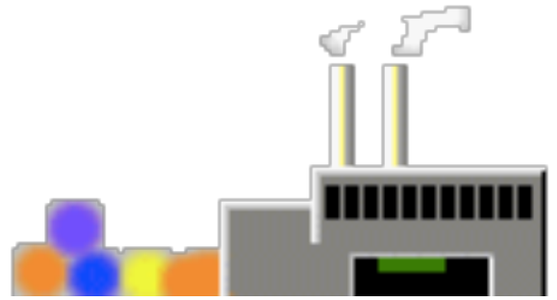
A recursive neural network chemical solver for fast long-term global simulations of atmospheric composition

Makoto Kelp & Daniel Jacob
AMS 20210113

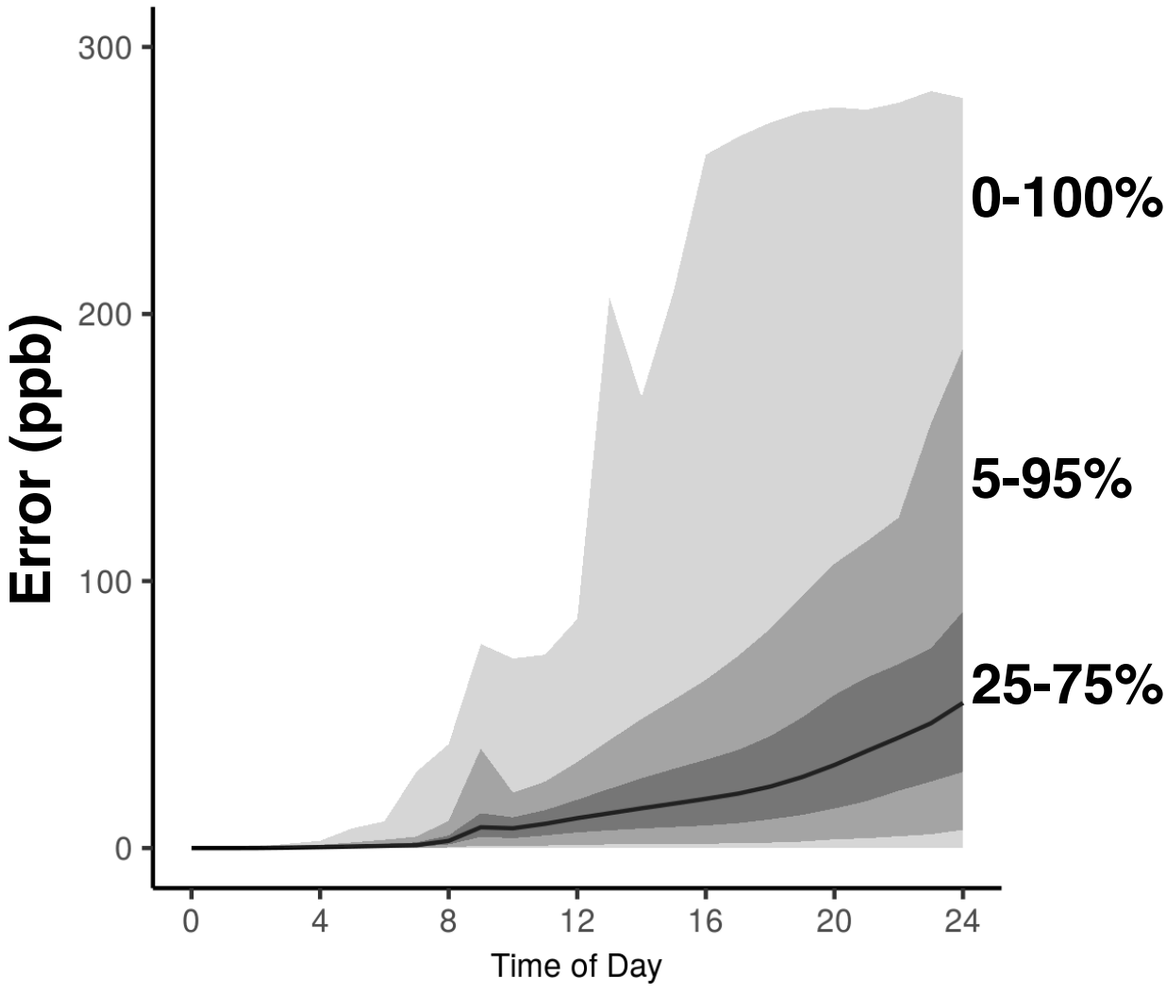


Application of machine learning to chemical solvers

Why	Why not
Chemical calculation is expensive	High dimensionality
Highly repetitive	Lack physics constraints
Fully deterministic	Error growth

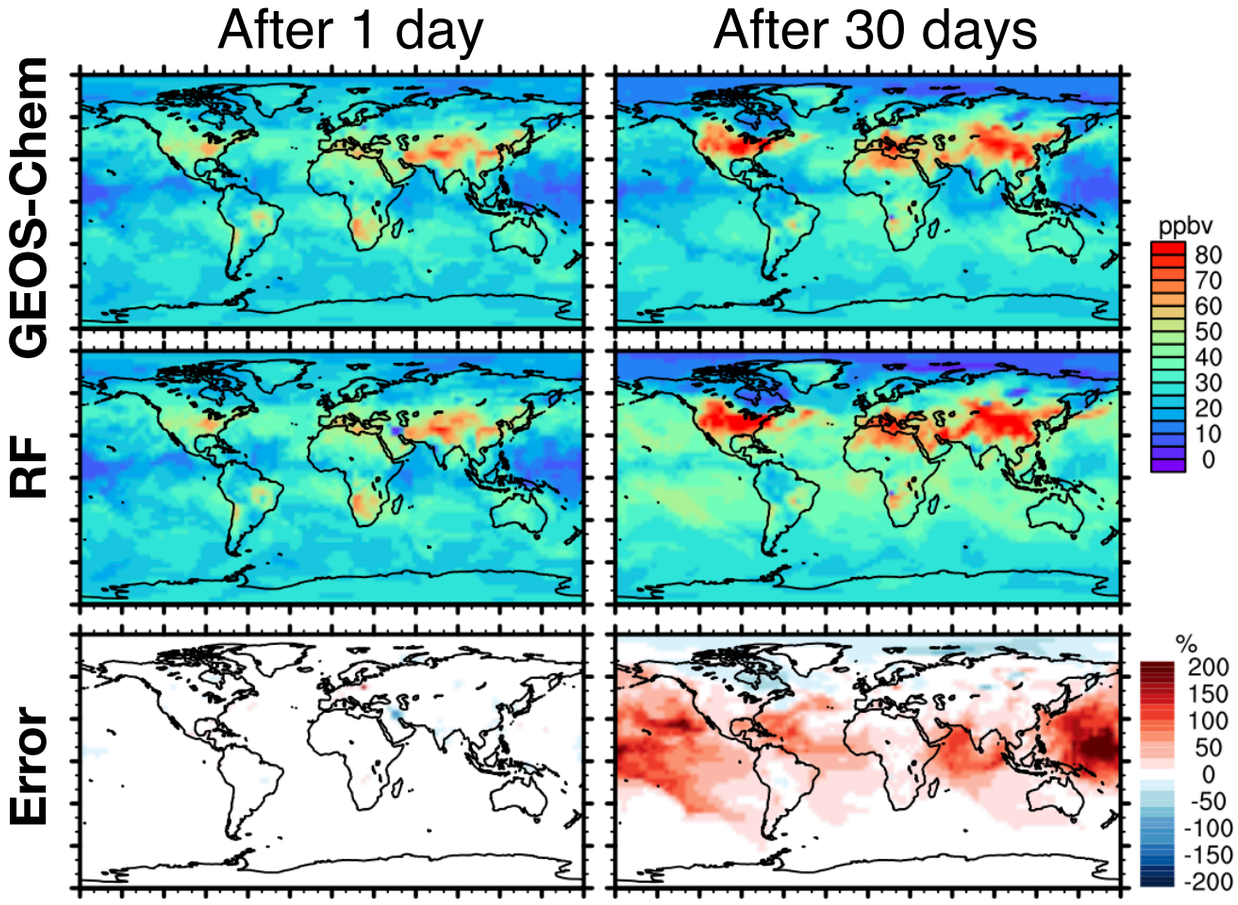
Past ML chemical solver attempts have been unsuccessful due to runaway error growth

