GNSS RO, CMIP6, AIRS, ECWMF and GFS Water vapor histograms Comparisons and Implications

E. R. Kursinski, A. L. Kursinski

Abstract

We have extended our examination of the hydrological cycle and our understanding of it by comparing low latitude, free tropospheric specific humidity histograms derived from COSMIC and COSMIC2 with those derived from ECWMF and GFS forecasts, AIRS version 7 Level 3 products and CMIP6 climate models.

We examined a number of CMIP6 models including HadGEM3, the CanESM5, MPI 1.2, CESM2, GFDL MC4, MIROC 6 and IPSL CMA6. In terms of water vapor histograms, the CMIP6 models are better than CMIP5 models although at certain pressure levels they can be worse. For daily output, no single model is best at the three different pressure levels we examined.

AIRS v7 is significantly better than versions 6 and 5 in the lower free troposphere but not as good as any of the high resolution analyses and reanalyses we examined. Version 7 is actually worse in the upper troposphere.

We quantified the dependence of the performance of the histograms on the temporal resolution of the output. Histograms from 6 hour averaging are only slightly worse than those from 1 hour averaging. However, 24 hour averaging does significantly degrade the histograms due to smoothing out the humidity extremes.

One surprise in our results is the specific humidity histograms from the 6 hour HadGEM3 AMIP run for 2007 with specified SST, which have not assimilated any observations, are as close the GPS RO derived histograms as are the 1 hour ERA5 reanalyses for 2007 which have assimilated an enormous number of observations.

We also note that in deriving the COSMIC-2 water vapor data set and histograms, we ran into a subtle problem where the error deconvolution of the histograms from our initial set of retrievals produced an unphysical spike in the driest bin. This was due to a subtle bias associated with using NWP pressure information. When we derived pressure minimizing the use of NWP information, along the lines used in the JPL RO processing system, the problem went away. This indicates the bias in carefully processed Direct water vapor retrievals is less than 0.03 g/kg, consistent with the estimate of Kursinski and Gebhardt (2014).

Two manuscripts on these results will be submitted shortly as part of a GVAP special issue.

2 Methods of GPS RO Water Vapor Retrievals

1. **1DVar retrievals**: Combine GPS RO refractivity or bending angle profile with forecast T, P, humidity + error covariances
2. **"Direct" Retrievals**: Combine GPS RO refractivity with analysis temperatures

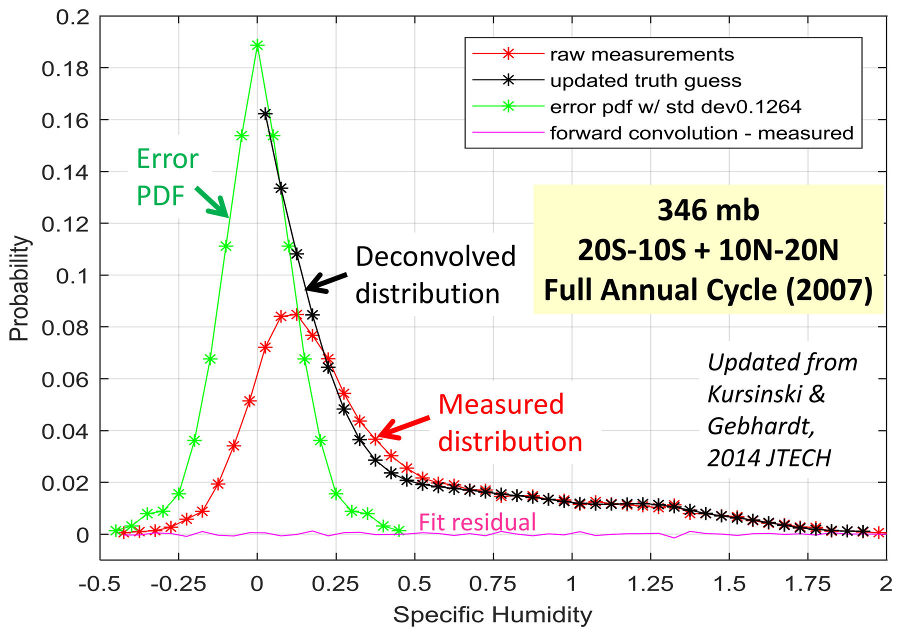
Moisture Histograms

Construct humidity histograms from GPS Direct retrieval profiles

Deconvolve the errors (see *Kursinski & Gebhardt 2014*)

⇒ Yields better humidity histograms w/o using background water vapor guess

plus it yields estimates of the Error PDFs



The Direct water vapor retrievals that we produce combine the RO refractivity profile with NWP temperature and a reference pressure level in the upper troposphere derived mostly from the RO. These Direct retrievals do **NOT** use the NWP water vapor because of unknown biases.

