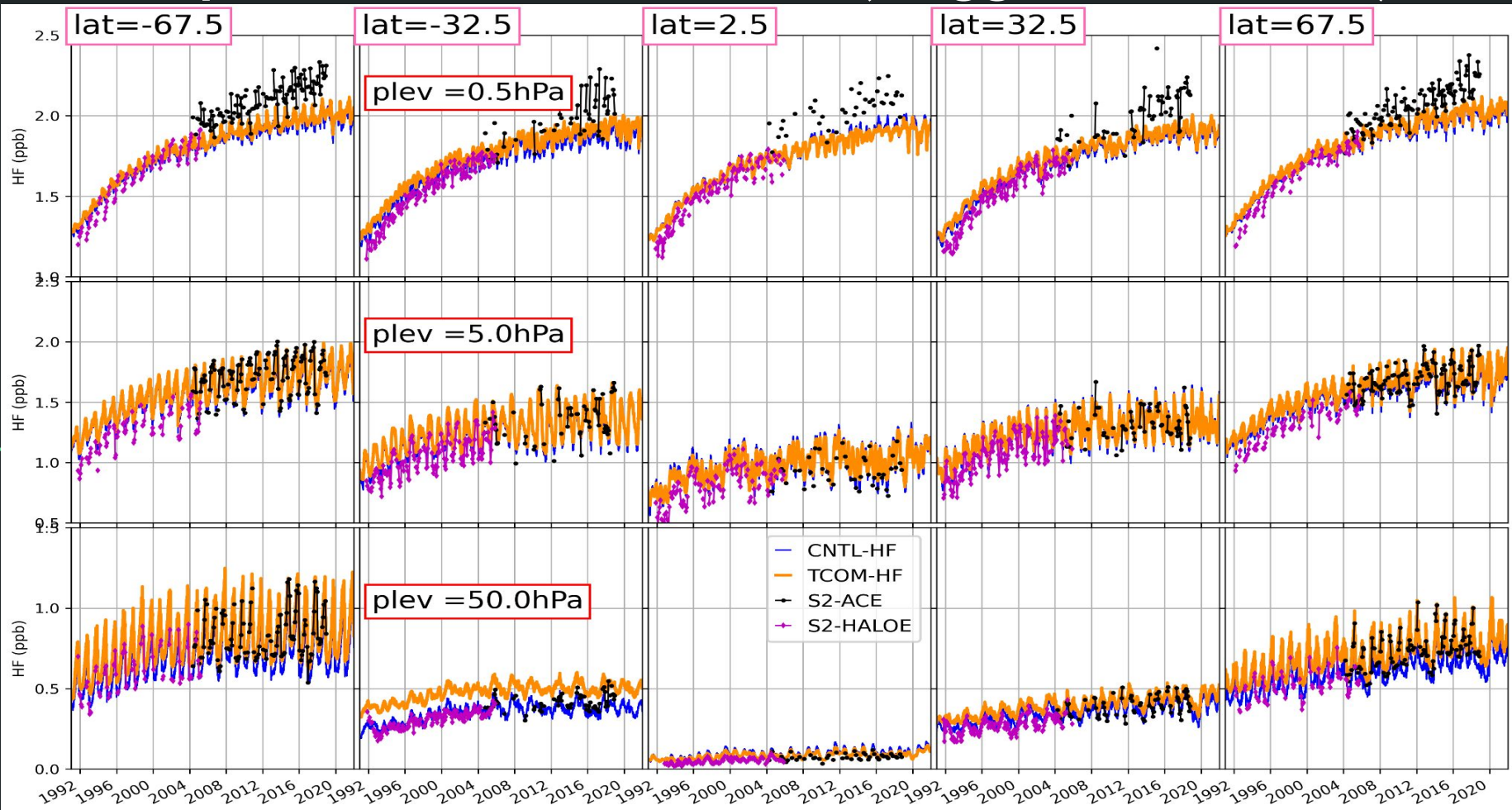


Comparison with SPARC data (Hegglin et al., 2021)



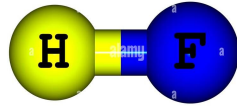
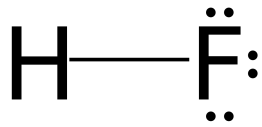
TCOM-HF : Daily global gap-free stratospheric hydrogen fluoride (HF) profile data set based on TOMCAT CTM and Occultation Measurements