CBM-Z box model: ozone



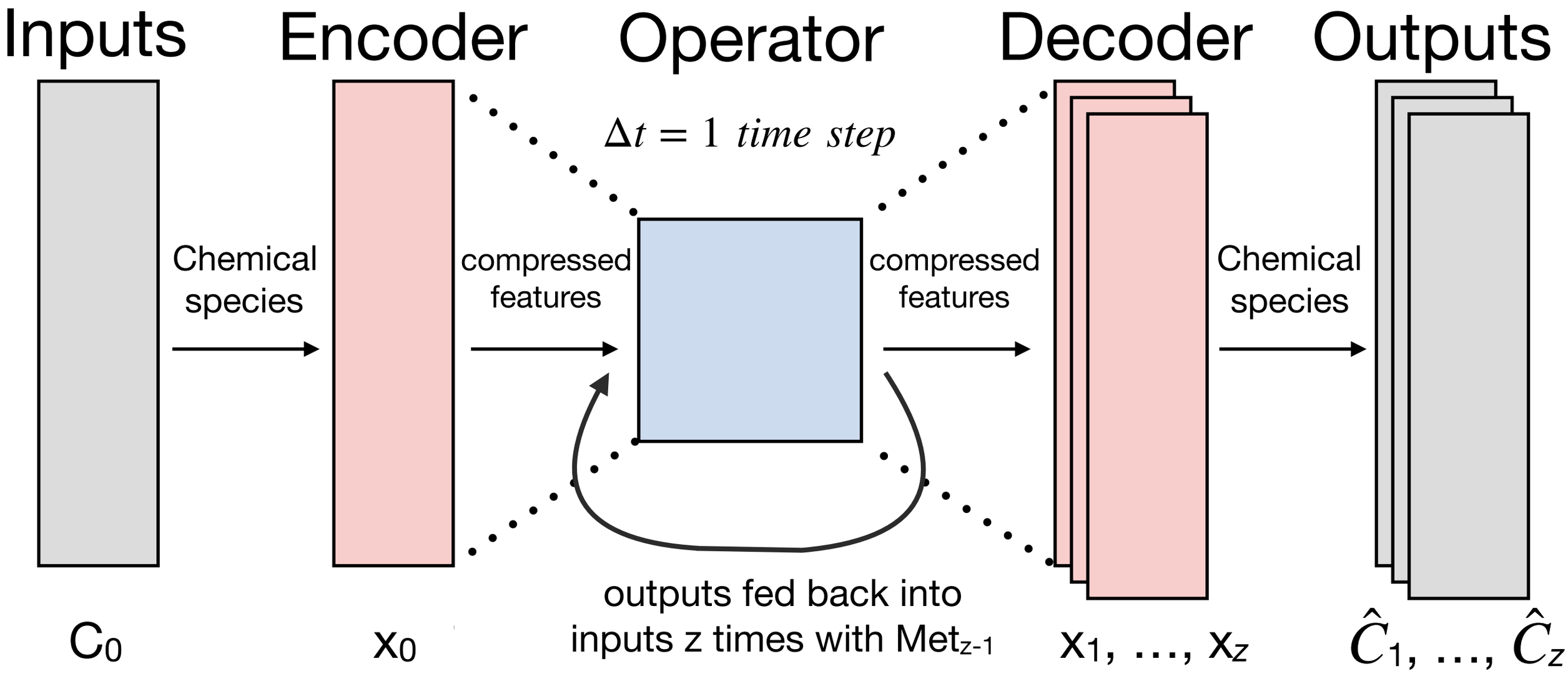
Dimensionality: 77 transported gas-phase
250 times faster, neural network
[Kelp et al. 2018, ArXiv]

GEOS-Chem: surface Ozone



Dimensionality: 51 transported gas-phase
85% slower, random forest
[Keller and Evans 2019]

New model framework 1) compresses dimensionality and 2) captures low-frequency chemical modes during ***training***



Mechanism: CBM-Z/MOSAIC Box model
101 species: 77 gas, 24 aerosol
4 meterological variables: T, P, RH, Solar angle