Absolute accuracy of GNSS RO Direct retrievals (COSMIC-2)

The reference pressure must be estimated carefully. Our initial COSMIC-2 Direct water vapor retrievals took the reference pressure from the 1DVar results. However, the resulting the error-deconvolved histograms contained a large, unphysical spike in the driest bin at all three pressure levels indicating a negative bias present in the histograms due to a pressure bias in the upper troposphere.

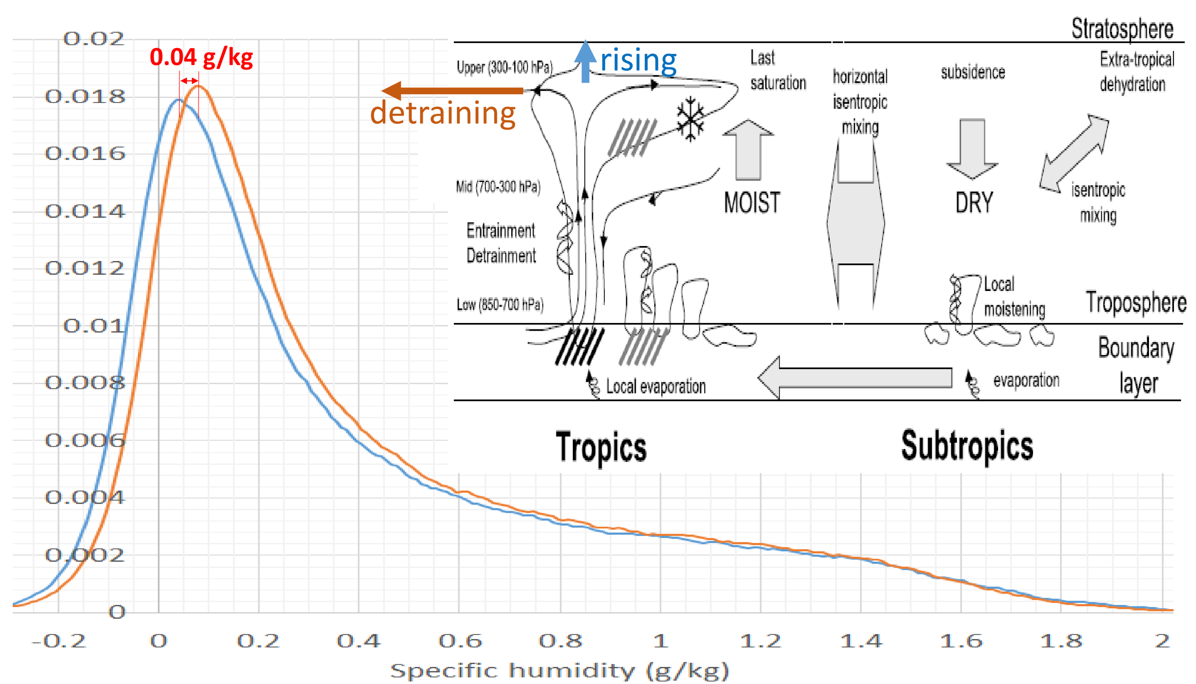
We then changed the reference pressure estimation using an updated version of the JPL approach:

1. Get an estimate of the pressure at 50 km from the NWP (~1 hPa).
2. Convert the RO refractivity to density using (1) NWP water vapor mixing ratio estimates in the stratosphere to account for the very small contribution of water vapor there and accounting for (2) the increasing CO_2 concentration and (3) the non-ideal gas behavior.
3. Perform a hydrostatic integral using that density from 50 km down to 12 km to derive a reference pressure at 12 km.

Direct water vapor results derived using this reference pressure approach showed no obvious bias.

