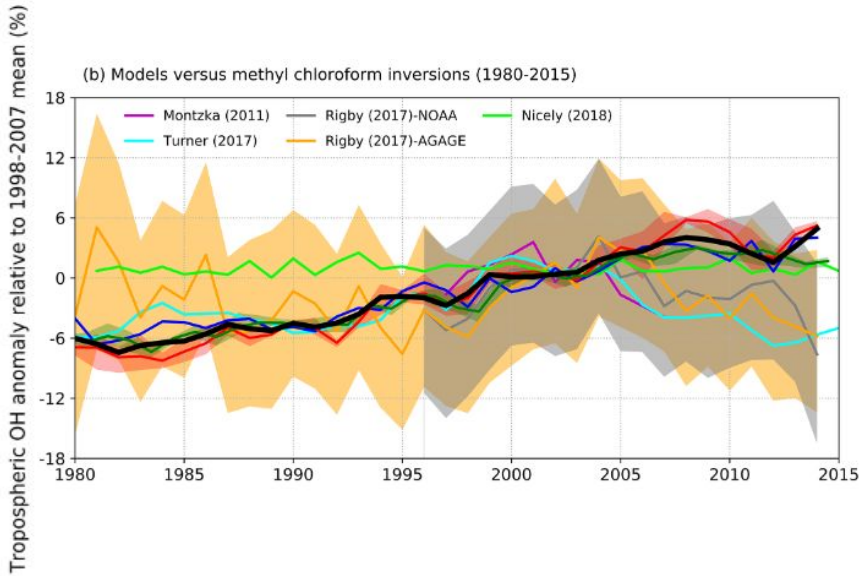
Can simple machine learning methods predict concentrations of OH better than state of the art chemical mechanisms?

Seb Hickman, Paul Griffiths, James Weber and Alex Archibald
Centre for Atmospheric Science, University of Cambridge

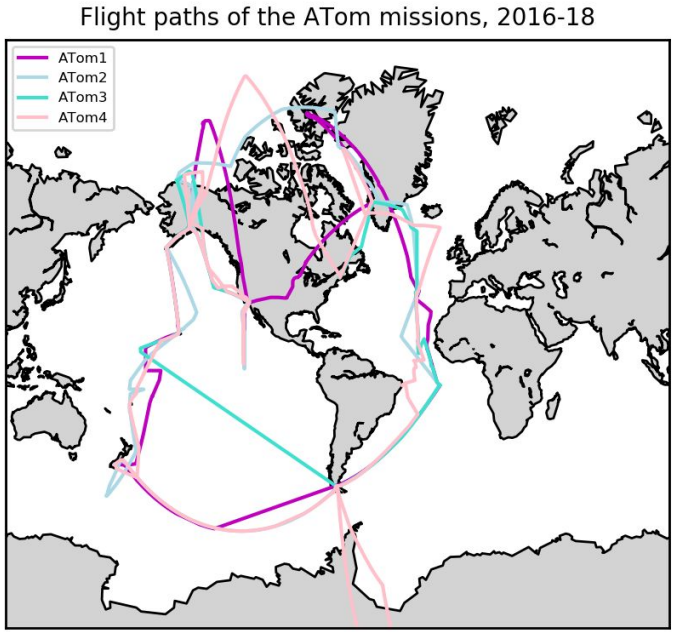
Why are we interested in OH?



- OH is the dominant oxidising agent in the troposphere
- Crucially, this means that it controls the lifetime of many important atmospheric species, particularly **methane**, a key greenhouse gas
- Chemistry-climate model uncertainty in global OH is significant

Stevenson et. al, 2020

NASA ATom Flights



- Flights sampled each season
- The flights constrain uncertainty in chemical composition and reactivity of the remote troposphere by making measurements on a vast scale
- We assimilated a dataset focussed on OH, consisting of ~40,000 individual **snapshots** of the atmosphere.