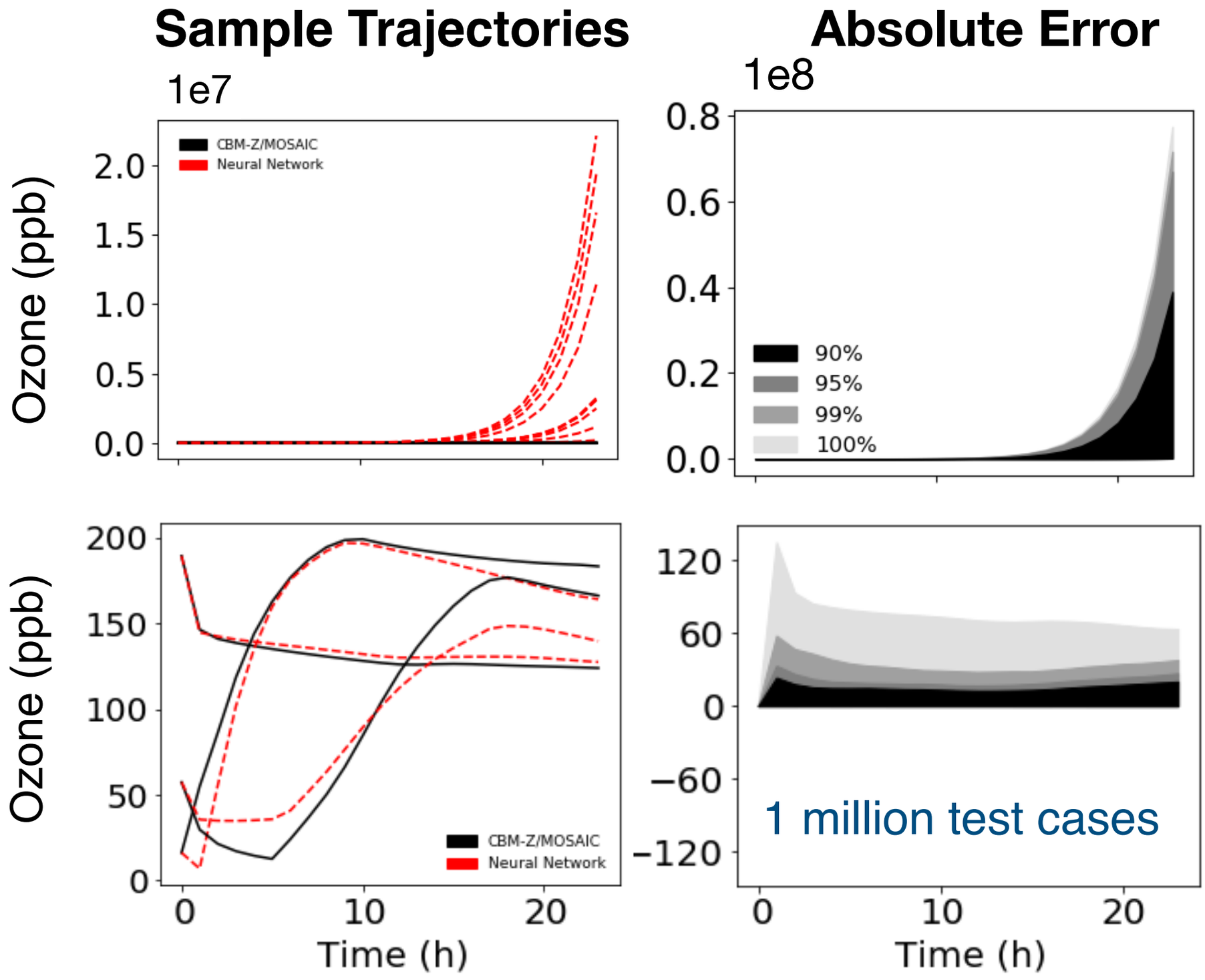
Kelp et al. [JGR 2020]

New ML model able to prevent error accumulation over time period of interest

Past Framework:

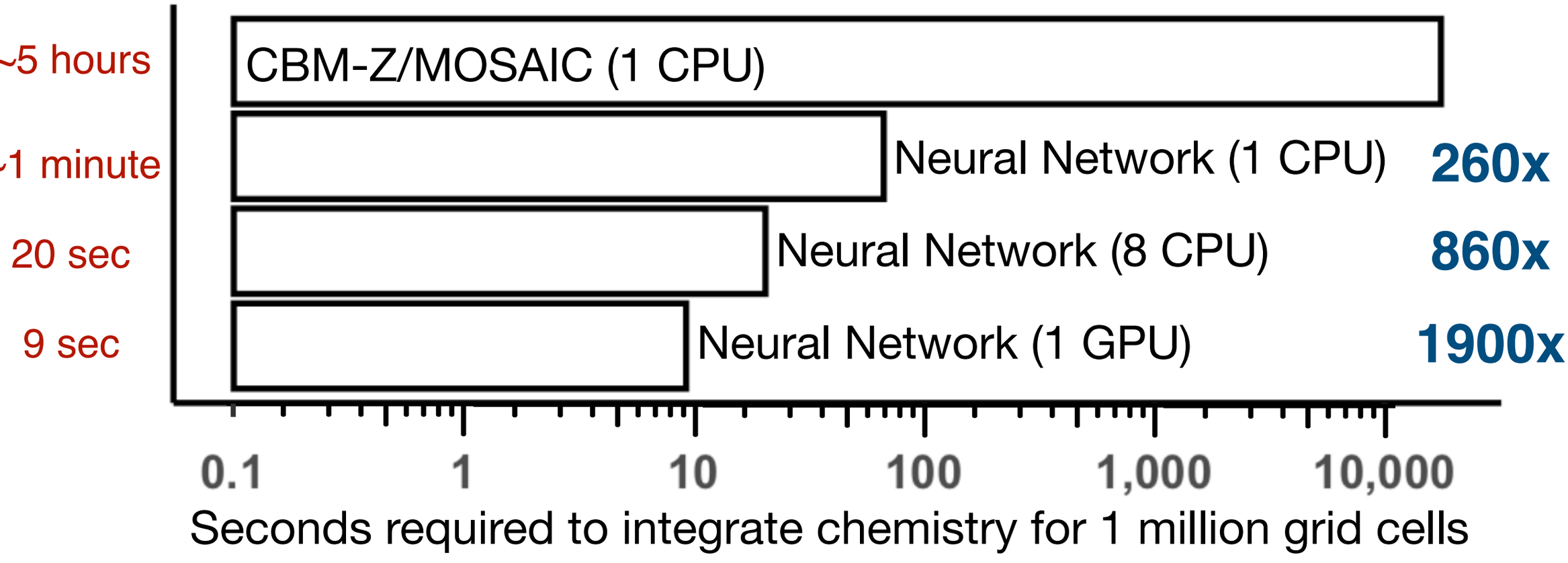
Keller and Evans, 2019
Kelp et al. 2018 preprint

Current Framework



Kelp et al. [JGR 2020]

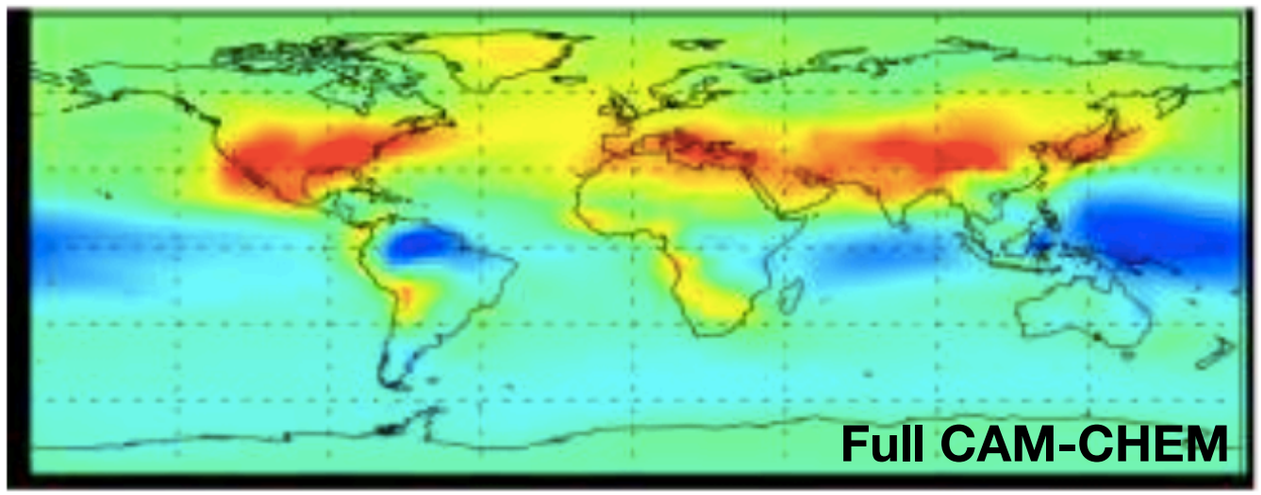
ML framework still achieves orders-of-magnitude speedup and can run operations on a GPU

