EnsAI: An Emulator for Atmospheric Chemical Ensembles

Michael Sitwell

Air Quality Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada

michael.sitwell@ec.gc.ca

Ensemble Forecasting of Atmospheric Chemical Constituents

➢ **Forecast Error Covariance Estimation**
- Quantifies level of uncertainty and correlations in a forecast
- Essential part of assimilation/inversion algorithms that combine forecasts with observational information

➢ **Ensemble Forecasting**
- Can provide flow/time-dependent covariances
- Common formulations do not require the model adjoint (unlike 4DVar)
- Can be very computationally expensive since it requires many model runs for each time step

➢ **Ensemble Forecasting of Chemical Constituents**
- Atmospheric chemical models are typically more computationally demanding then NWP models
  - Computation of chemical reactions
  - Advection of many chemical concentrations
- Very computationally costly to run a chemical ensemble with an adequate number of members

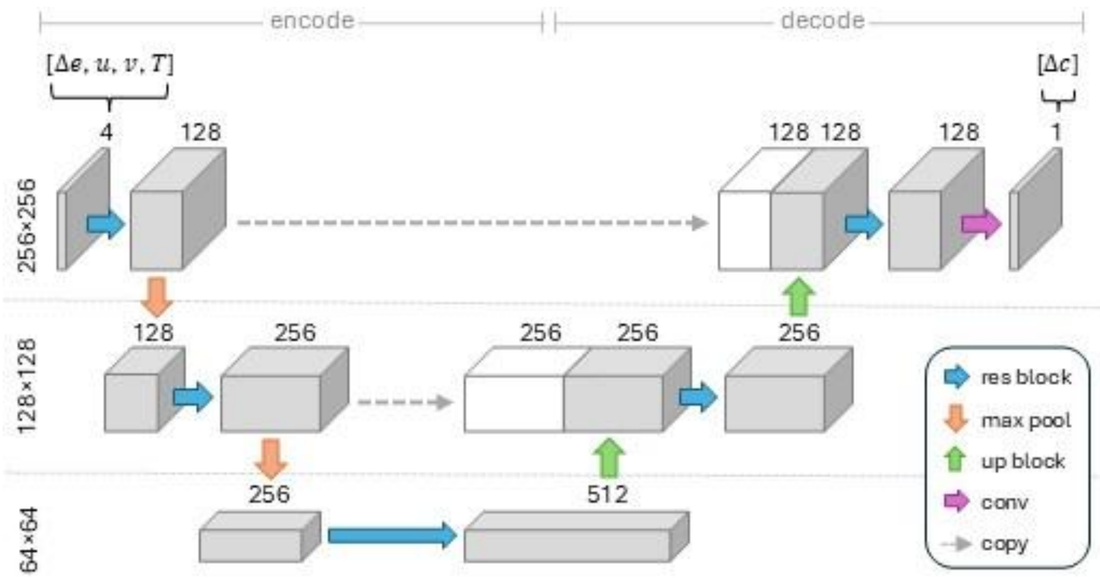
Ensemble Generation with AI Emulator

➢ **Ensemble Emulators**
- AI models can be trained using ensembles of forecasts to build an ensemble emulator
- Once trained, AI models can run very quickly
- Goal is to generate ensemble members faster and potentially have ensembles with larger sizes

➢ **Ammonia Surface Concentration Ensemble**
- A pre-existing ensemble of atmospheric ammonia (NH_3) concentrations that was used in an emissions inversion system was available to use for training
  - Produced by the GEM-MACH air quality model
  - NH_3 monthly emissions perturbed to yield perturbed hourly NH_3 atmospheric concentrations
  - 60 ensemble members

Ensemble Emulator for Atmospheric Chemical Constituents (EnsAI)

➢ **EnsAI**
- The ensemble of NH_3 surface concentrations was used to train an AI-based ensemble emulator, referred to as *EnsAI*
- Uses the U-Net convolutional neural network architecture
  - Inputs: emissions perturbation (Δe), surface horizontal winds (u and v), and surface temperature (T)
  - Outputs: surface ammonia concentration perturbation (Δc)
- Data sets formed for training/validation/testing:
  - Training set: Jan-Dec 2016
  - Validation set: Jul 2014
  - Test set: Jan-Dec 2015
- To be used for:
  - Generation of the ensemble used in NH_3 emissions inversions
  - Provides a starting point for further work generating ensembles for other chemical species (NO, NO_2, SO_2, etc…) to be used in data assimilation