Shown below are two raw histograms from Direct retrievals derived using the two different reference pressures, 1DVar shown in orange and JPL-style shown in blue. The bias using the 1DVar reference pressure is about -0.04 g/kg. The negative bias was revealed by the unphysical spike it caused in its error-deconvolved histogram. Peak mixing ratios using $P_{ref\text{-}1DVar}$ are physically too small corresponding to detrainment temperatures of ~212K (13 km/183 hPa) which are so cold that such air would have then radiatively **warmed** and risen into the stratosphere. Peak mixing ratios using $P_{ref\text{-}"JPL"}$ correspond to detrainment temperatures of ~220K (12km/213hPa) which are high enough to radiatively cool and descend down into the troposphere and are therefore physically reasonable.

- Thus, biases of 0.04 g/kg and subtle UT pressure errors are readily recognizable and correctable.
- **This implies the Direct water vapor retrievals have accuracy sufficient for climate applications.**



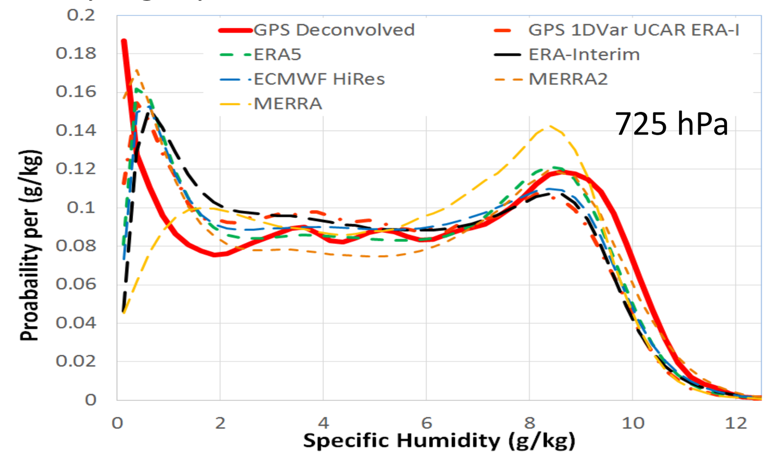
Climate Models

We examined several CMIP6 models including HadGEM3, CanESM5, MPI 1.2, CESM2, GFDL MC4, MIROC 6 and IPSL CMA6.

- In terms of water vapor histograms, CMIP6 are better than CMIP5 models, although at certain pressure levels can be worse.
- For daily output, no one model is best at all three pressure levels we examined. The best are Canada ESM5 at 346, MPI1.2 at 547, GFDL MC4 at 725 and HadGEM3 on average.
  - IPSL has improved steadily from CMIP3 to CMIP6 but is still worst
  - MPI CMIP6 is better in mid troposphere and worse at 725
  - GFDL MC4 is better at 725 and worse in mid troposphere
  - HadGEM3 improved at 725 and mid troposphere is now best overall
- Clearly it is challenging to improve the humidity and get it correct across the depth of the troposphere

Analyses & Reanalyses

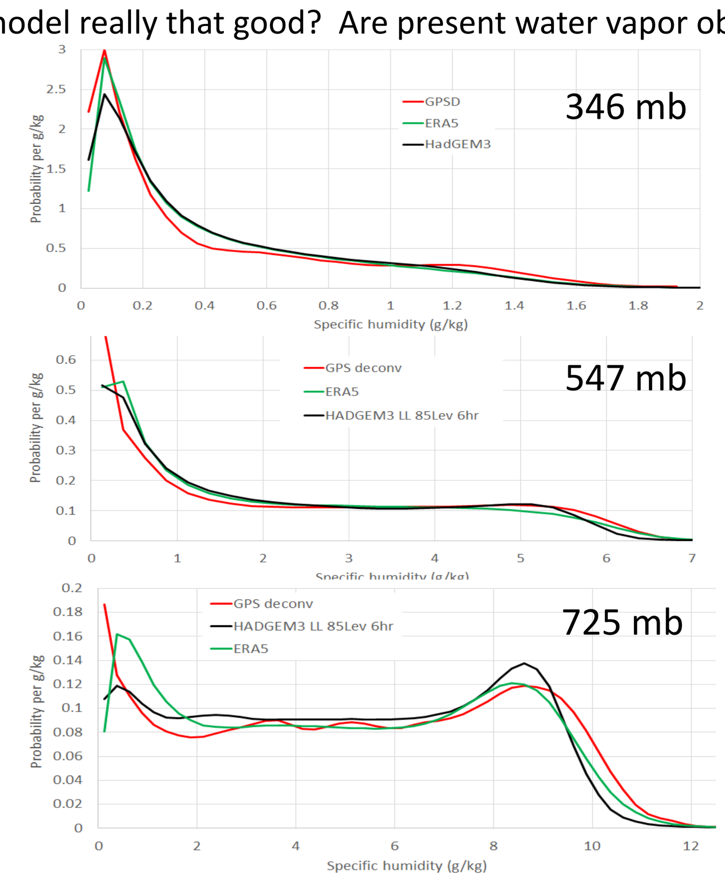
- Analyses do pretty well but tend to underestimate extreme dry and wet air
- ECMWF operational analyses in 2007 better than ERA-I due better resolution?
- ERA5 clearly better than ERAI and 2007 operational analyses and it improved at all 3 levels with the biggest improvement in the mid troposphere
- MERRA (extinct) is best in the upper and mid troposphere but poor at 725
- MERRA2 is much improved at 725 but worse than MERRA at the other two levels
- 1DVar is only slightly better than ERAI



