National Science Challenges

THE DEEP SOUTH

Te Kōmata o Te Tonga

Machine learning to create filled total column ozone data sets

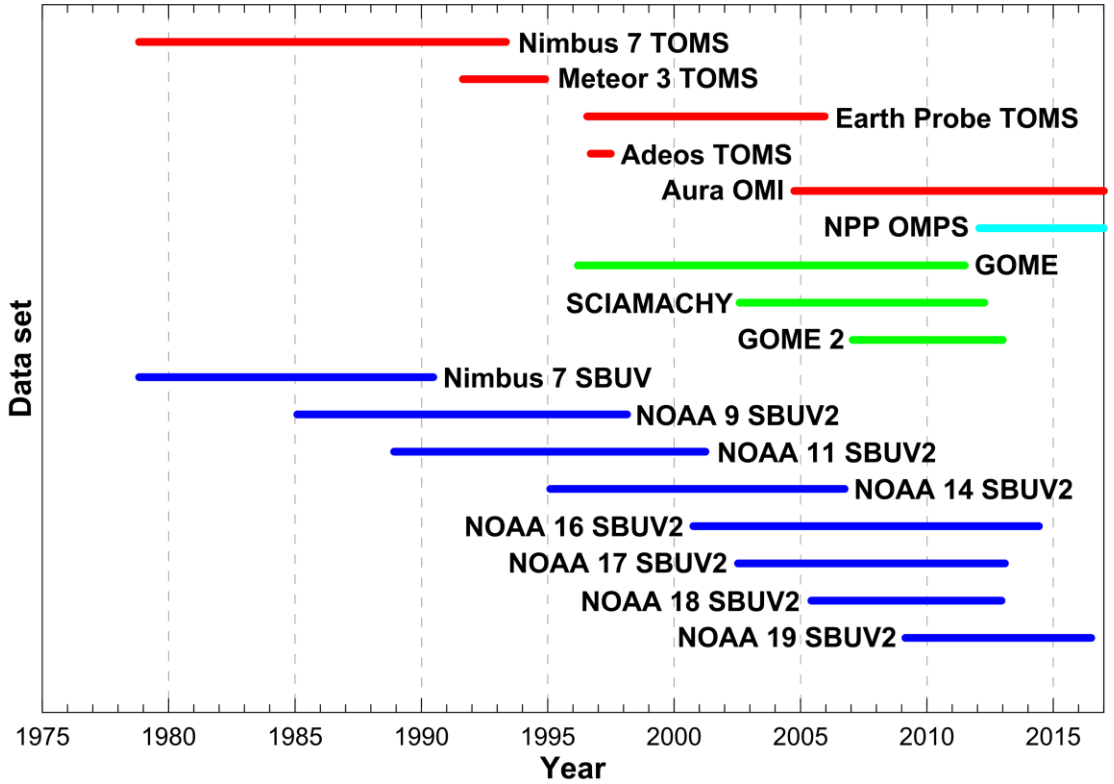
Greg Bodeker

Presented at the Deep South National Science Challenge Symposium

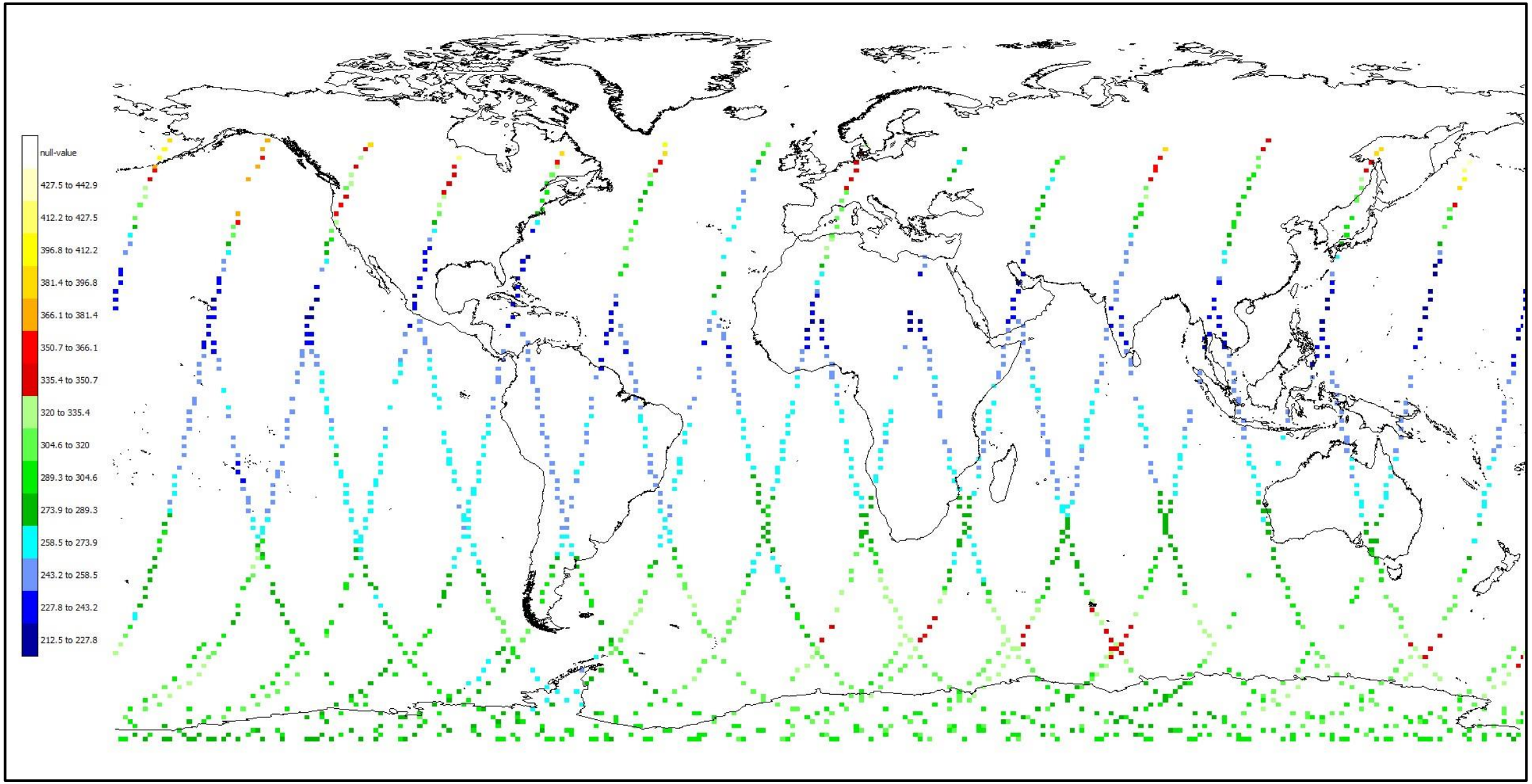
Auckland, 7 May 2019