Sandip Dhomse (University of Leeds, Leeds, UK)

s.s.dhomse@leeds.ac.uk

Why do we need long-term data sets

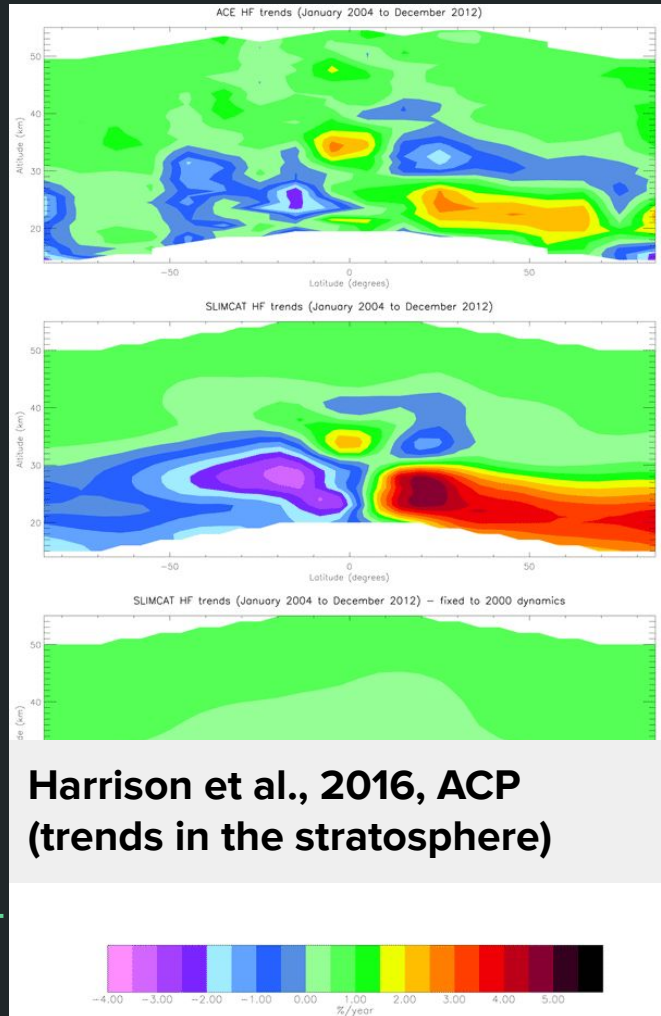
1. HF distribution better insight about the CFC, HCFC & HFC loss processes → anthropogenic sources
2. Satellite missions - shorter duration/different methodologies
3. In-situ Observations - limited spatial coverage
4. Chemical models : biases (parameterisations)
5. But can run CTM forced with reanalysis data representing our best knowledge about the past
6. Machine learning model is trained for the model/observational data differences so that we can estimate corrections needed for all model grids



Troposphere : Largest sources are volcanos : eg. Mt Etna

Stratosphere : About 20,000 tonnes/year from degradation of CFCs, HFCs, HCFCs

Removal: Wet & dry deposition



Satellite data from two solar Occultation instruments

- **ACE-FTS (2003–2021)**
- **HALOE (1991-2005)**

Model simulation

- TOMCAT - detailed stratospheric chem.
- Forced with ERA5.1
- LBC from WMO2022

For HF chemistry in TOMCAT: See
Harrison et al., 2016, ACP

Here we use XGBoost Regression model
(assuming all the model-observation eros
can be linked to model tracers)

$dHF = O_3 + CH_4 + N_2O + HF +$
 $ClONO_2 + PV + temp + \dots$ (13 tracers
from TOMCAT)

