
u	10.474	10.534	10.654	10.404	10.867
v	9.477	9.736	9.620	9.320	9.881
Height	6.719	7.658	7.414	7.968	9.248
T	8.663	9.016	9.135	8.711	9.515
$T - T_d$	1.738	2.034	2.022	2.013	2.287
Average	7.414	7.796	7.769	7.683	8.360

Conclusions with regard to the genetic algorithm

Positive aspects:

- discrete parameters (switches) can be handled.
- interesting results were obtained for the mixing length estimates and the hybrid gain parameter γ.
- unexpected correlations between parameters have helped to diagnose an issue with the treatment of humidity in the EnKF analysis (not shown).
- longer-range forecasts issued with the optimised ensembles also improved.

Negative aspect:

- Many parameters were not identifiable (their estimated distributions did not evolve substantially from their initial distributions).

Issues: comparison with other methods

The update rule is ad hoc. The noise term could, for instance, be sampled from a multi-dimensional distribution (for the continuous parameters).

The parameters α and 32 could be tuned.

The use of an EnKF context with 6h forecasts is historical. What is the optimal forecast length to tune model parameters?

Alternative methods, such as the iterative ensemble smoother, could be used for the parameter estimates.

The method is expensive and cumbersome with individual experiments taking several weeks to complete.

Summary of the use to improve the Canadian global ensemble prediction system

The genetic algorithm has supported a planned upgrade to the Canadian global EPS. Conclusions regarding a humidity bias in the EnKF, the mixing length algorithm, and the hybrid gain parameter, have been accepted. The expected error reduction is 7 %.

We start using Stochastic Parameter Perturbations (SPP). Here, Markov chains are used to perturb model parameters. It would be natural to use information from the genetic algorithm to inform on the distributions used in SPP. This, however, has not yet been done.

Use of the genetic algorithm normally leads to additional model development. Often, this is what leads to the improvements in the system. The genetic algorithm is perhaps just directly linked to a 0.5 % reduction in error.

The end

Thank you