

How the Emissivity Principal Components (EPC) Profiling Algorithm Adapts to All Surface Conditions (Part 1)

Joe Turk (1) and Nobuyuki Utsumi (2)

(1) JPL/Caltech, Pasadena, CA

(2) Kyoto University of Advanced Science, Kyoto, Japan

19 August 2020

The work contained in this presentation was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. © 2020 all rights reserved.

Background

Original idea from Ziad Haddad, then Yalei You did early work on this, while he was a summer intern at JPL.

Previous work have demonstrated a Bayesian-based precipitation retrieval framework based on an *a-priori* dataset formed by the principal components of the emissivity vector and the environmental state (surface temp, 2-m air temp, column water vapor). “emissivity principal components” or EPC.

Using a large set of DPR and DPRGMI (CORRA) data for each radiometer type (GMI, ATMS, etc), transformation coefficients are derived that allow the EPC vector to be calculated from the observed TB.

The EPC is also used to index and guide the *a-priori* database searches, to isolate candidates that are most congruent to the observations.

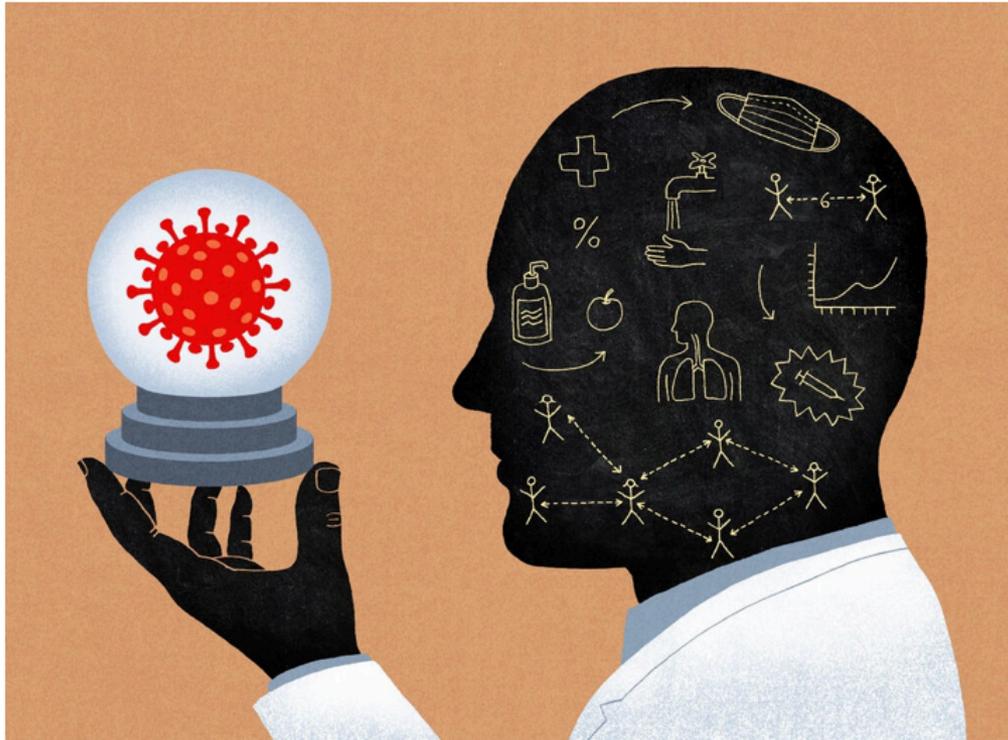
¹ Turk F. Joseph, Haddad Z. S., Kirstetter P., You Y., & Ringerud S. (2018). An observationally based method for stratifying a priori passive microwave observations in a Bayesian-based precipitation retrieval framework. *Quarterly Journal of the Royal Meteorological Society*. <https://doi.org/10.1002/qj.3203>

Utsumi, N., F. J. Turk, Z.S. Haddad, P. Kirstetter, H. Kim, 2020: Evaluation of Precipitation vertical profiles estimated by GPM-Era satellite-based passive microwave retrievals. *J. Hydrometeorology*, in review.

Utsumi, N., F. J. Turk and Ziad. S. Haddad, 2020: Consistency of the passive microwave-based precipitation vertical profiles and surface precipitation rates estimated from the GPM constellation satellites (In preparation) .

How to Think Like an Epidemiologist

Don't worry, a little Bayesian analysis won't hurt you.



James Steinberg

New York Times
Aug 4 2020

<https://www.nytimes.com/2020/08/04/science/coronavirus-bayes-statistics-math.html>



There is a statistician's rejoinder — sometimes offered as wry criticism, sometimes as honest advice — that could hardly be a better motto for our times: “Update your priors!”

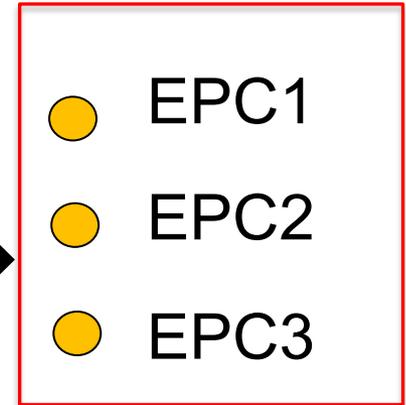
In stats lingo, “priors” are your prior knowledge and beliefs, inevitably fuzzy and uncertain, before seeing evidence. Evidence prompts an updating; and then more evidence prompts further updating, so forth and so on. This iterative process hones greater certainty and generates a coherent accumulation of knowledge.

Principle

Ideally, if one could obtain a reasonable estimate of the emissivity vector at the observation time, this formulation would better accommodate day-to-day (or shorter) changes in the surface (land, ocean, mixed...) emissivity properties

emissivity principal component analysis (EPC)

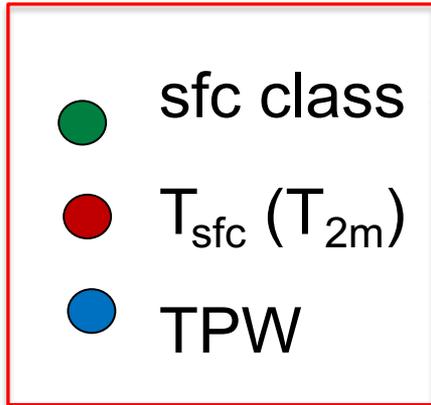
first three EPC explain the majority of the joint variability



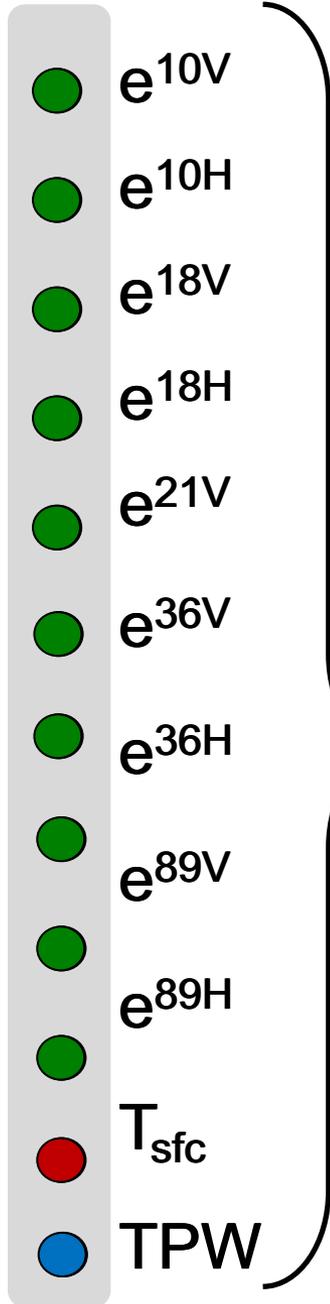
EPC search and index terms

(other radiometers have different channels)