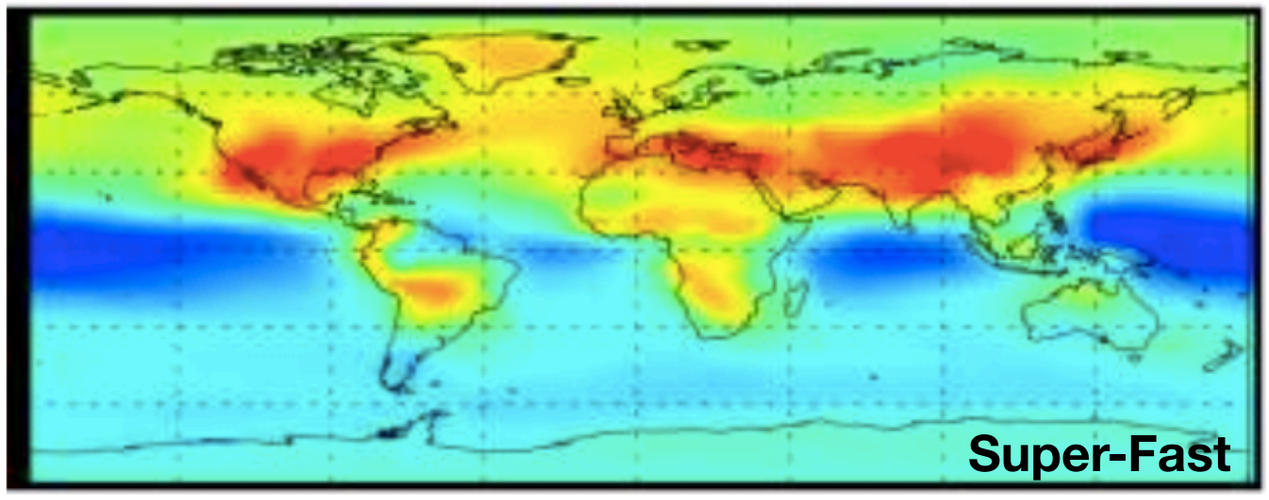


Kelp et al. [JGR 2020]

The ‘Super Fast’ chemical mechanism will allow us to better define ML methods and understand limitations in a full 3-D global modeling framework

Ozone





0 ppb *Cameron-Smith et al., 2006* 60 ppb

- Global mechanism with ~15 species [Brown-Steiner et al., 2018]
- Benchmarked in GEOS-Chem v12.0.0
- 4x5° resolution

Training Data:

1-hour chemical time step output

23 variables:

3 physical var: T, H_2O, air density

6 photolysis frequencies

14 gas-phase species

1 month dataset would contain:

lon×lat×lev×days×hours =

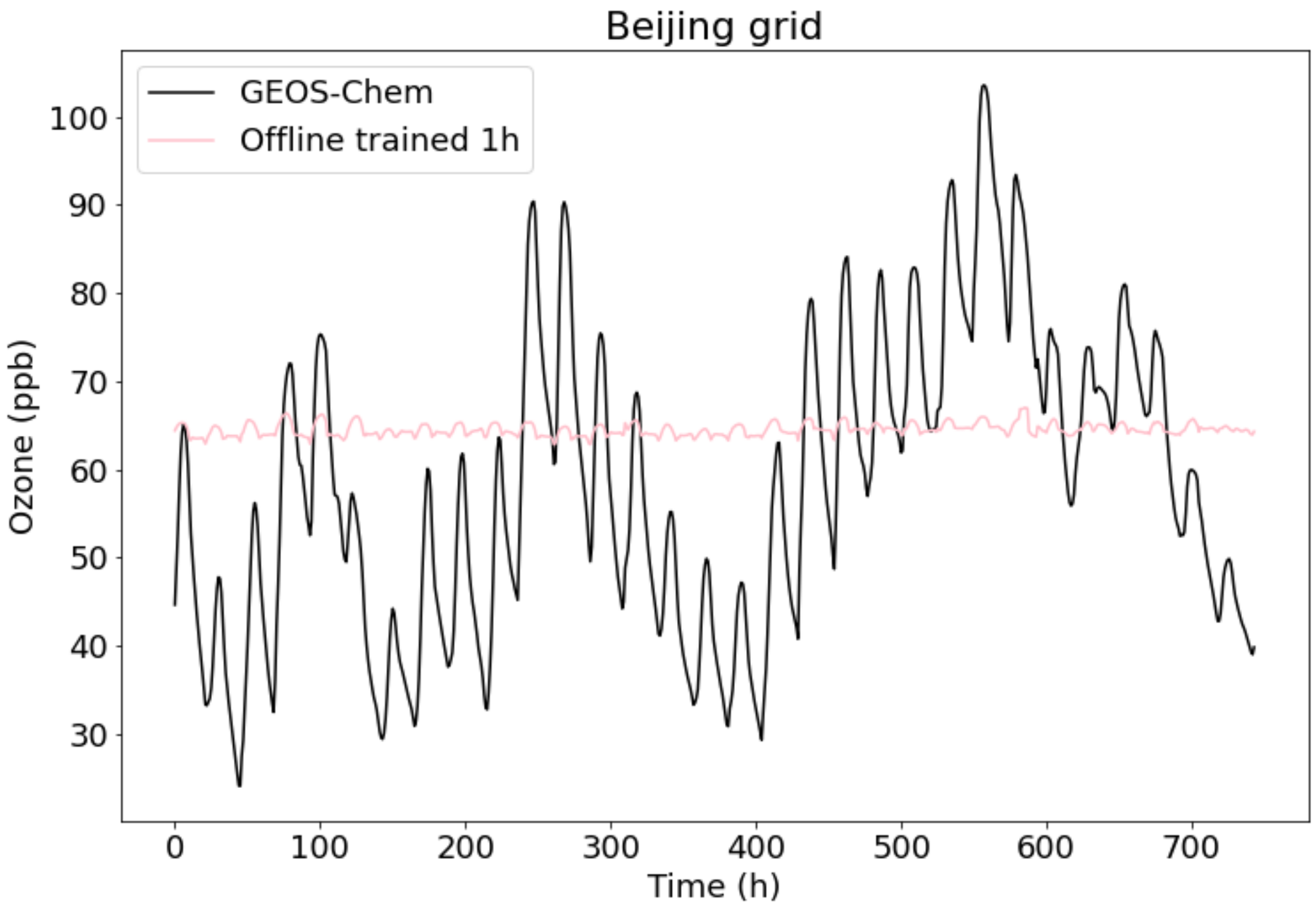
46×72×20×31×24 **->**~49 million samples