The NIWA-BS combined total column ozone (TCO) database

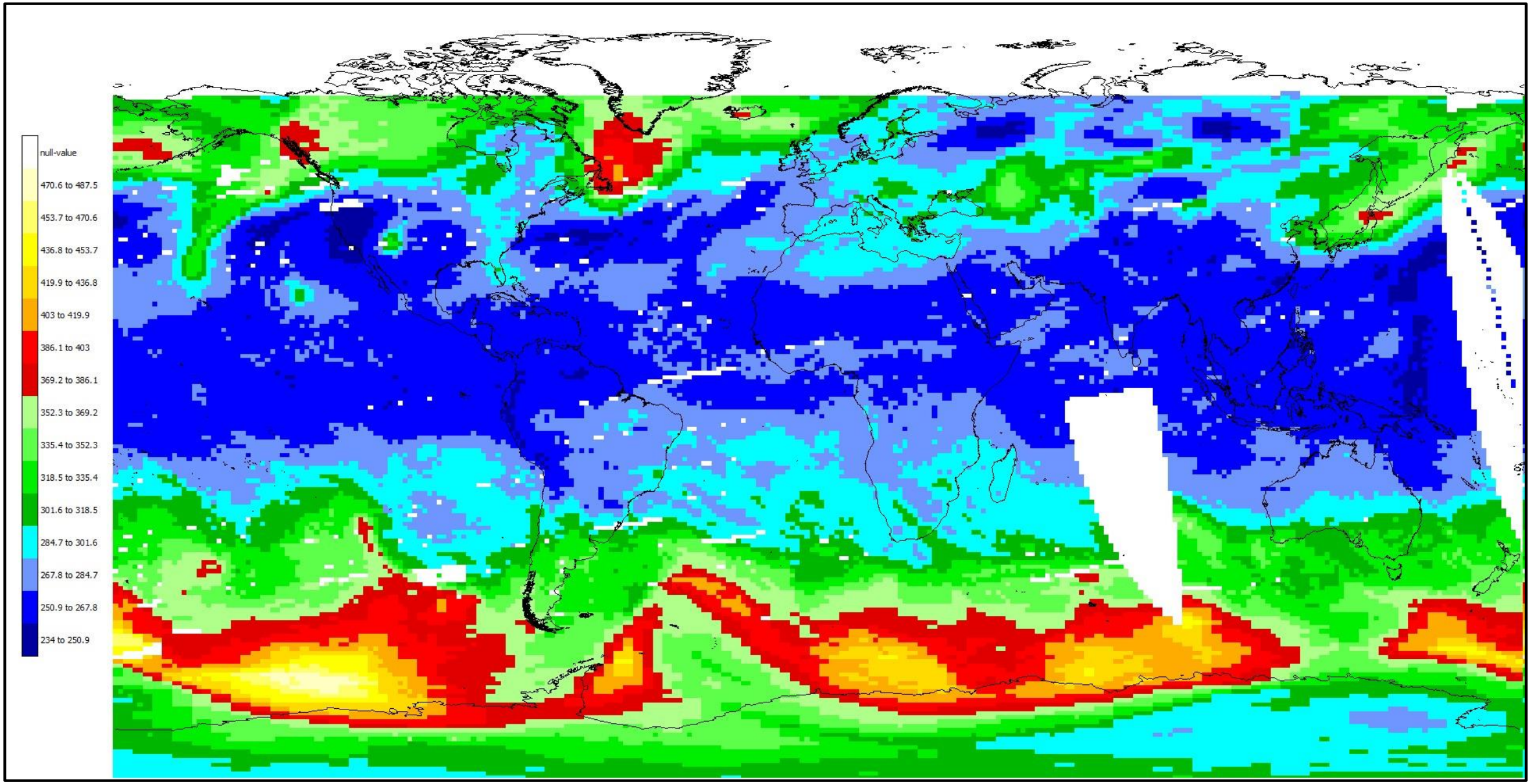
- Long history going back to Bodeker, G.E.; Scott, J.C.; Kreher, K. and McKenzie, R.L., Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978-1998, *J. Geophys. Res.*, 106, 23029-23042, **2001**.
- Used in the Deep South Challenge to benchmark the performance of the NZESM.
- Used by more than 50 other research groups around the world for various purposes.
- Combines measurements from 17 different satellites and removes offsets and drifts through comparisons with the ground-based global Dobson and Brewer spectrophotometer network.