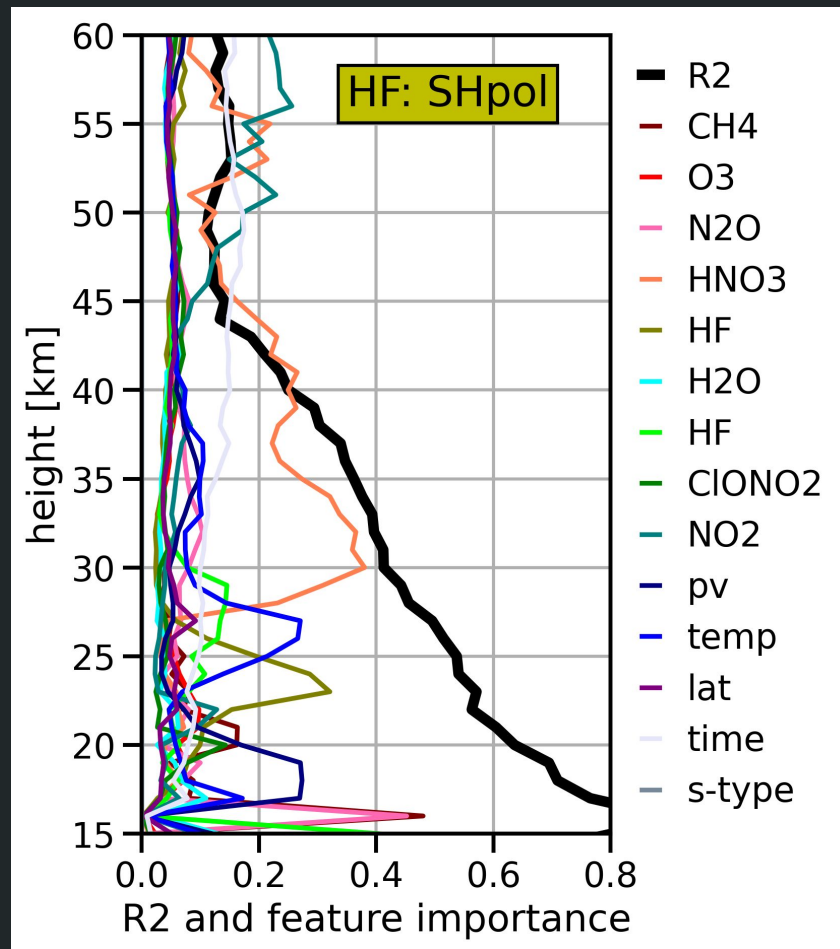
Methodology

1. Sample model output at measurement collocation
2. Calculate model - observation differences
3. Train machine learning algorithm for the a given altitude and latitude range
4. Get the correction coefficient and apply to all the grid points

70% training and 30 % testing

XGBoost shows best performance in
the lower -middle stratosphere

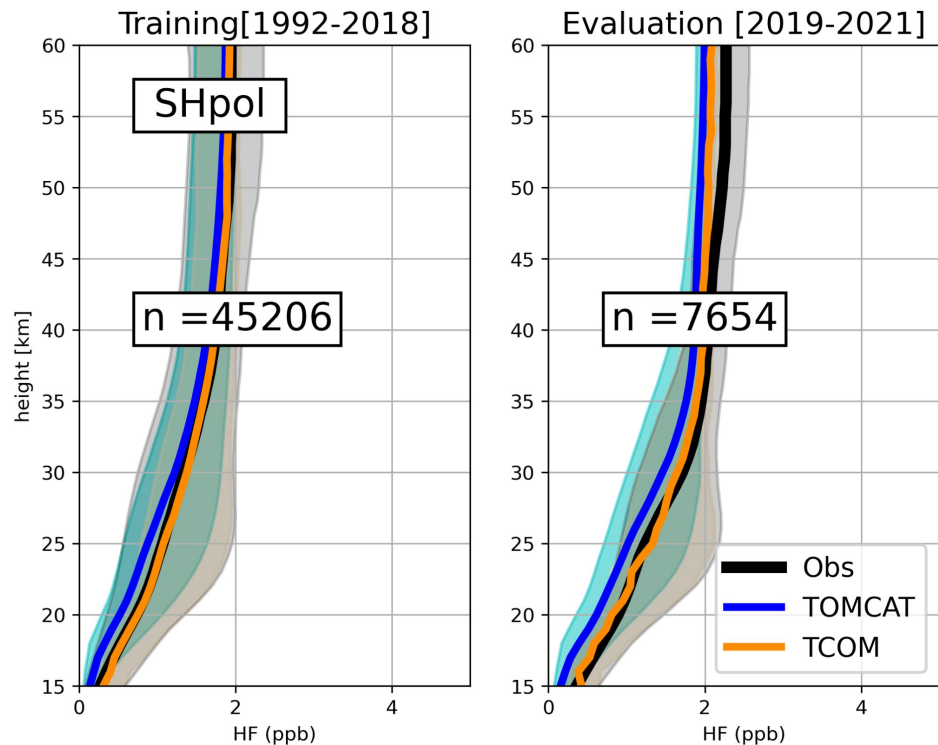
(R2 is shown for the testing data)