GPROF search and index terms



highly correlated

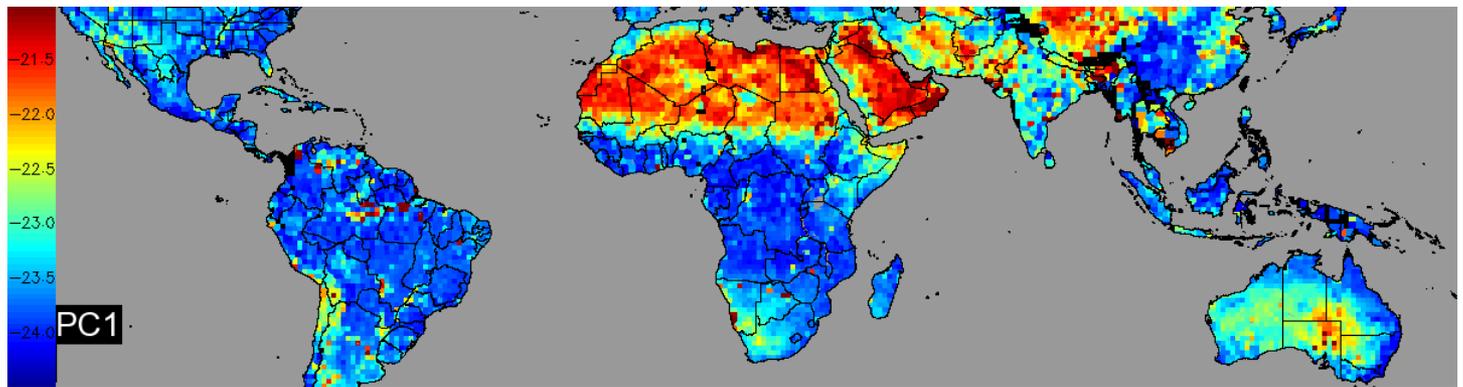


TELSEM Surface Classes Used in GPROF Database index and search

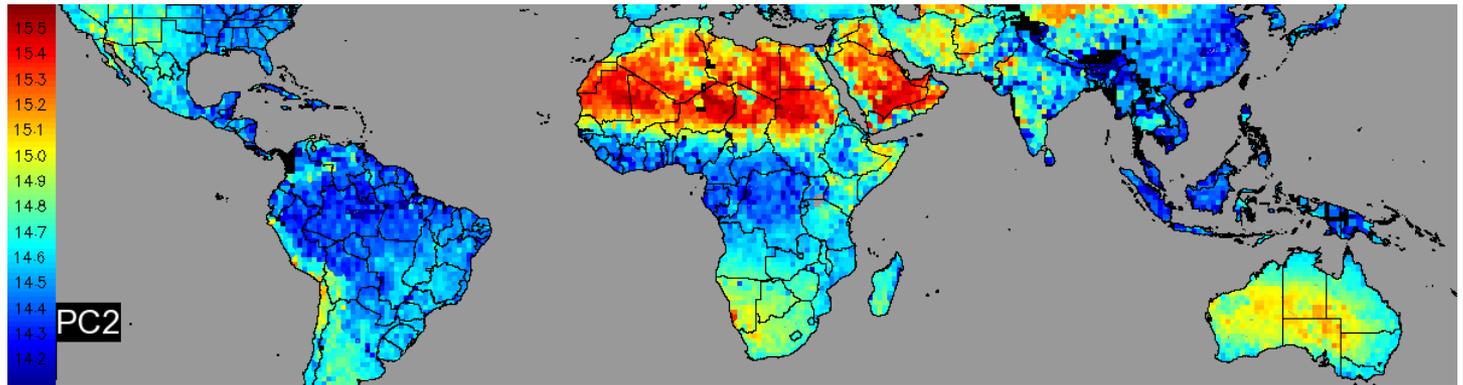
- 1 Ocean / Large inland Water
- 2 Sea Ice
- 3 –7 Decreasing Vegetation Covered
(3=Amazon, 7= Sahara Desert)
- 8 –11 Decreasing Snow Covered
(8=Antarctica, 11= lightly snow covered)
- 12 Inland Water / Rivers/ Estuaries
- 13 Coastlines (land/ocean boundary)
- 14 Ocean / Sea-ice Boundary

Daily land snow gets assigned 10.