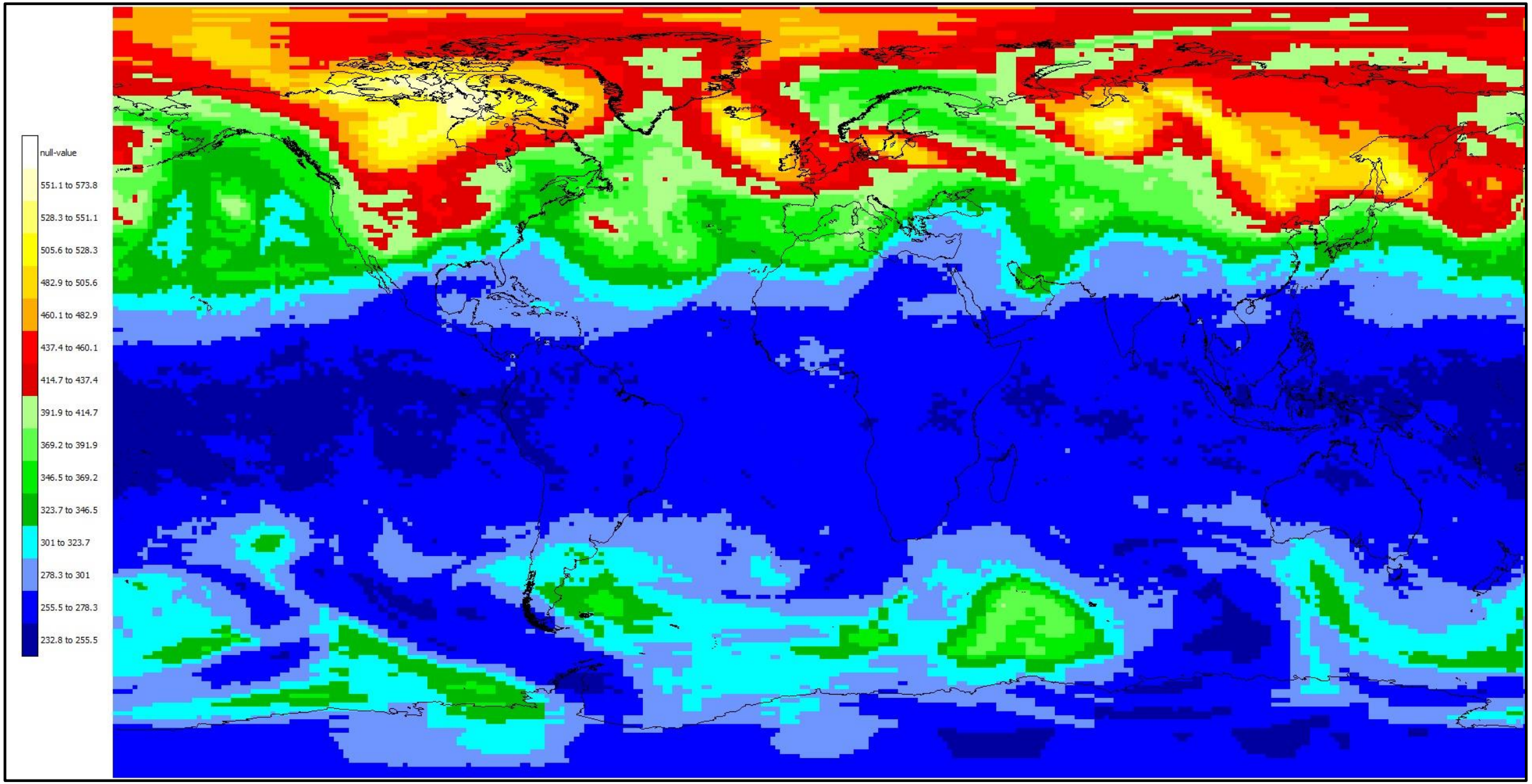
Sometimes the TCO fields look like this:



Or if you are lucky like this:



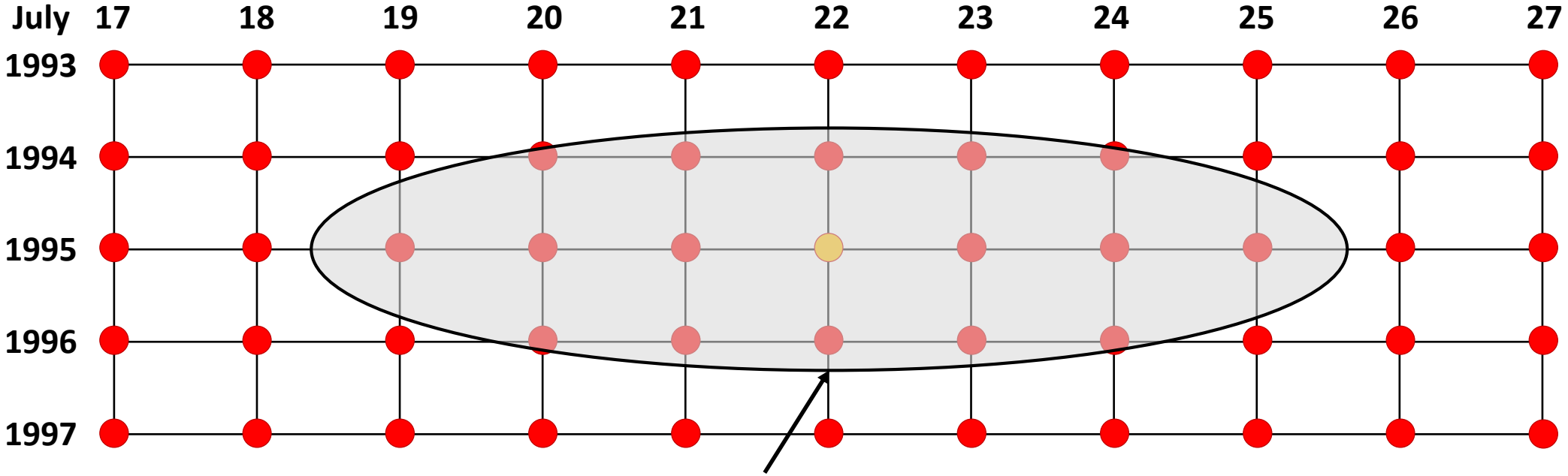
When what you really want is this:



Machine learning to the rescue

Regression analysis → uses **potential vorticity (PV)** and **tropopause height (TH)** fields as predictors of the TCO field. Dependence of TCO on predictors changes with time, latitude and longitude.