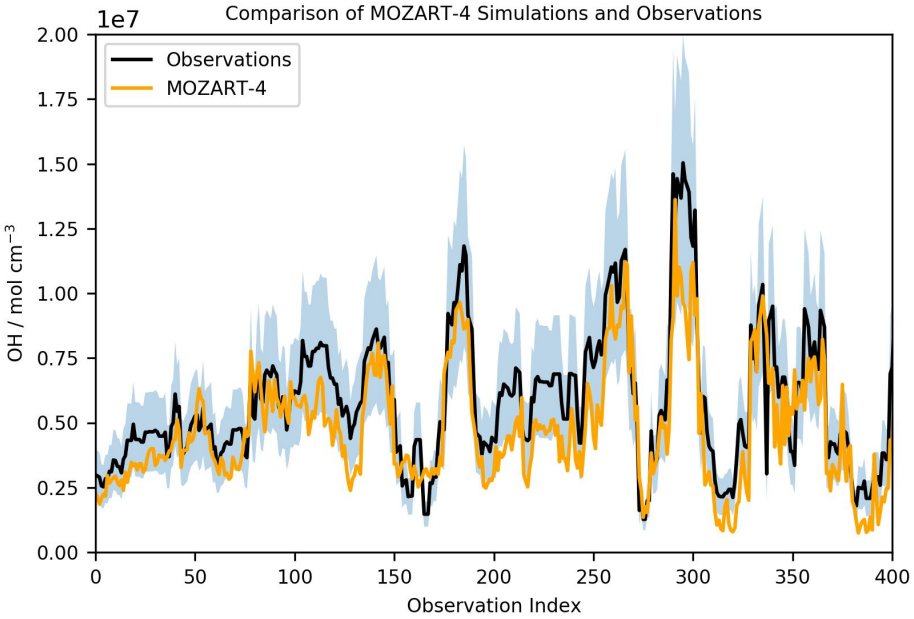
Simulating OH with a chemical process model?

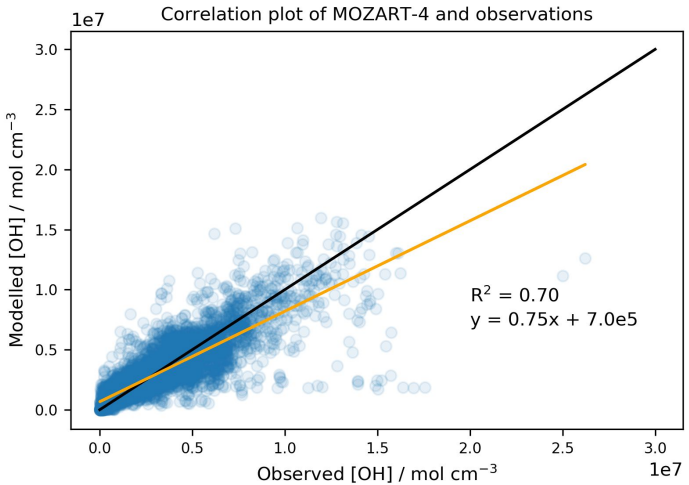
Challenges:

1. OH is controlled by a vast number of chemical species and environmental conditions on a short timescale
2. How do we evaluate our mechanisms with this unusual high speed aircraft data?

- We need a forward model to simulate this…**but**
- Including all species in an explicit chemical mechanism is expensive

MOZART-4 is good…but not perfect!