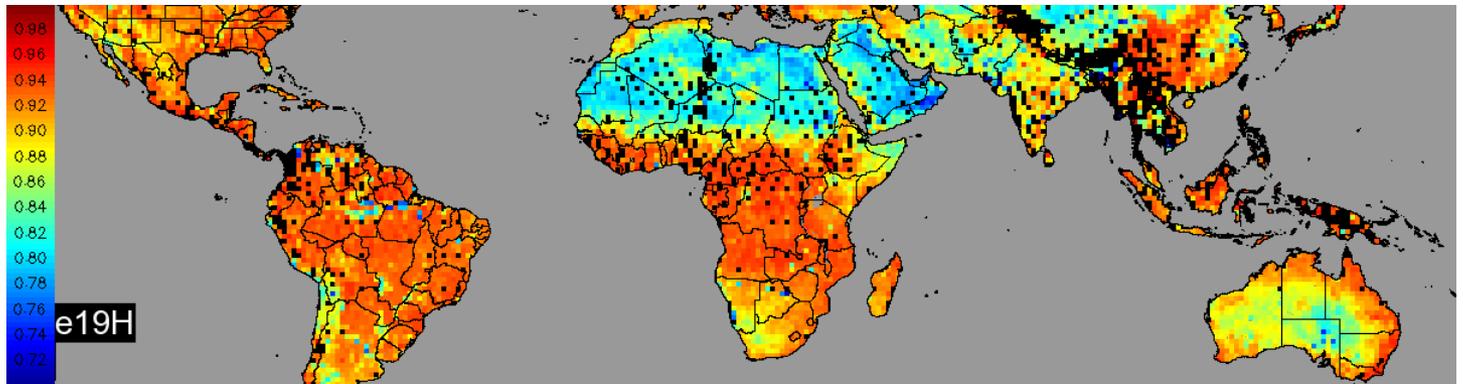
First EPC



Second EPC

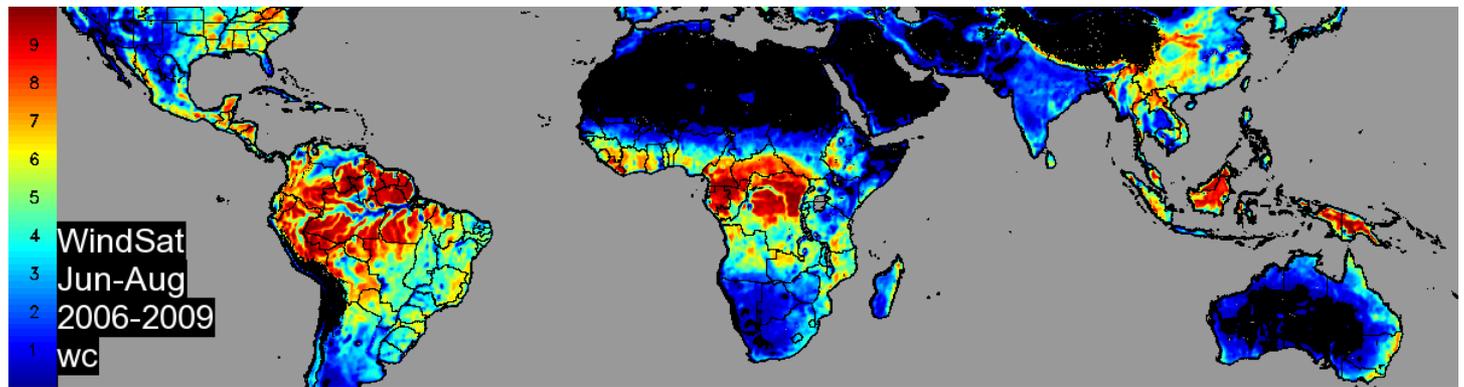


TRMM-TMI
emissivity
19H

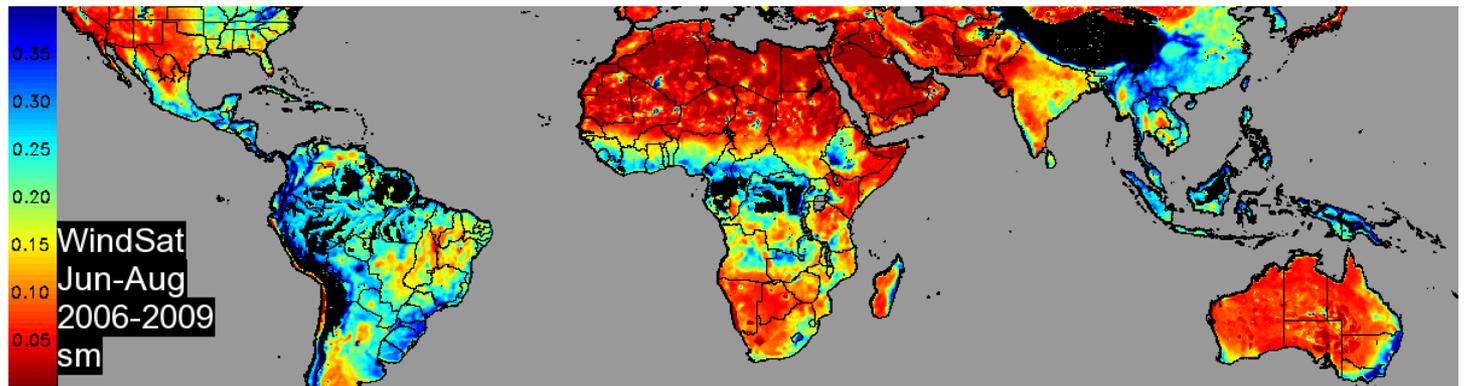


Turk, F. J., Haddad, Z. S. & You, Y. Principal components of multifrequency microwave land surface emissivities. Part I: Estimation under clear and precipitating conditions. *Journal of Hydrometeorology* **15**, 3–19 (2014).

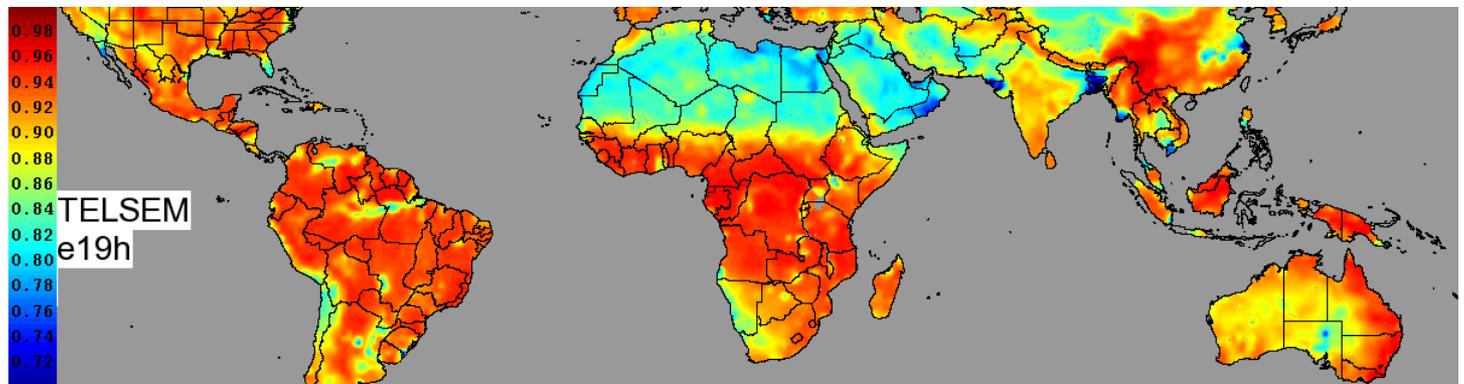
WindSat-
derived
vegetation
water
content