Trained on JJA of 2016

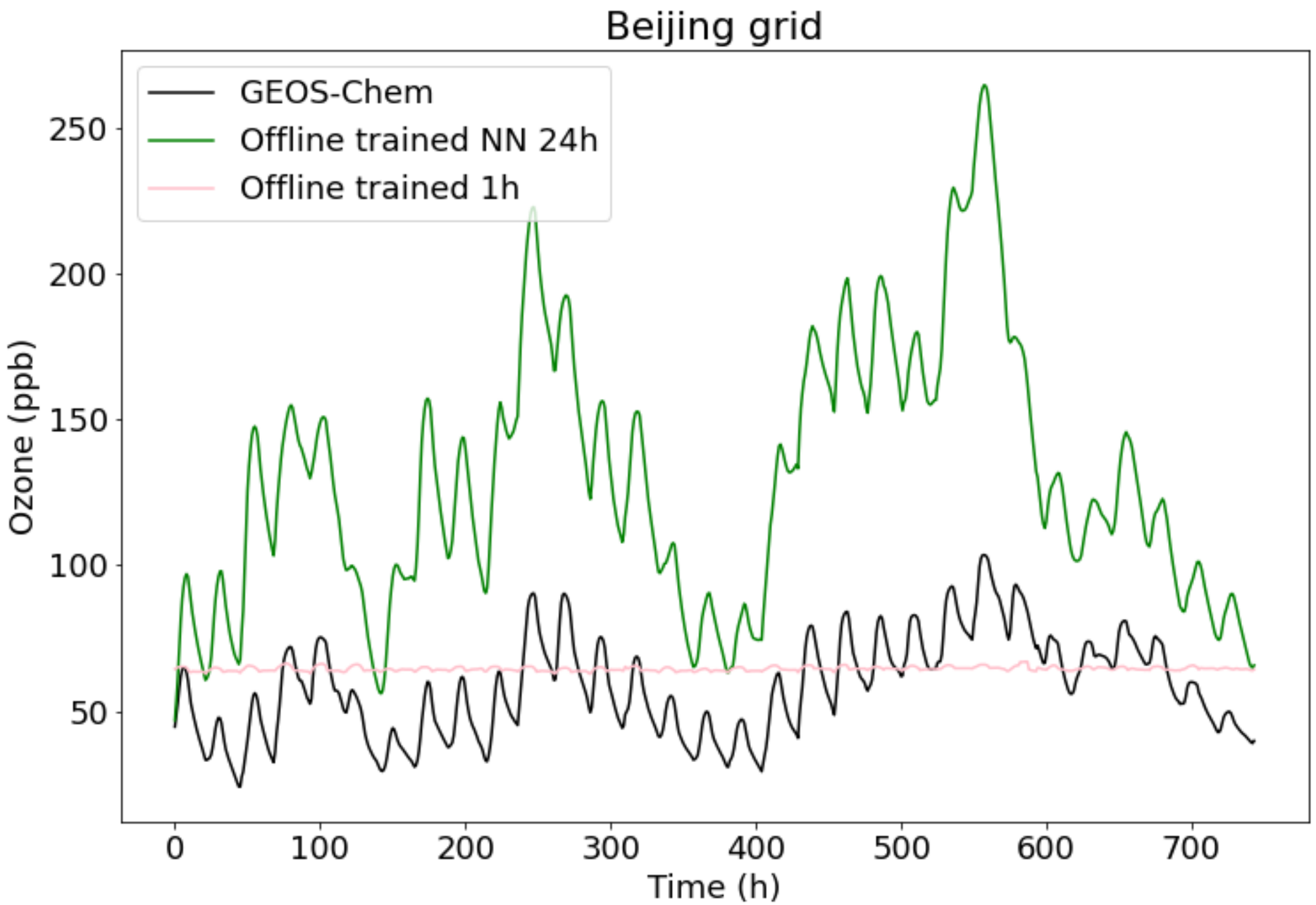
Testing:

Online testing in GEOS-Chem v12.0.0 for July 2017

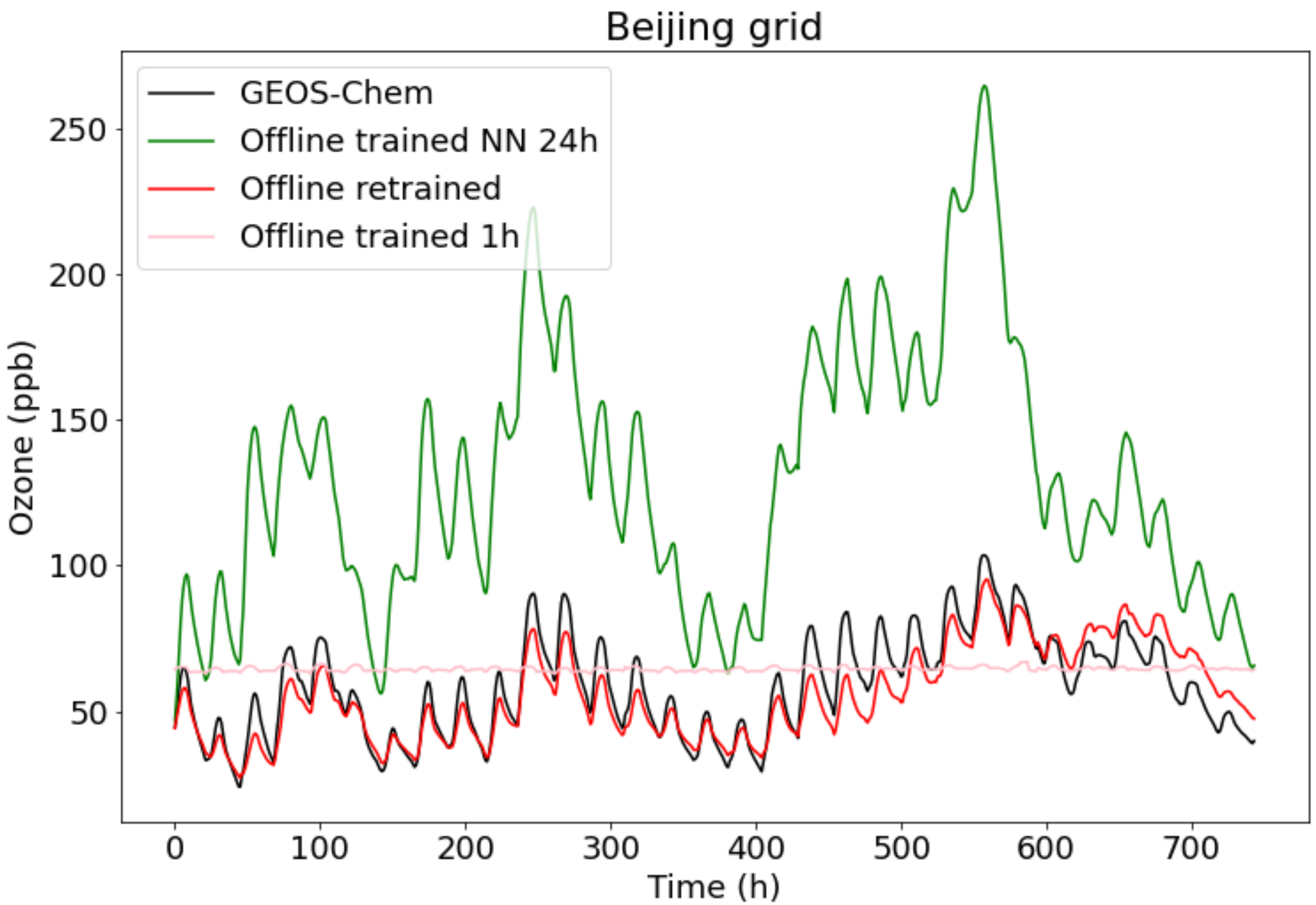
Offline 1h training leads to prediction of global average (naive approach)