22 July 1995 as an example…

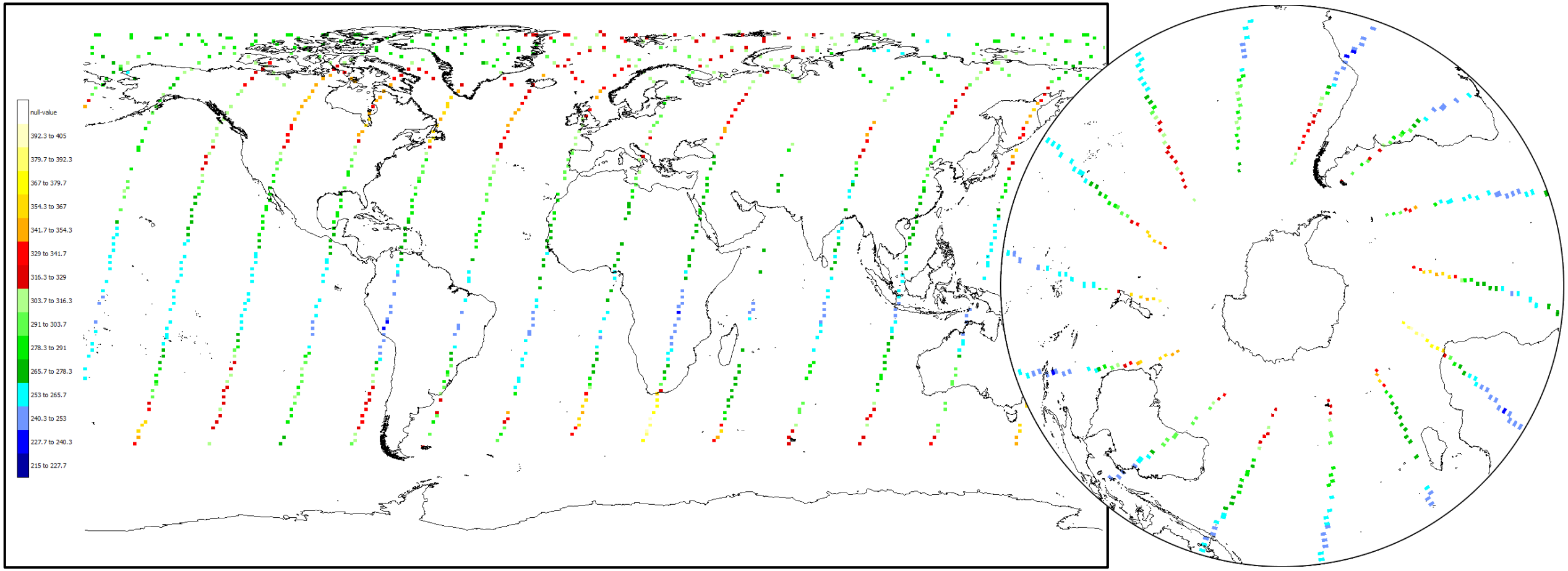


- Training in this domain (~30 minutes/day).
- Geographical structure in dependence of TCO on PV and TH captured by expanding model coefficients in spherical harmonics.
- Expansion indices are data adaptive.

When there are insufficient TCO data we expand the search ellipse.