WindSat-
derived soil
moisture



August
TELSEM
emissivity
19H

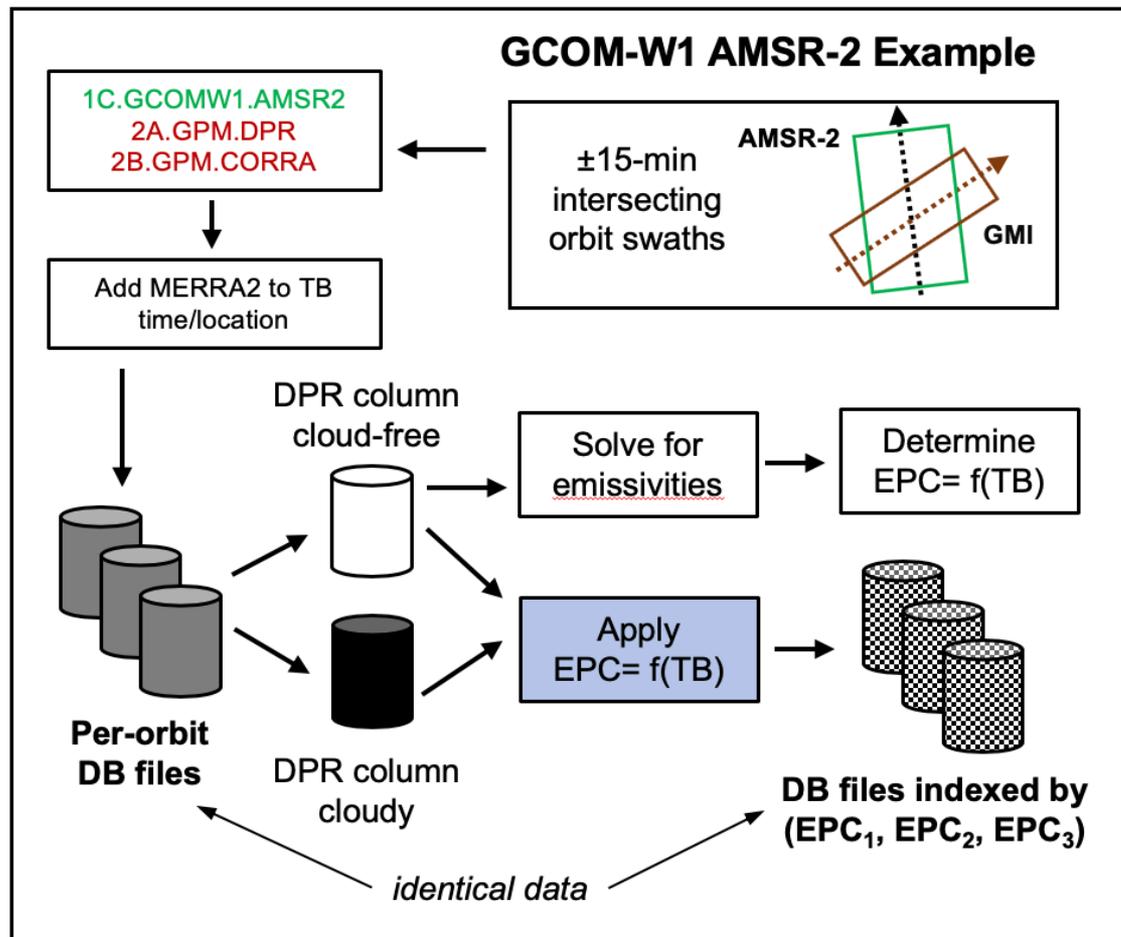


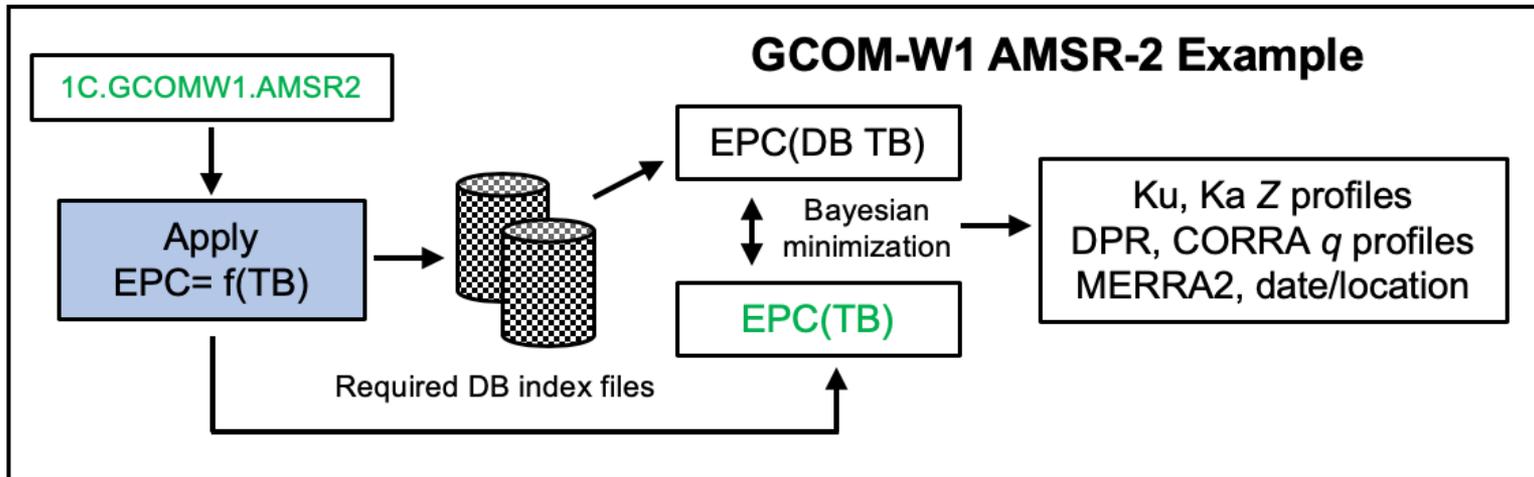
Turk, F. J., Haddad, Z. S. & You, Y. Principal components of multifrequency microwave land surface emissivities. Part I: Estimation under clear and precipitating conditions. *Journal of Hydrometeorology* **15**, 3–19 (2014).

This methodology was expanded to **all surfaces** (wherever GMI observed) after about 4 yrs of GPM data was collected.

A principal component analysis uses the large collection of DPR cloud-free observations to compute the coefficients that transfer the TB observations to EPC space (blue boxes).

The *a-priori* database is indexed by the first three EPC elements.





The solver in the EPC retrieval is similar (nearly identical) to the facility GPROF algorithm, except that the observations and database entries are weighted in EPC space.

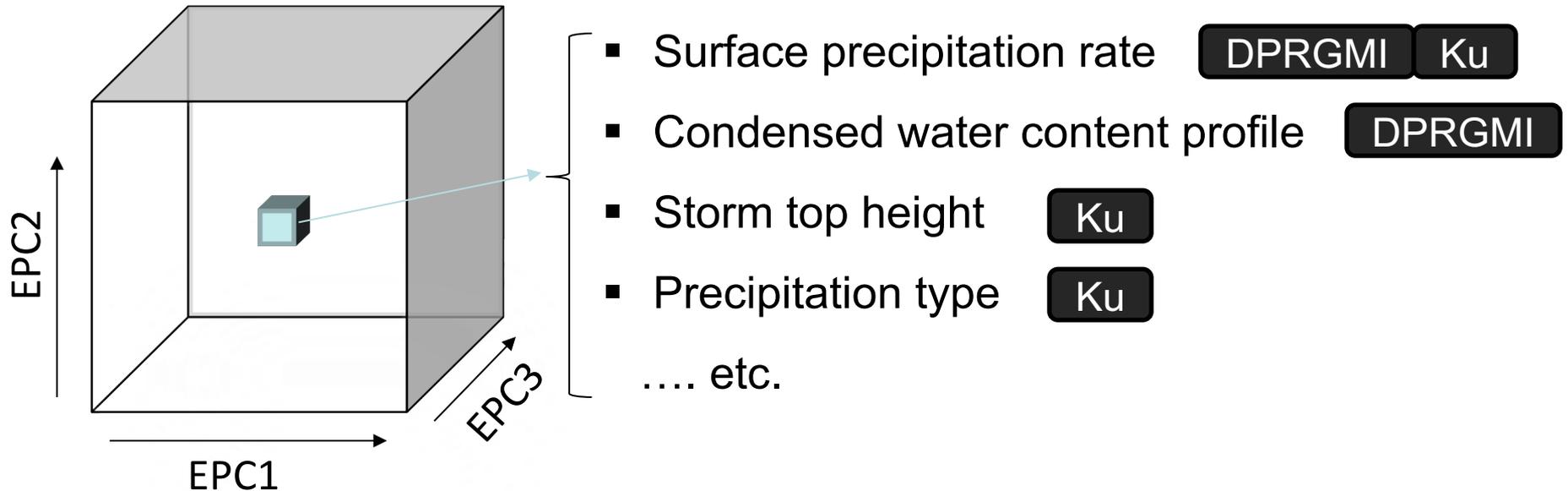
The surface precipitation(s) and the DPR and CORRA vertical Ku/Ka-band structure is estimated at the same time.

The EPC writes out netCDF4 files which are formatted nearly identical to GPROF files (runs on my Macbook with just a C++ compiler and python3).

EPC writes out **many types** of diagnostics some of which I will show next.

a-priori Database (DB) binned by EPC

A-Priori DB, indexed by EPC, was developed from DPR & each of the constellation of radiometer matching scenes.