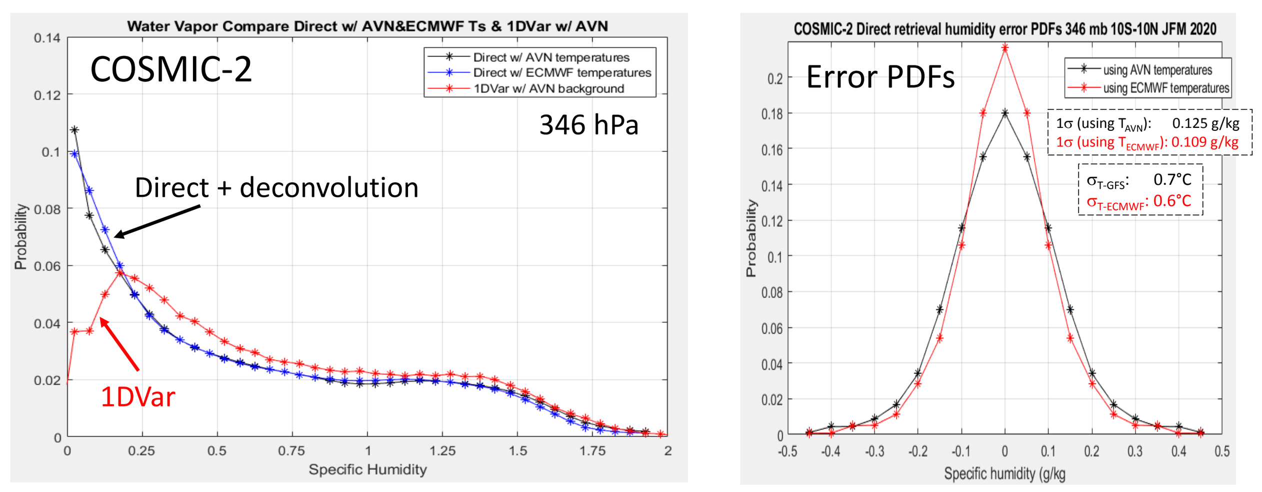
HadGEM3's remarkable performance

HadGEM3 scores as well as ERA5! One surprise is the 2007 specific humidity histograms from the 6 hour HadGEM3 AMIP (specified SSTs) score as close to the GPS RO derived histograms as the 1 hour ERA5 reanalyses for 2007 which have assimilated an enormous number of observations, despite HadGEM2 not having assimilated any data. Questions:

Is the MOHC model really that good? Are present water vapor obs not providing much constraints?



1DVar vs Direct and ECMWF vs GFS temperature uncertainty in UT



Narrower error PDF of Direct retrievals using ECMWF temperatures indicates ECMWF temperatures are less noisy than NOAA GFS temperatures

Evolution of analysis and climate model performance

- Newer versions of systems are generally better than previous versions
- There are signatures of models focused on getting the PWV right: MERRA 2, AIRS v7, GFDL
- MERRA2 has focused on using the PWV to improve its moisture field. That clearly improved it in the lower troposphere but resulted in worse realism in the mid and upper troposphere
- See that it is tricky to get the water vapor correct across the entire depth of the troposphere

To be added

- **Add ROMSAF results**
- **Create CLIMATE quality database**
- **Uncertainty summary**
- **Resolution: Vertical more important than horizontal**
- **how good is good enough?**