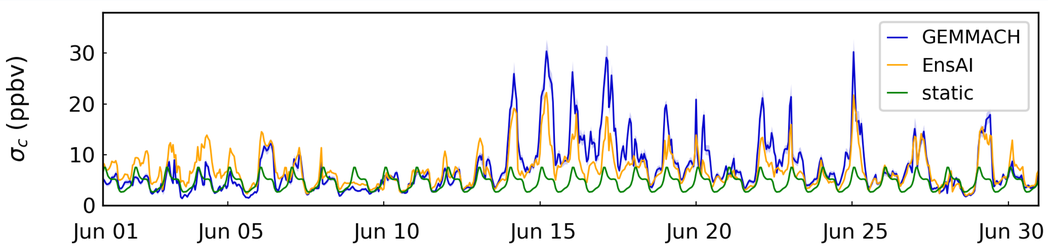


➢ **Training**
- 366 days × 24 h day^{-1} × 60 members = 527, 040 samples for the training set
- Mean squared error (MSE) loss function used for training
  - No unintentional smoothing observed
- Training took approximately 18 days through 50 epochs on a NVIDIA A100 GPU

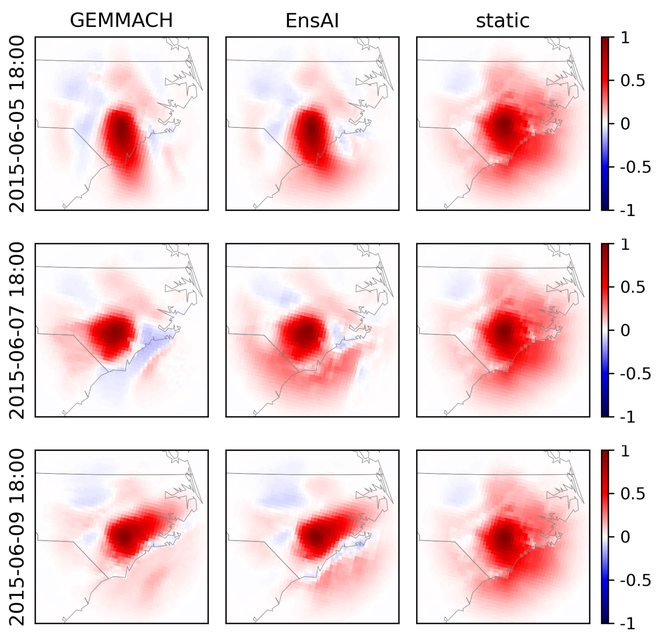
➢ **Runtimes**
- Wall time to generate a week's worth of simulated data per ensemble member:
  - 6.5 hour for GEM-MACH running on a CPU cluster using 720 computing nodes
  - 7 seconds for EnsAI running on a single GPU
- EnsAI runs about 3, 300 times faster than GEM-MACH

Results: Time-Dependent Forecast Error Covariances and Emissions Inversions

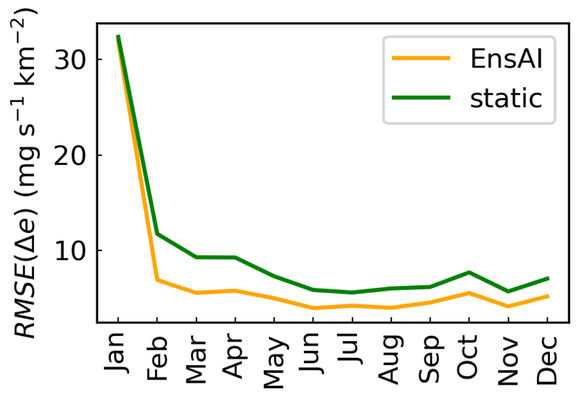
- A set of flow-independent covariances that are similar to those used at ECCC operationally (labeled 'static') are used as a reference when comparing ensemble covariances
- Covariances from EnsAI are consistently closer to the original GEM-MACH covariance than the 'static' covariances are
- NH_3 emission inversion results using the EnsAI ensemble are much closer to the inversions using the original GEM-MACH ensemble than when using the static covariances



surface concentration standard deviation at a location in North Carolina for June 2015



example of the emissions/surface concentration spatial correlations at different times



RMSE of NH_3 emissions increments between inversions made with the original GEM-MACH ensemble and inversions made with the EnsAI ensemble (orange) or with the static covariances (green)

Sitwell, Michael. "EnsAI: An Emulator for Atmospheric Chemical Ensembles." arXiv preprint arXiv:2504.16024 (2025).

5th ECMWF-ESA Machine Learning Workshop - April 13-17, 2026 - Bologna, Italy

Environment and Climate Change Canada / Environnement et Changement climatique Canada

Canada