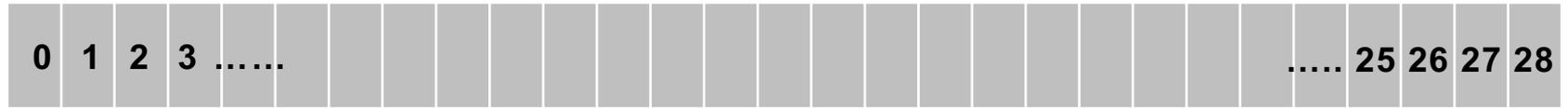
- EPC is calculated from TB at each observation time.

$$EPC = f(TB) \leftarrow \text{Predefined regression function relating TB combinations to each EPC}$$

- The 3-D "cube" is built and indexed in a 1-D fashion.

Binning the a-Priori Database (DB)

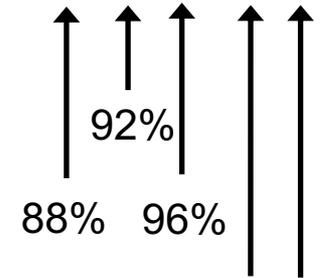
←..... 29 bins, covering the expected range of EPC1→



In between, bins 2-26 hold the DB entries
in 4% increments of the CDF of EPC1

*Same procedure for
EPC2 and EPC3*

Bins 0 and 1 hold the DB
entries when the CDF of
EPC1 reaches the 0.001%
and 0.1% level



Bin 27 and 28 hold the DB
entries when the CDF of
EPC1 reaches the 99.9%
and 99.999% level

Example: For a given TB, its EPC1, EPC2 and EPC3 fall into bins 10, 8, and 25.

$$\begin{aligned} \text{DB index} &= (29)^2 \text{EPC}_1 + (29)^1 \text{EPC}_2 + (29)^0 \text{EPC}_3 \\ &= (29)^2 10 + (29)^1 8 + (29)^0 25 = \mathbf{8667} \end{aligned}$$

The DB index ranges from 0 to $(29)^3 - 1$
(0 to 24389)

**This is important
for two reasons
(next slide)**

Binning the a-Priori Database

The database is sparsely populated in some of the 3-D areas.

Example: An observed TB falls in DB index 10000, but there are only 20 entries (insufficient). The database search can be expanded outward (9999, 10001, 9998, 10002, etc.) until a sufficient number of DB entries are reached.

Since this changes the smaller EPC3 bin (EPC2 if needed), *the database expansion search moves slowly and smoothly through the joint variability in surface emissivity and environmental conditions.*

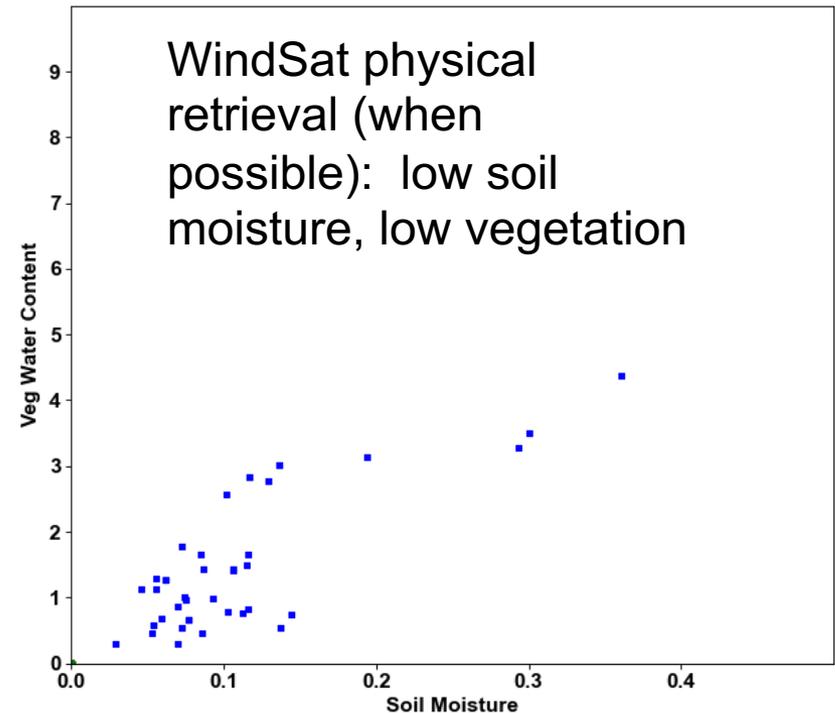
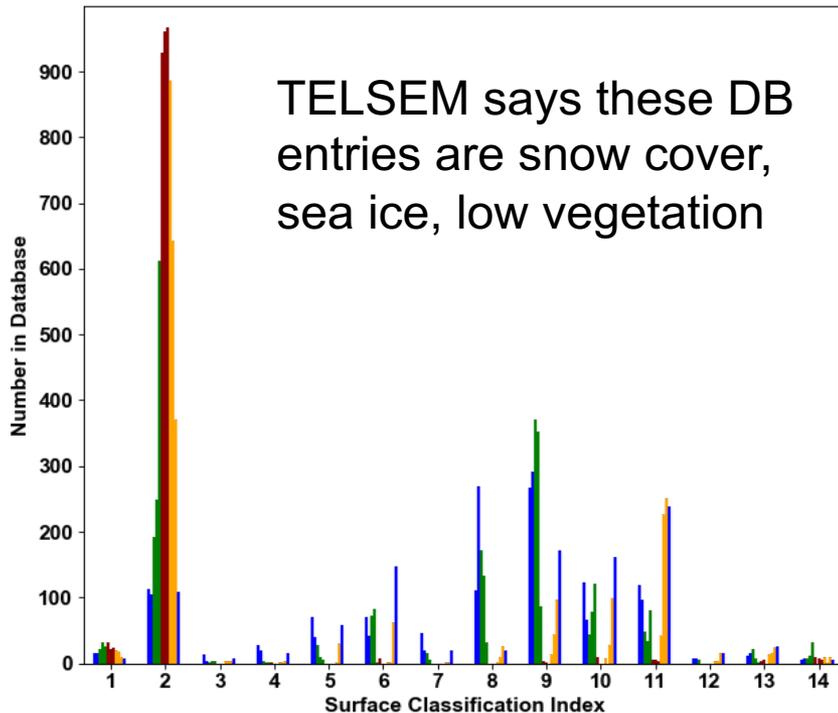
The extreme cold 89 GHz TB (corresponding to the extreme precipitation observations) tend to cluster in the first (0, 1) or last (27, 28) EPC bins.

Using logarithmic bin spacing at the end bins isolates the extreme precipitation events. When the EPC computed from an observed TB falls in one of these bins, it gets associated with more extreme precipitation (ie, *fewer non-extreme events are included in the Bayesian weighting*).

Example

Database file 03208 (EPC1 is small)

Only DB entries whose month is February-
Haven't finished this yet!