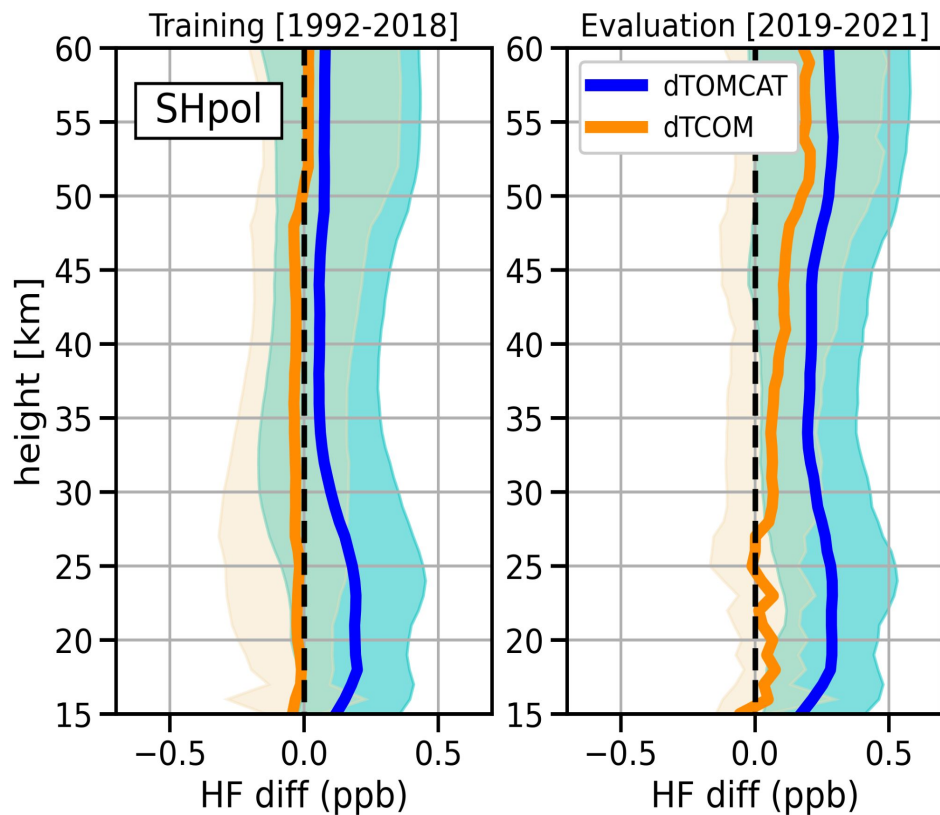


Median profiles (Training vs Evaluation data)

Good performance for the
evaluation period (2019-2021)

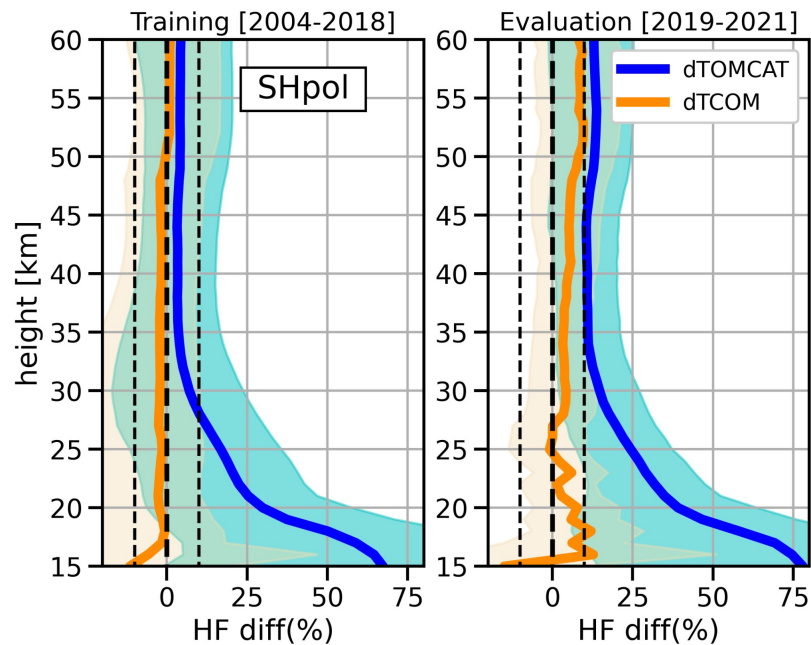
Shaded region indicate 10 & 90 percentiles