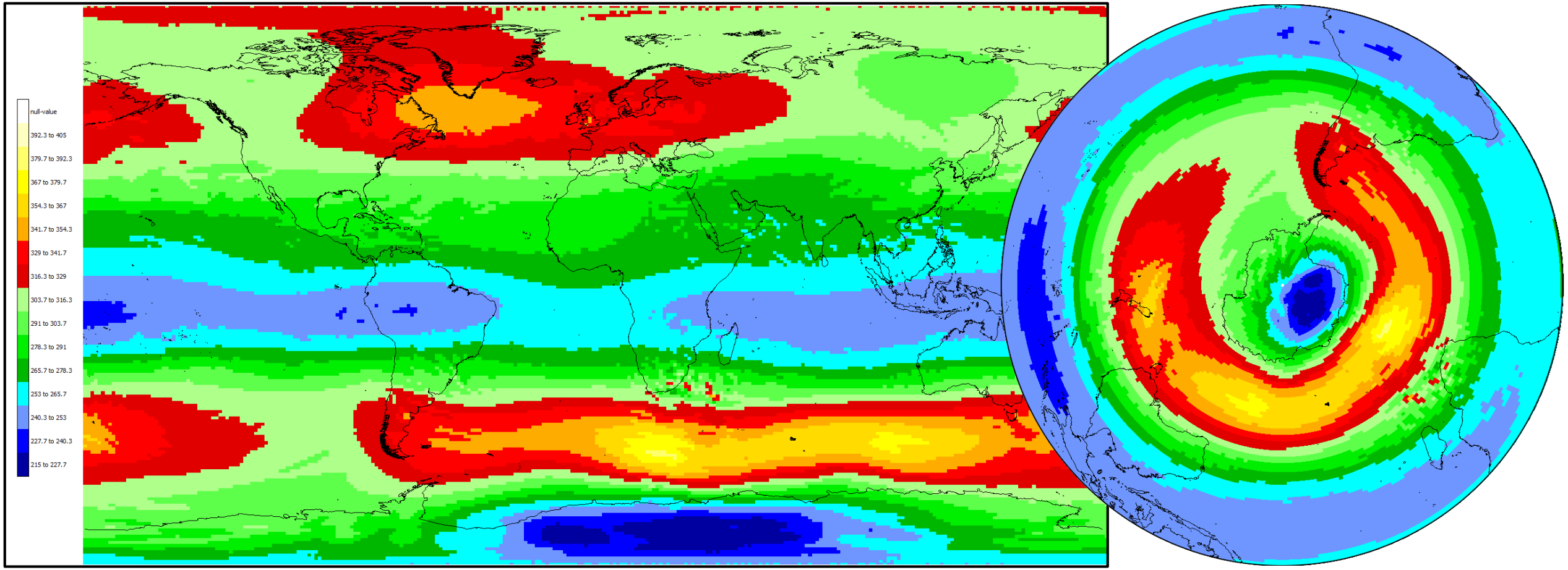
So lets see how this works. We start with original data…