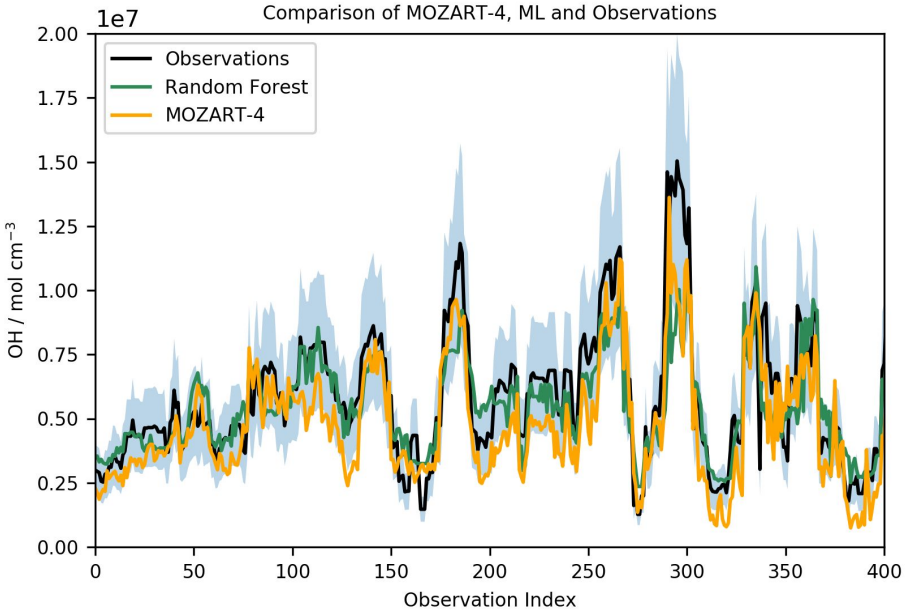


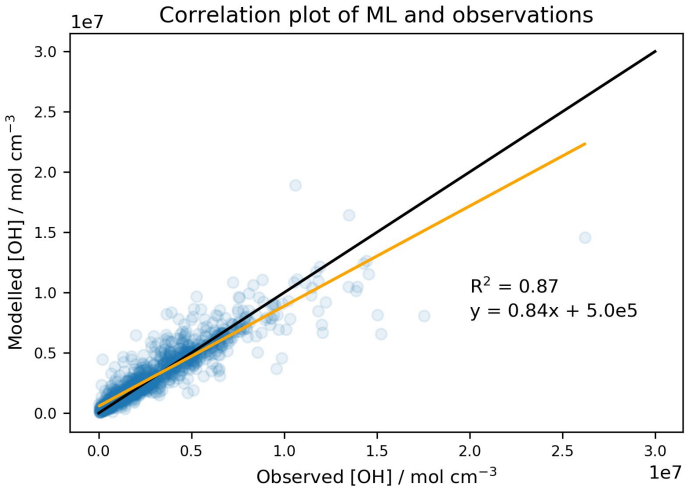


- BOXMOX, MOZART-4 chemical mechanism, initialised with in-situ data from ATom
- Model integrated forward for 20 mins with a 1s timestep.

- This low bias is consistent with an under-represented source at high OH
- Very tricky to diagnose this problem in the chemical model

Can simple machine learning models reproduce OH?

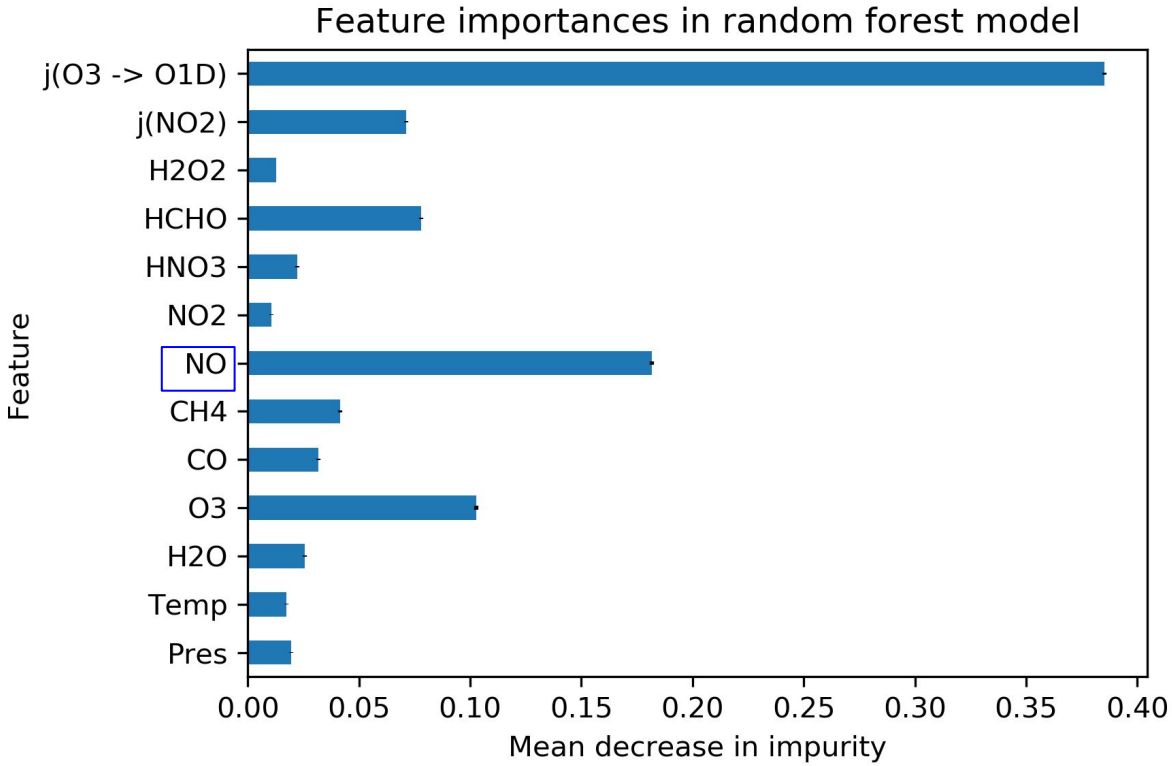




- Random forest model trained, validated and tested on ATom data
- 13 inputs from ATom measurements