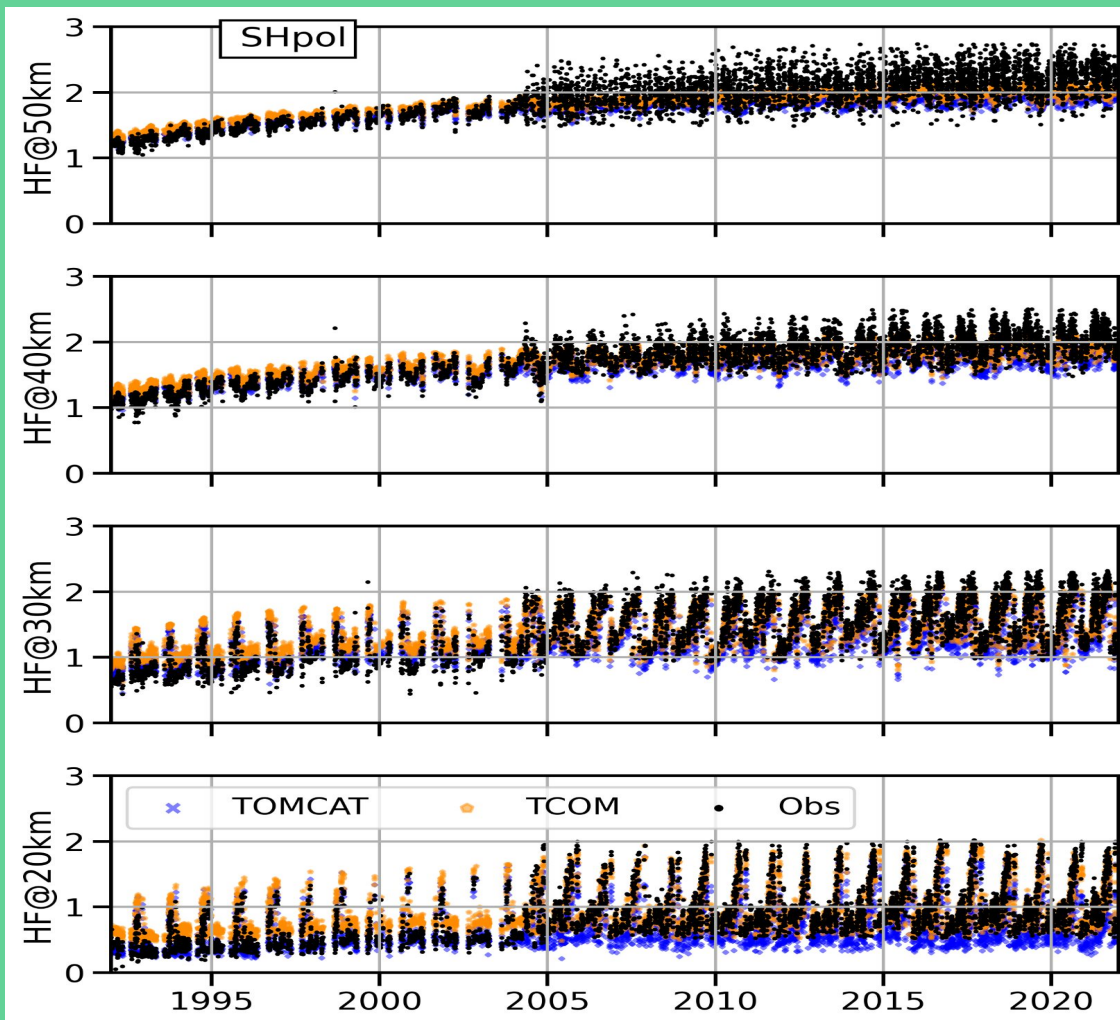


Absolute
differences

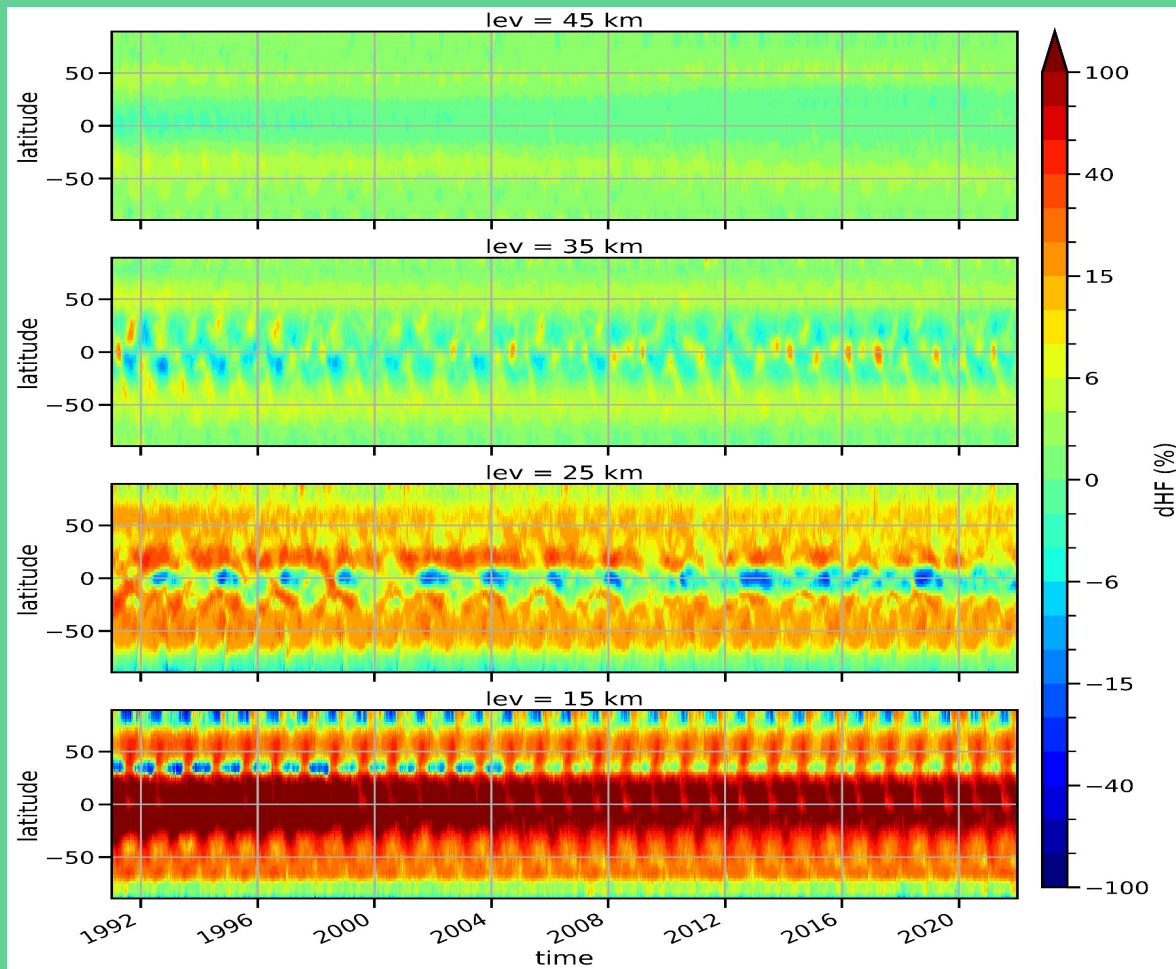
Percent
differences



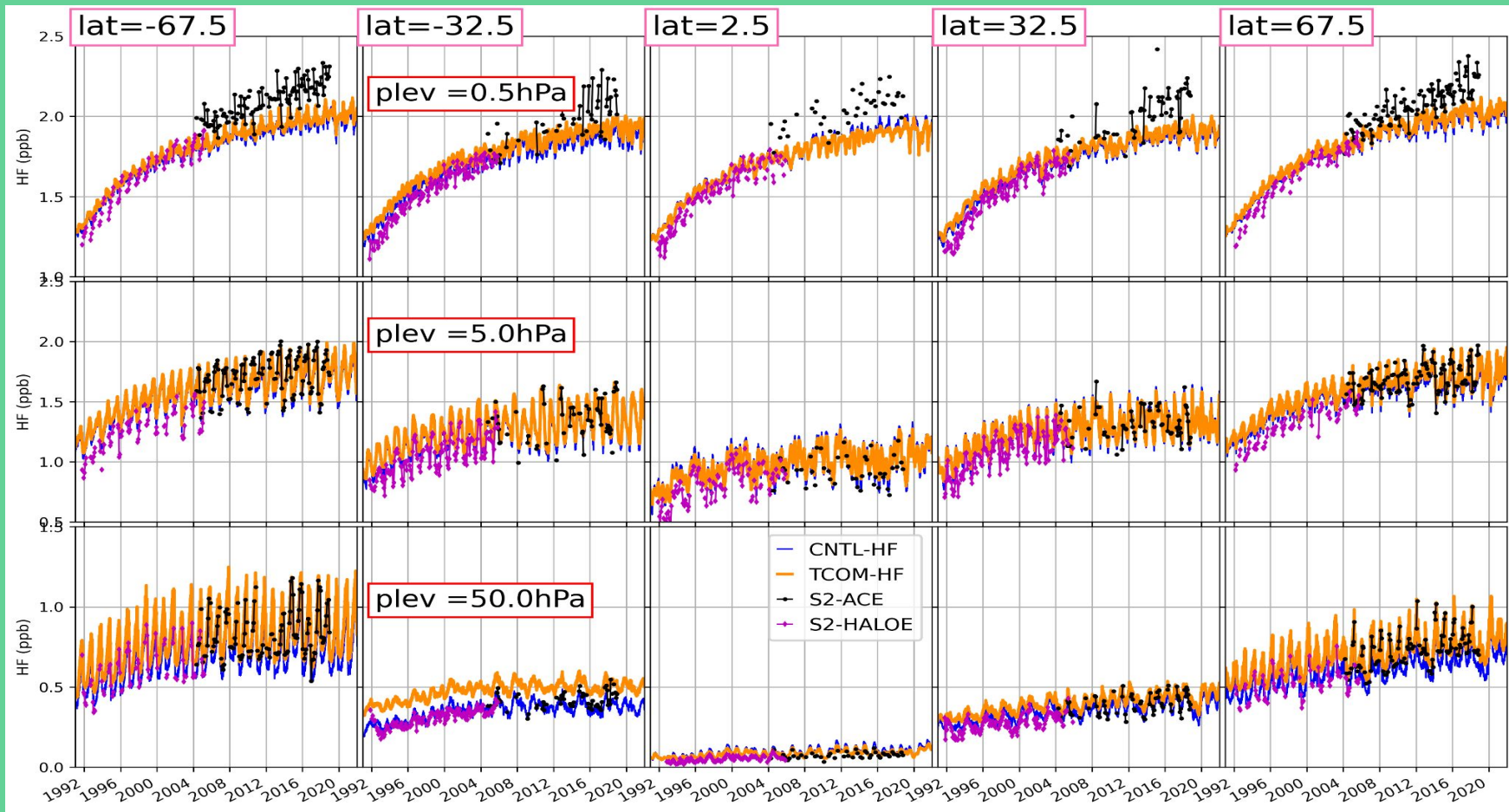
Comparing evaluations at difference height levels



Diff. (in %) between TCOM & TOMCAT HF



Comparison with SPARC data set (Hegglin et al., 2021)



Conclusion

- We have constructed long-term gap-free HF data set
- Data is based on HF profile measurements by HALOE and ACE satellite instruments and TOMCAT CTM output
- We use XGBoost regression model to estimate correction terms needed for TOMCAT output
- Evaluation data set suggests errors are less than 10 % throughout the stratosphere