22 July 1995



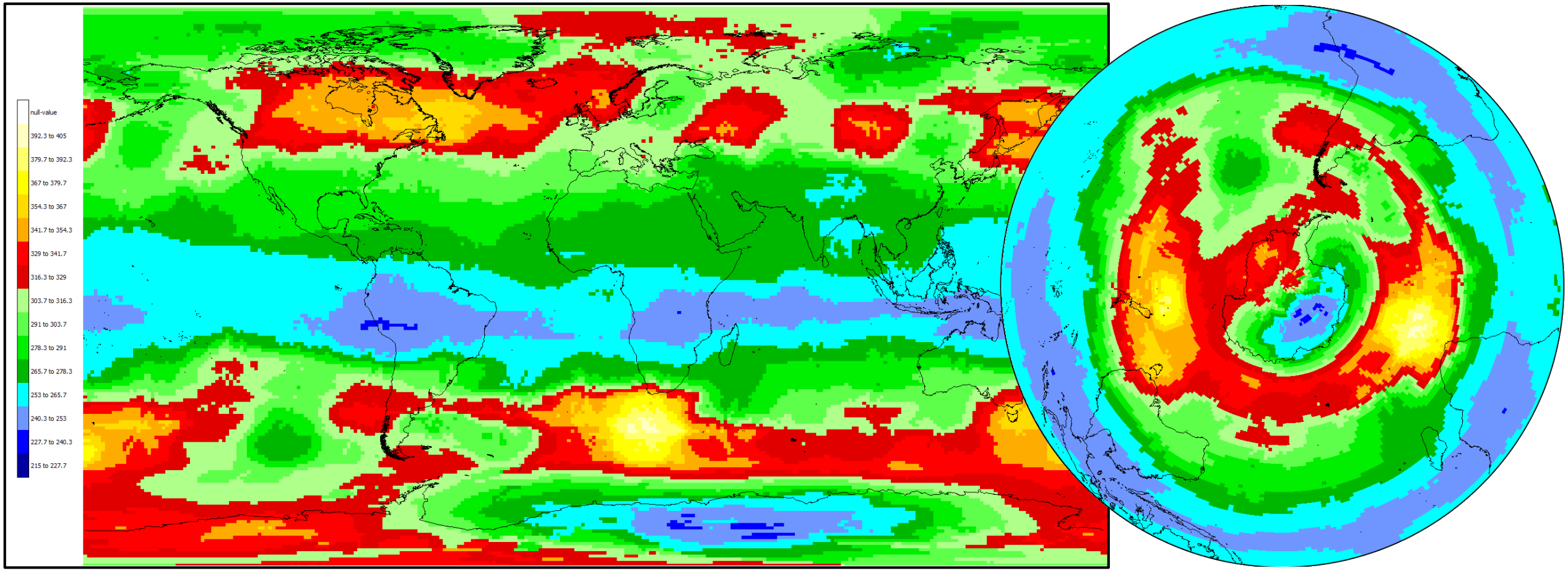
Machine learning output

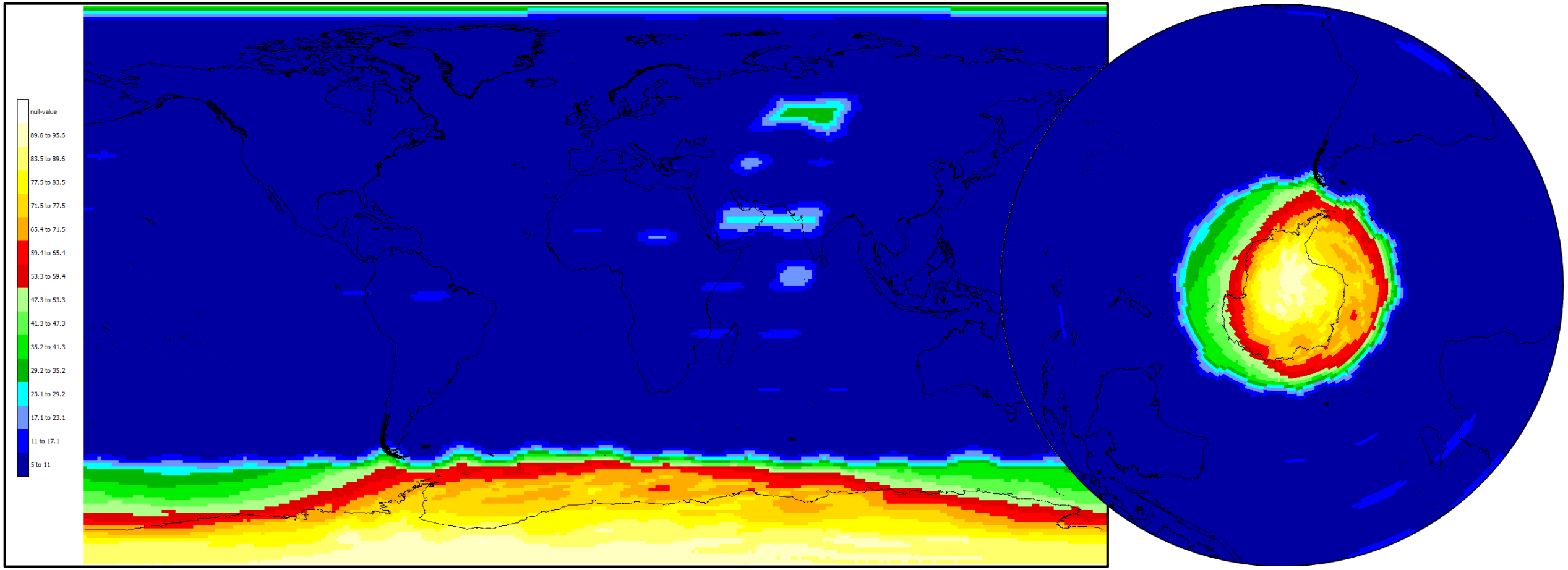
22 July 1995



‘Blend’ the two fields to make the best of both worlds

22 July 1995



What about uncertainties…

22 July 1995