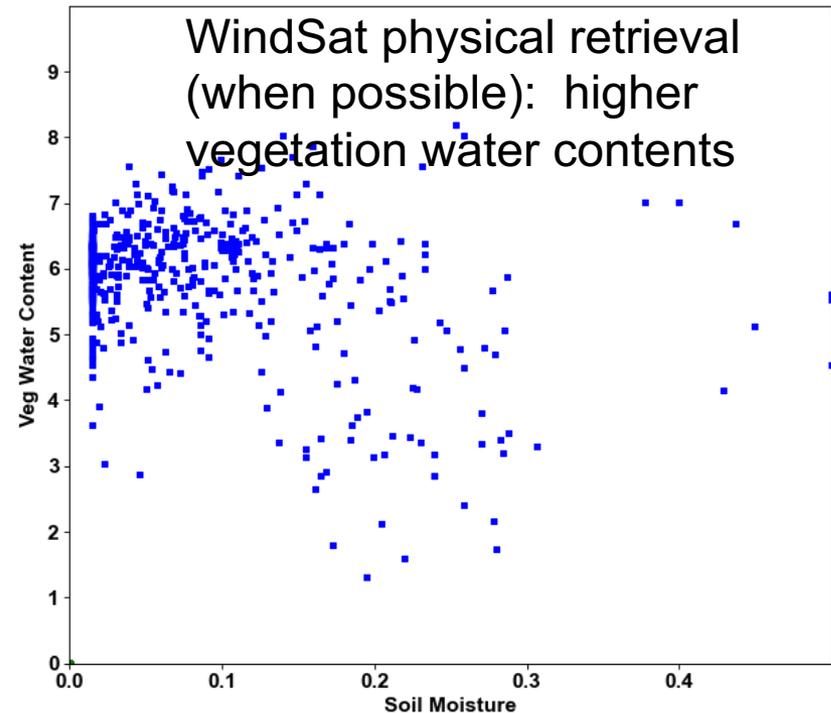
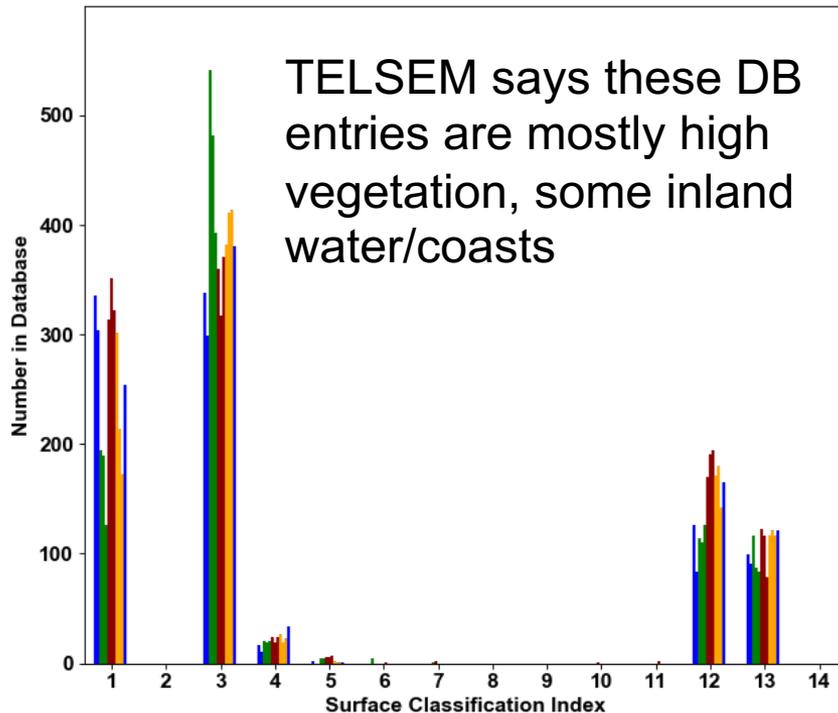
Mixture of what TELSEM says these pixels were (TELSEM is a classification-may or may not be what the surface looks like at observation time)

Database has mixture of raining and non-raining; the Bayesian search is blind to this

Example

Database file 19208 (EPC1 is larger than before)

Only DB entries whose month is February-
Haven't finished this yet!



Mixture of what TELSEM says these pixels were (TELSEM is a classification-may or may not be what the surface looks like at observation time)

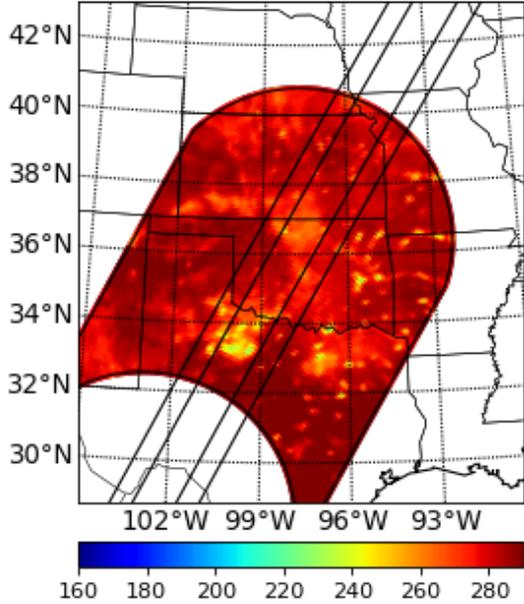
Database has mixture of raining and non-raining; the Bayesian search is blind to this

GPM Overpass

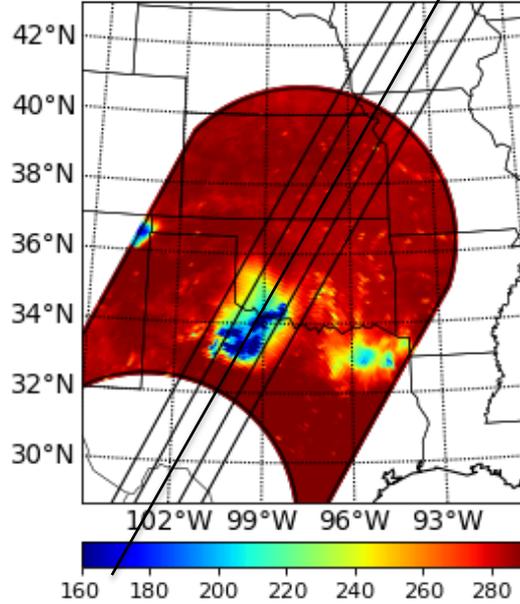
8 June 2018 0043 UTC

Nadir track

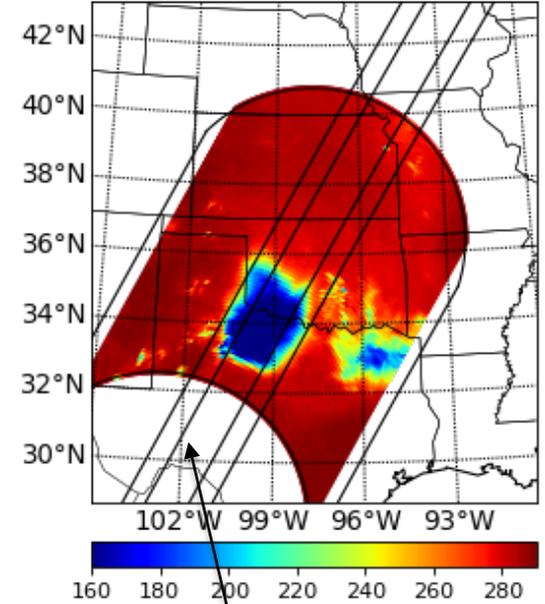
GMI 18.7H (K)