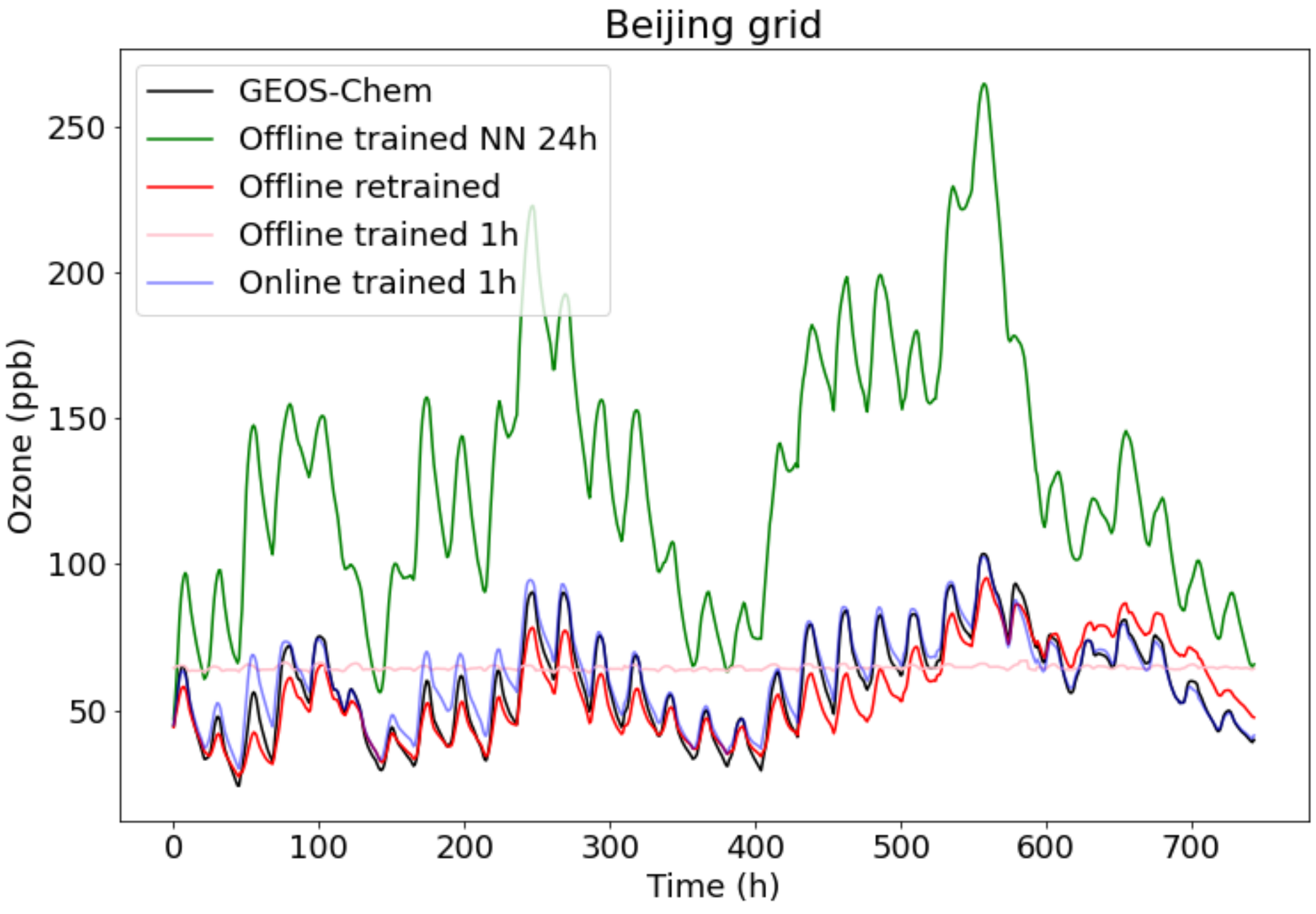
Offline 24h recursive training unable to dynamically account for operator splitting