Comparisons are in terms of **Rectilinear Distance** (RD) where the deconvolved GPS RO PDFs are the reference PDF

$$RD_{21} = \sum_{i=1}^{N} |PDF_2(i) - PDF_1(i)|$$

CMIP 3 & 5 (CMIP 5: GFDL MC3 – IPSL cm5a; CMIP 3: Miub echo-g – IPSL cm4; DM07: Dessler & Minschwaner saturation-advection model)

P (mb)	GFDL MC3	MPI	HadGEM 2	CESM1.0	MIROC 5	IPSL cm5a	Miub echo-g	MRI cgm2	IPSL cm4	DM07
346	29.0%	27.3%	23.8%	28.7%	42.0%	57.4%	41.9%	40.0%	56.4%	37.8%
547	29.5%	30.7%	30.4%	26.9%	39.7%	56.5%	45.0%	46.7%	57.5%	62.1%
725	20.8%	22.9%	30.3%	33.4%	36.6%	29.1%	29.7%	46.3%	38.6%	
Avg	26.4%	27.0%	28.2%	29.7%	39.5%	47.6%	38.9%	44.4%	50.8%	50.0%

CMIP 6

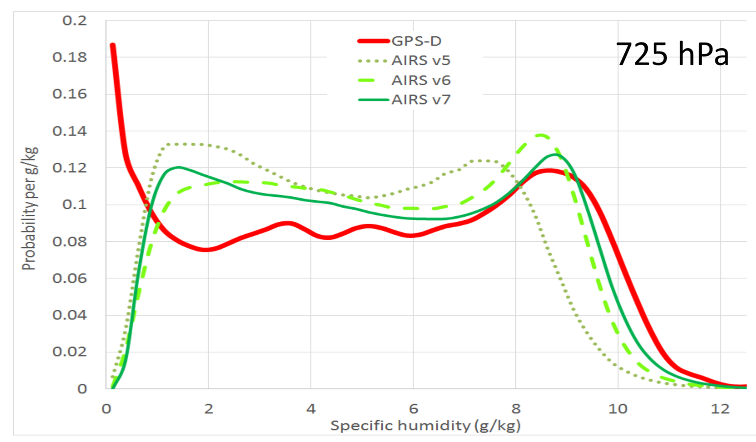
P (mb)	HadGEM3 LL 85Lev 6hr	HadGEM3 LL 85Lev day	Can ESM5	MPI 1.2	CESM2	GFDL MC4	HadGEM3 MM 19Lev	MIROC 6	IPSL CMA6
346	**19.0%**	27.3%	**19.9%**	28.9%	27.9%	29.4%	34.1%	41.2%	47.2%
547	**17.6%**	23.8%	26.3%	**21.5%**	22.7%	33.3%	36.3%	41.4%	45.0%
725	**14.8%**	20.0%	29.0%	28.5%	29.0%	**16.8%**	23.5%	24.7%	27.5%
Avg	**17.1%**	**23.7%**	25.1%	26.3%	26.5%	26.5%	31.3%	35.8%	39.9%

Analyses/ Reanalyses (Analyses & reanalyses: MERRA – NCEP FNL; AIRS Level 3: AIRS v7 – v5)

Level (mb)	MERRA (deceased)	ERA5 hrly37 lev	MERRA 2	ECMWF 2007	GPS 1Dvar	ERA-I	ECMWF Lo-Res	NCEP FNL	AIRS v7	AIRS v6	AIRS v5
346	**8.9%**	19.0%	19.5%	19.5%	29.8%	23.7%	48.7%	42.7%	30.4%	26.3%	**25.8%**
547	**12.6%**	18.2%	20.3%	25.0%	24.5%	30.6%	55.1%	56.9%	44.6%	45.6%	**34.7%**
725	20.7%	14.0%	**12.1%**	15.5%	17.1%	19.1%	30.3%	46.9%	**24.6%**	31.4%	45.0%
Avg	**14.0%**	17.1%	17.3%	20.0%	23.7%	24.5%	44.7%	48.9%	**33.2%**	34.5%	35.2%

AIRS Level 3 versions 5,6,7

In the lower free troposphere, AIRS v7 is significantly better than v6 and v5. However, v7 is not as good as any of the high resolution analyses and reanalyses we examined. All AIRS versions miss the dry air at 725 hPa. Also v7 is worse than v5 and v6 in the upper troposphere.