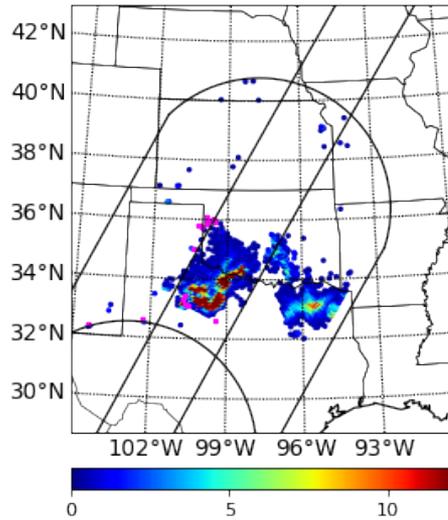
GMI 89.0H (K)



GMI 166H (K)

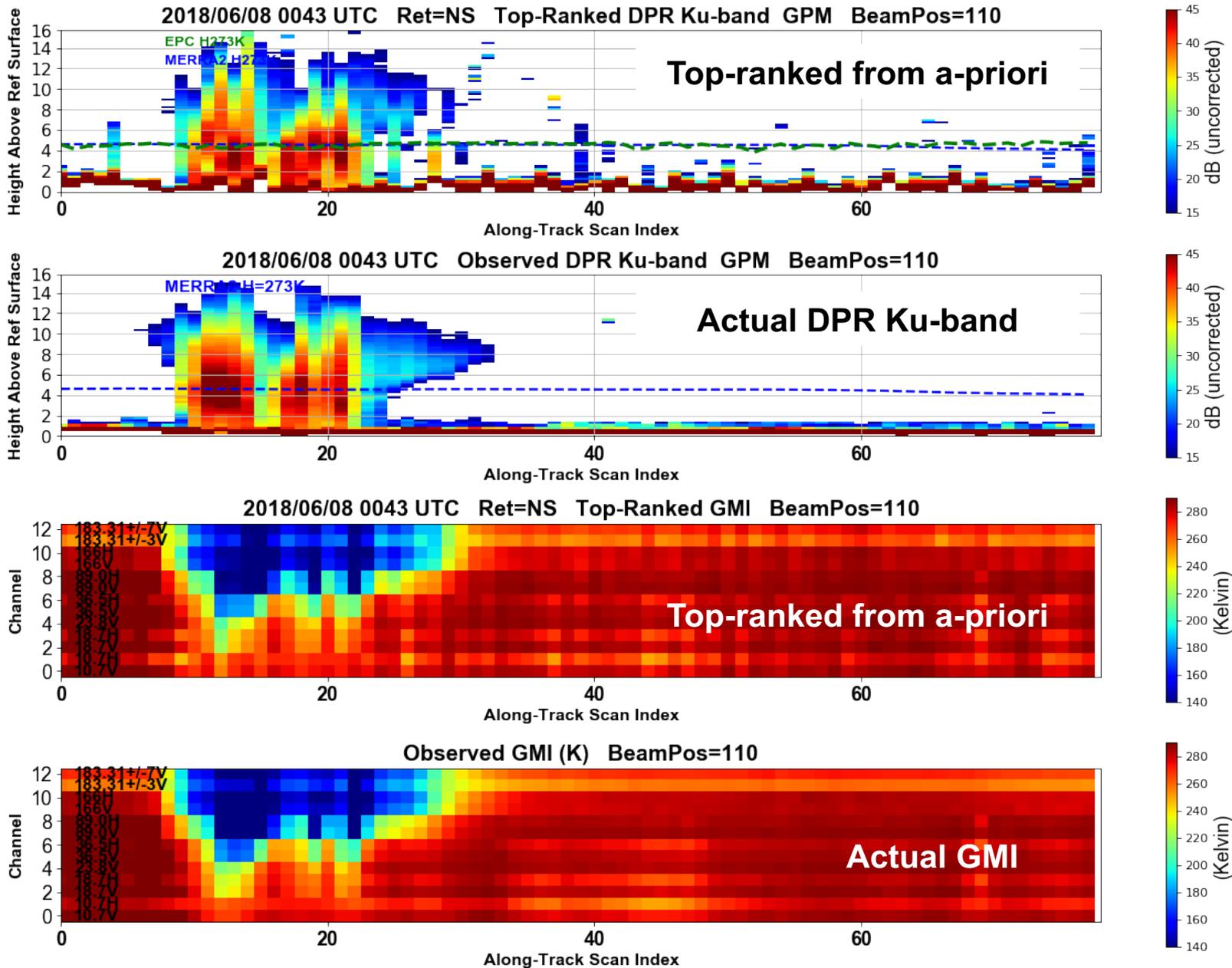


EPC Precip (CMB-NS)