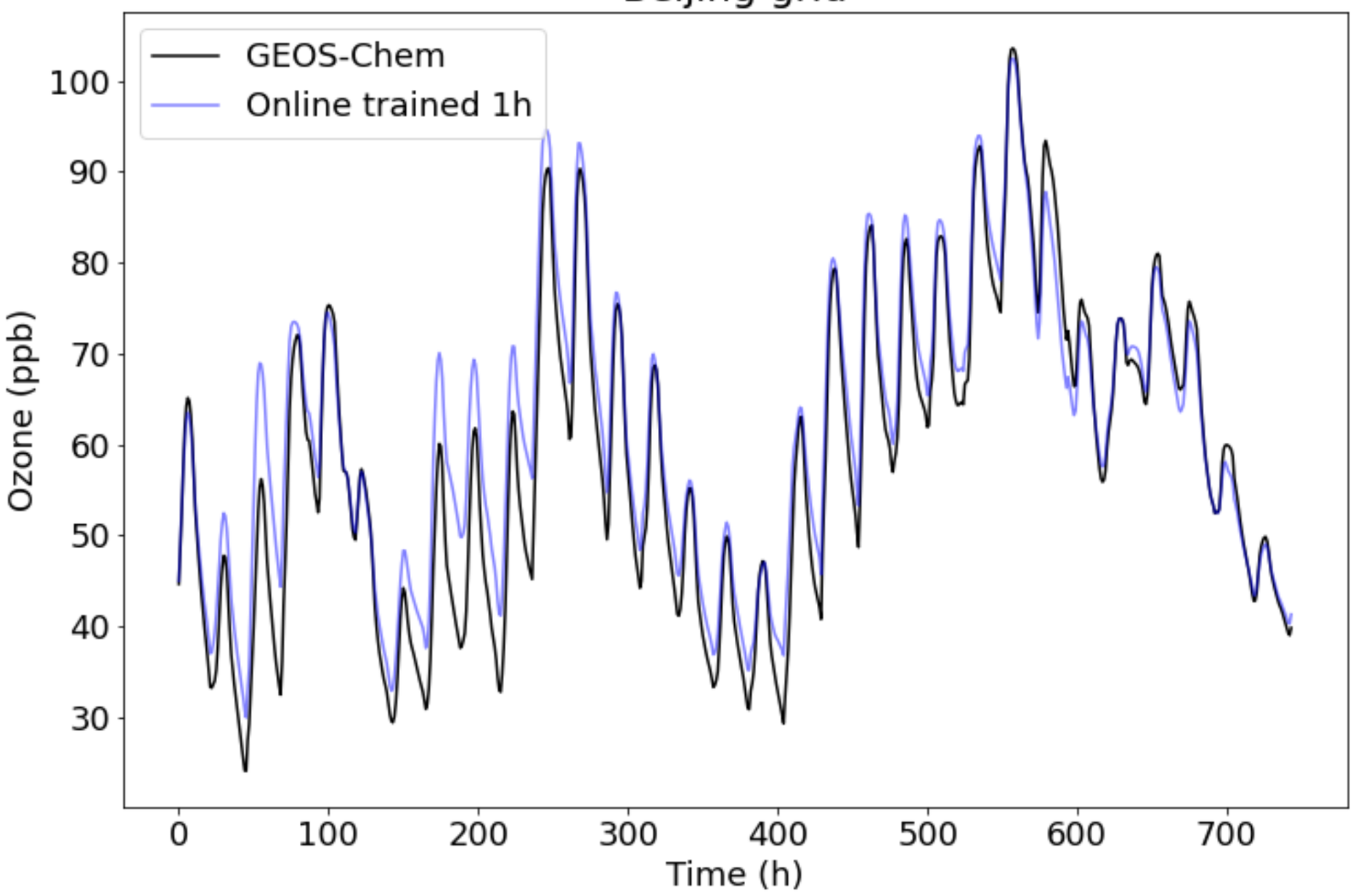


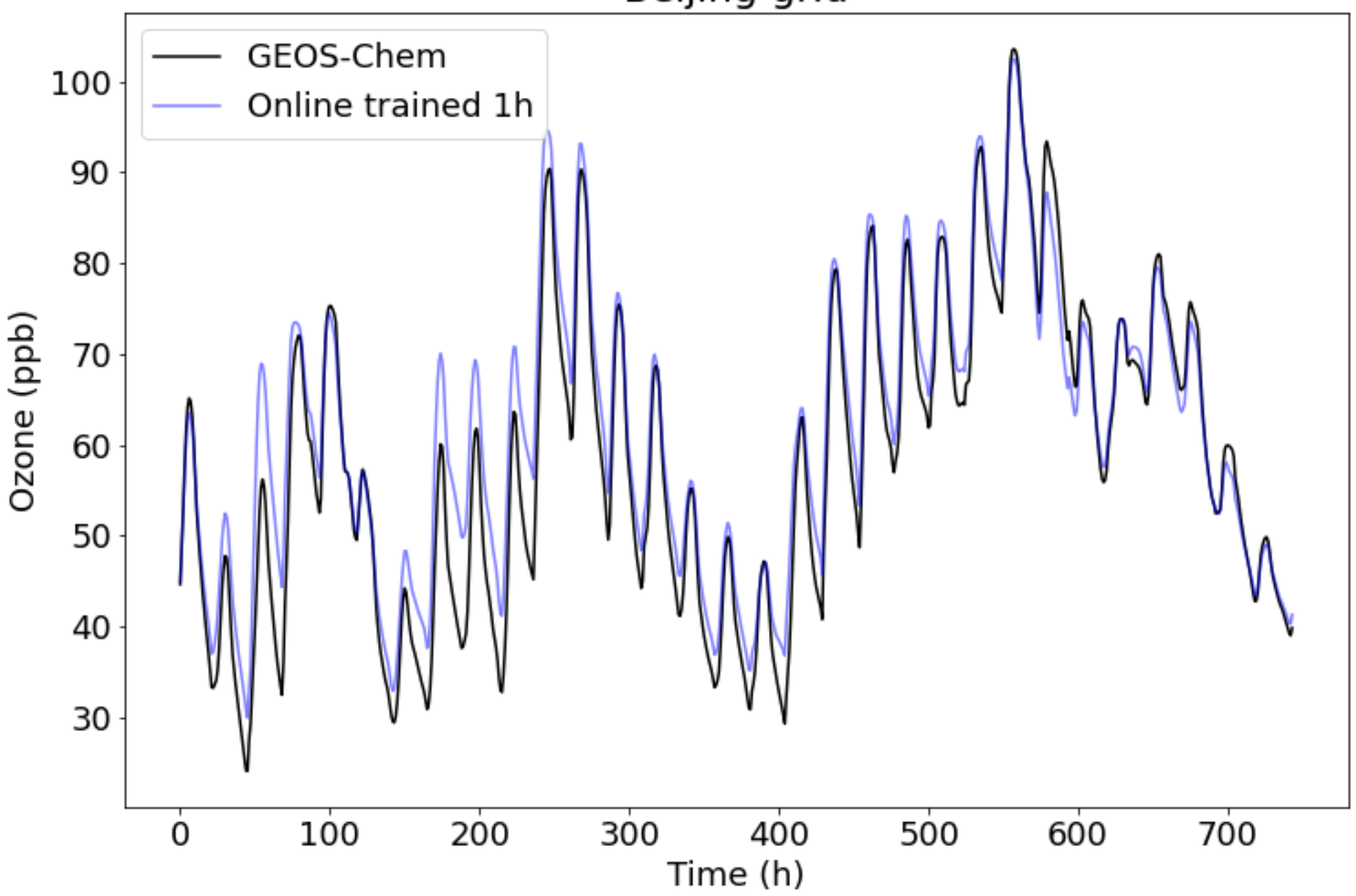
Offline 24h ML model re-trained online corrects model toward GEOS-Chem