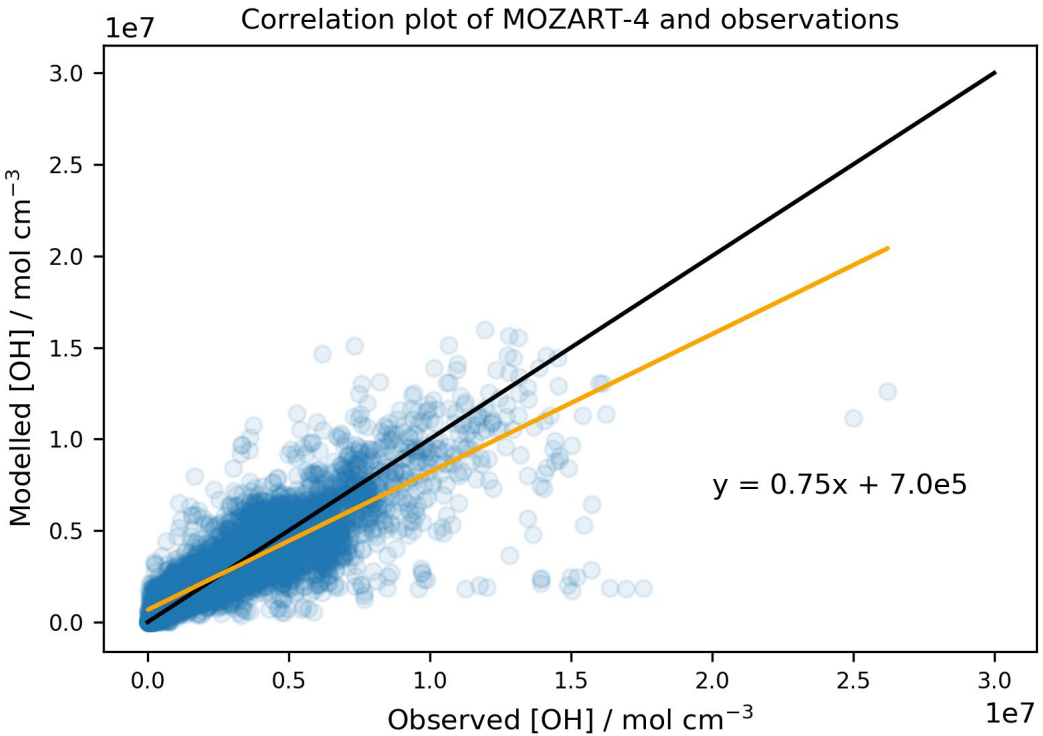
- Random forests are better able to reproduce OH
- These models generalise reasonably across flights
- Show a smaller low bias at high OH

Can we learn anything from these data-driven methods?