Temporal Resolution's Impact on Moisture Extremes

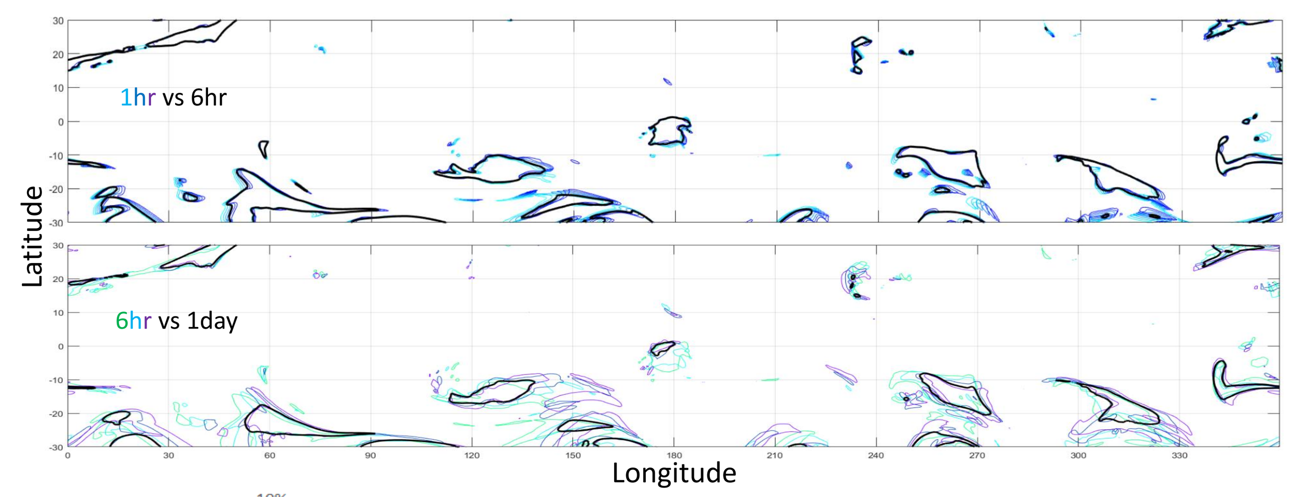
HadGEM3 6 hr 85 Level is better than any other AMIP at all three pressure levels examined, a surprising result, not seen previously. A key clue is that HadGEM3 6 hr is the only CMIP6 output with 6 hour resolution. All the rest are daily output. Averaging HadGEM3 6 hr down to 24 hour resolution, resulted in degrading its RD scores. It still scored best overall but is not best at any of the three pressure levels like HadGEM3 6hr is.

We quantified the dependence of the performance of the histograms on the temporal resolution of the output by averaging the ERA5 from 1 hr to 6 hr to 24 hr and averaging the HadGEM3 from 6 hr to 24 hr.

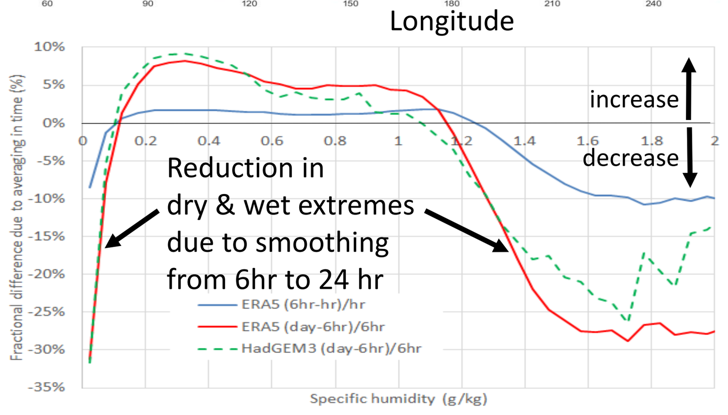
- 1 hr to 6 hr averaging: Degraded ERA5 histograms slightly, increasing RD by ~**2%**.
- 6 hr to 24 hr averaging: Degraded ERA5 & HadGEM3 histograms, increasing RD by ~**7%** due to smearing out much of the air with extreme humidity extremes.

The poorer RD scores associated with temporal averaging is primarily due to **reduced extreme moisture**

Contours of driest air at 346 hPa



% changes in PDFs at 346 hPa due to temporal averaging



Temporal averaging compresses the PDFs.

Note how similar the fractional changes are in going from 6hr to 24hr resolution in both the ERA5 and HadGEM3 results