Direct online training of ML model offers greatest promise

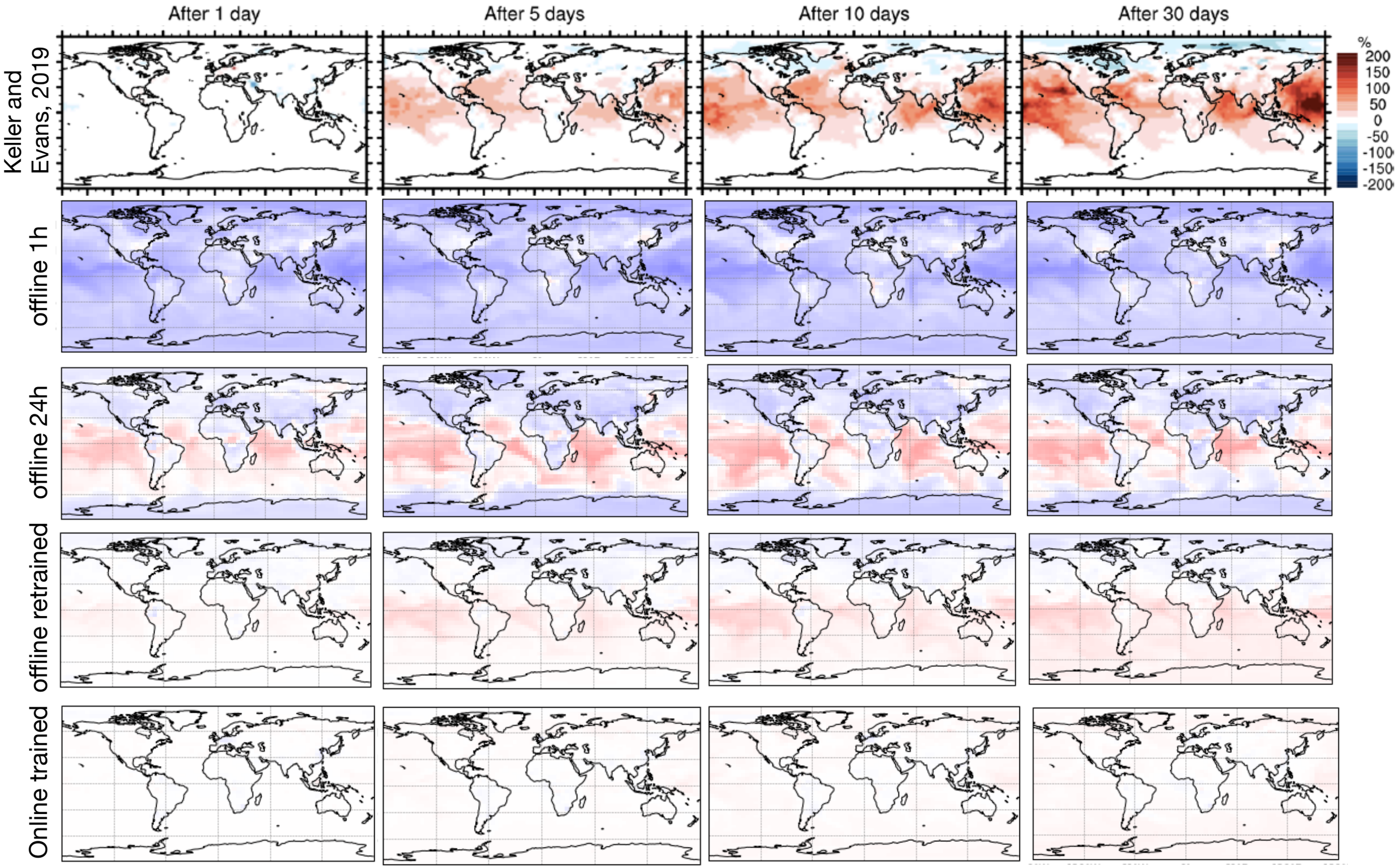



Beijing grid
GEOS-Chem
Online trained 1h
Ozone (ppb)
100
90
80
70
60
50
40
30
0
100
200
300
400
500
600
700
Time (h)

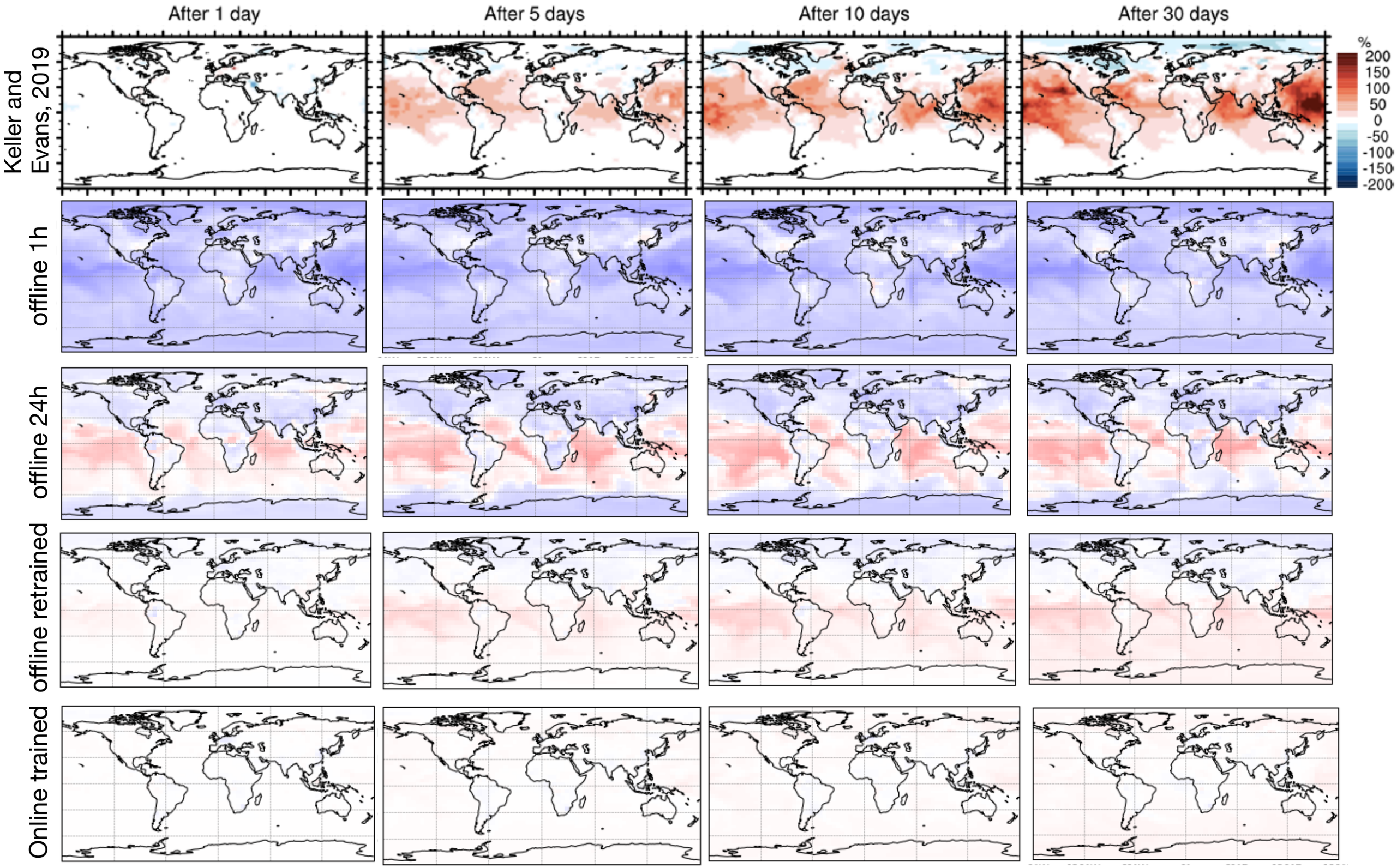


- No need to generate large offline data archives
- Incremental training allows us to:
  - Train on representative realizations
  - Avoid overfitting by training on future data
- Account for non-stationary distributions of data due to operator splitting

Error in simulating surface ozone in 30-day simulation



Error in simulating surface ozone in 30-day simulation



Offline training ≠ online performance

Next steps

-Quantify error for longer time scales (1 year, all 4 seasons)

-Achieve similar performance for all species in Super Fast mechanism

-Apply to full GEOS-Chem mechanism

-Implement into GEOS-CF for short term forecasting and data assimilation