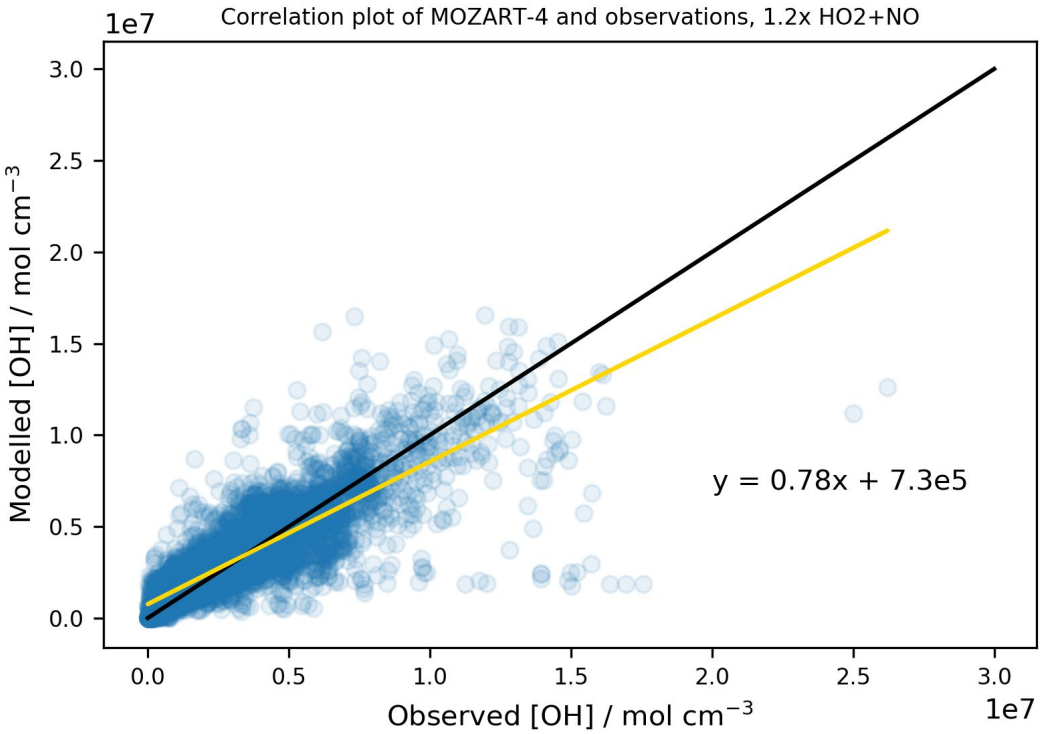


- Across ML methods, **nitric oxide** was a key feature
- This prompted us to think more about nitric oxide in our chemical model

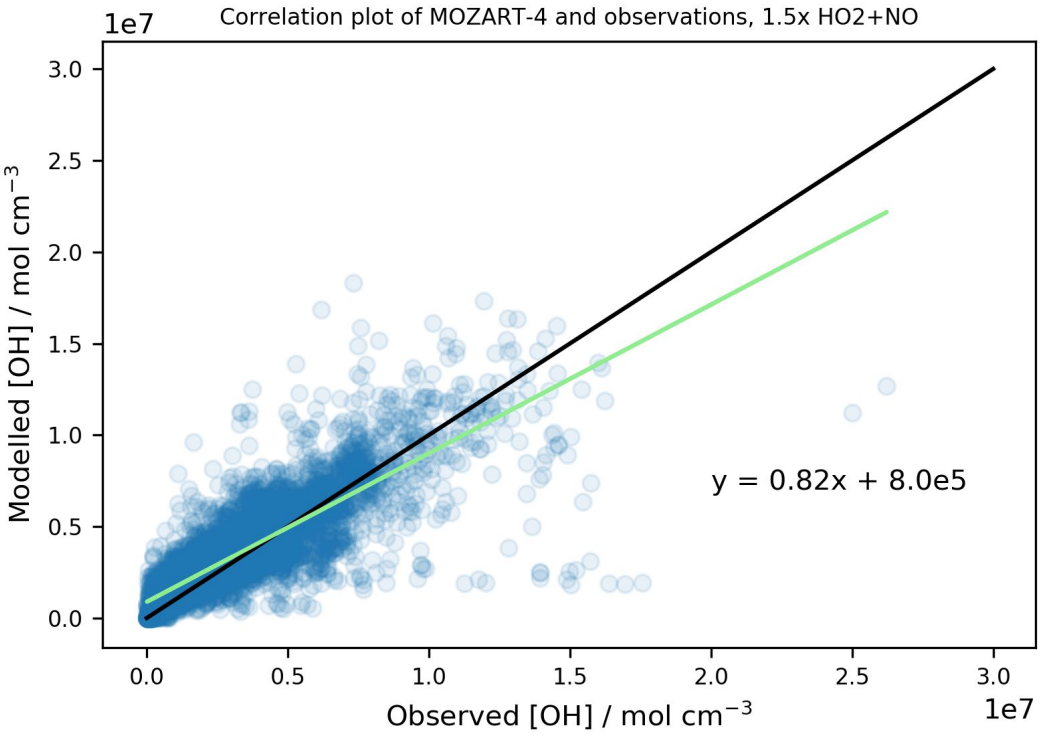
Does low bias stem from an under-represented process?



Does low bias stem from an under-represented process?



Does low bias stem from an under-represented process?



Overall?

- Evaluated the skill of a reduced chemical mechanism in determining OH concentrations
- Explored machine learning approaches as a complementary alternative, finding simple non-linear methods accurate and robust

- Feature importance in our ML models prompted it us to look further into nitric oxide
- Found that it had a considerable effect on our chemical model!