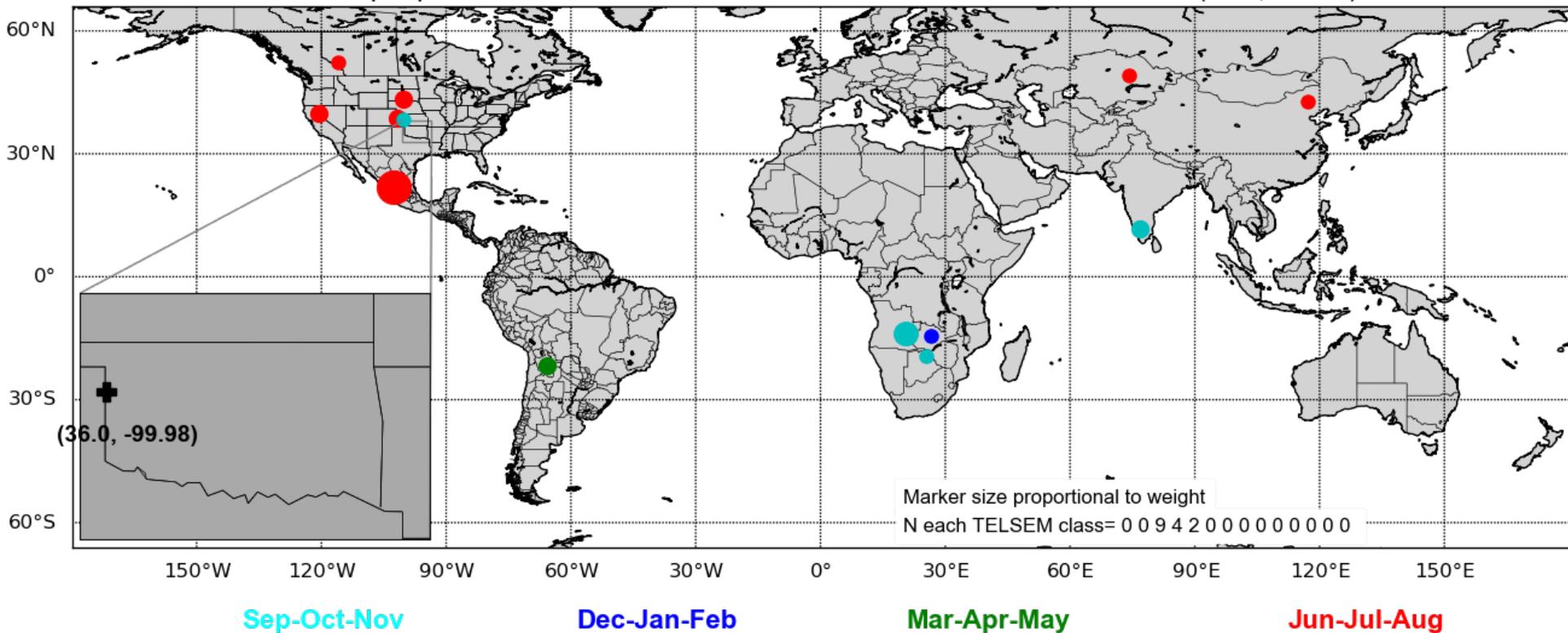
Swaths of Ku- and Ka-band radar

GPM Overpass 8 June 2018 0043 UTC



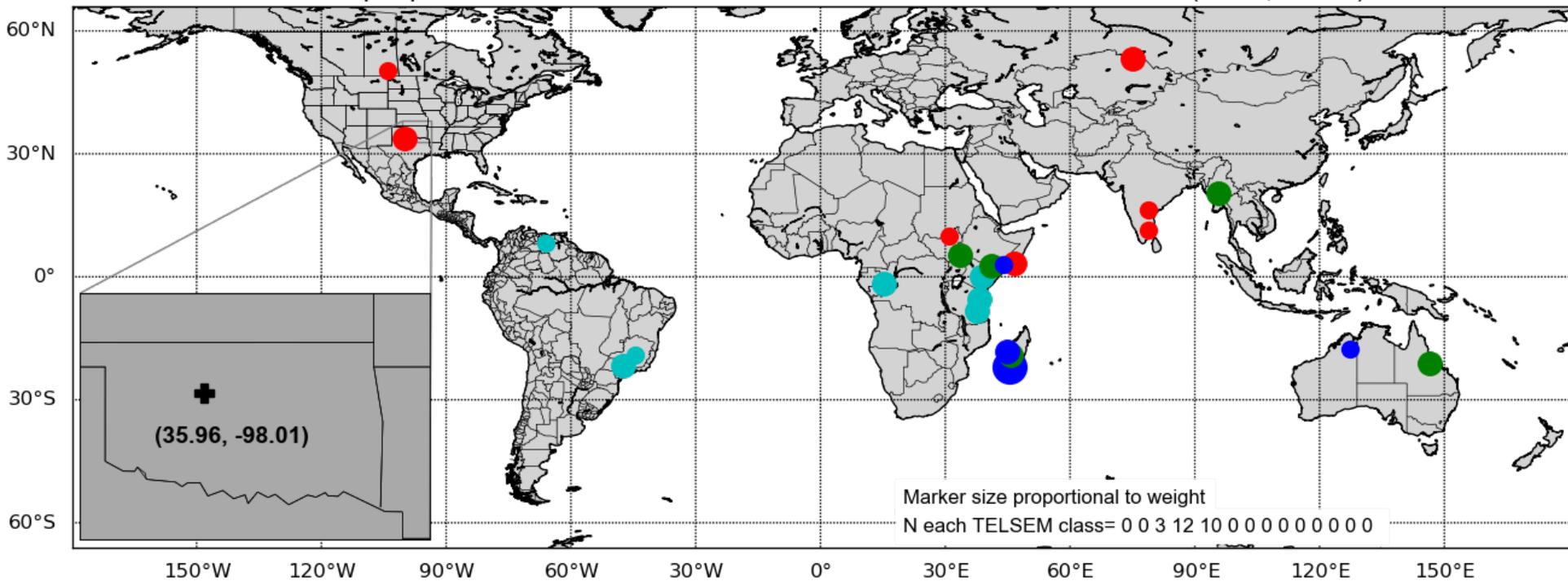
Geographical locations of top ranked *a-priori* dataset candidates (GMI pixel nearest 36N 100W)

Locations of top a-priori DB entries GPM GMI Rev=024290 2018/06/08 0043 UTC (36.0, -99.98) R=0.0



Geographical locations of top ranked *a-priori* dataset candidates (GMI pixel nearest 36N 98W)

Locations of top a-priori DB entries GPM GMI Rev=024290 2018/06/08 0043 UTC (35.96, -98.01) R=0.0



Sep-Oct-Nov

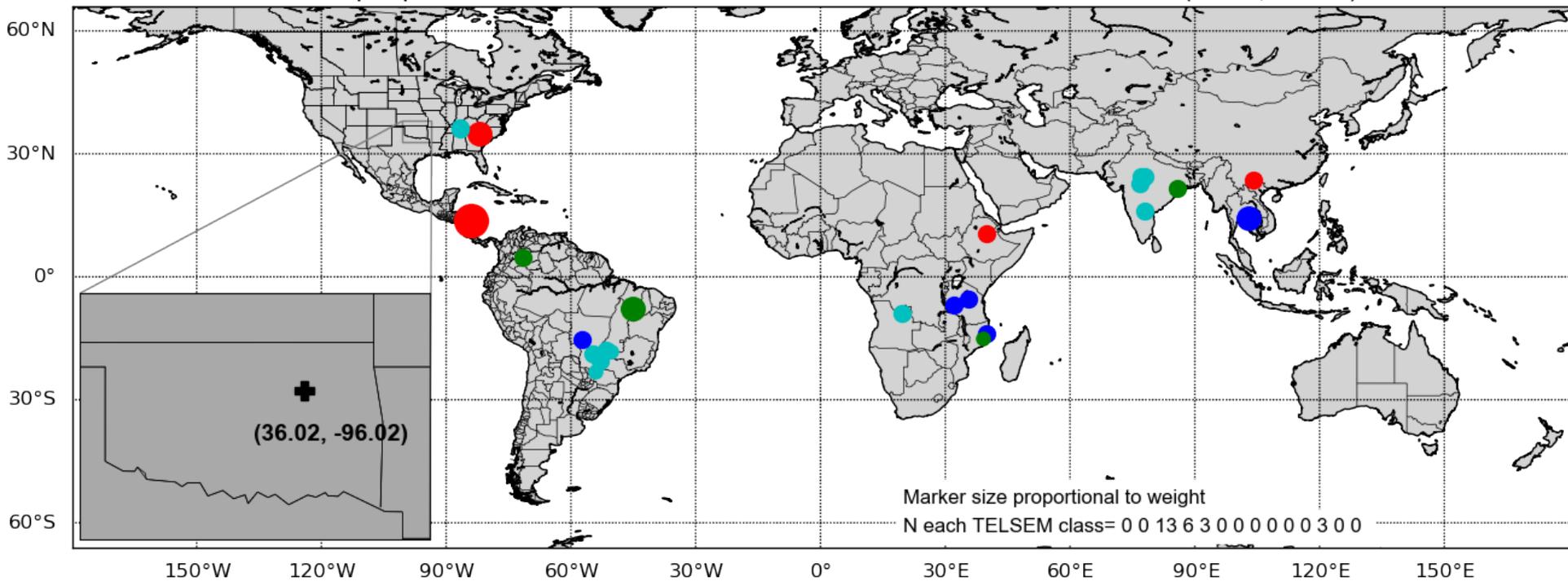
Dec-Jan-Feb

Mar-Apr-May

Jun-Jul-Aug

Geographical locations of top ranked *a-priori* dataset candidates (GMI pixel nearest 36N 96W)

Locations of top a-priori DB entries GPM GMI Rev=024290 2018/06/08 0043 UTC (36.02, -96.02) R=0.06



Sep-Oct-Nov

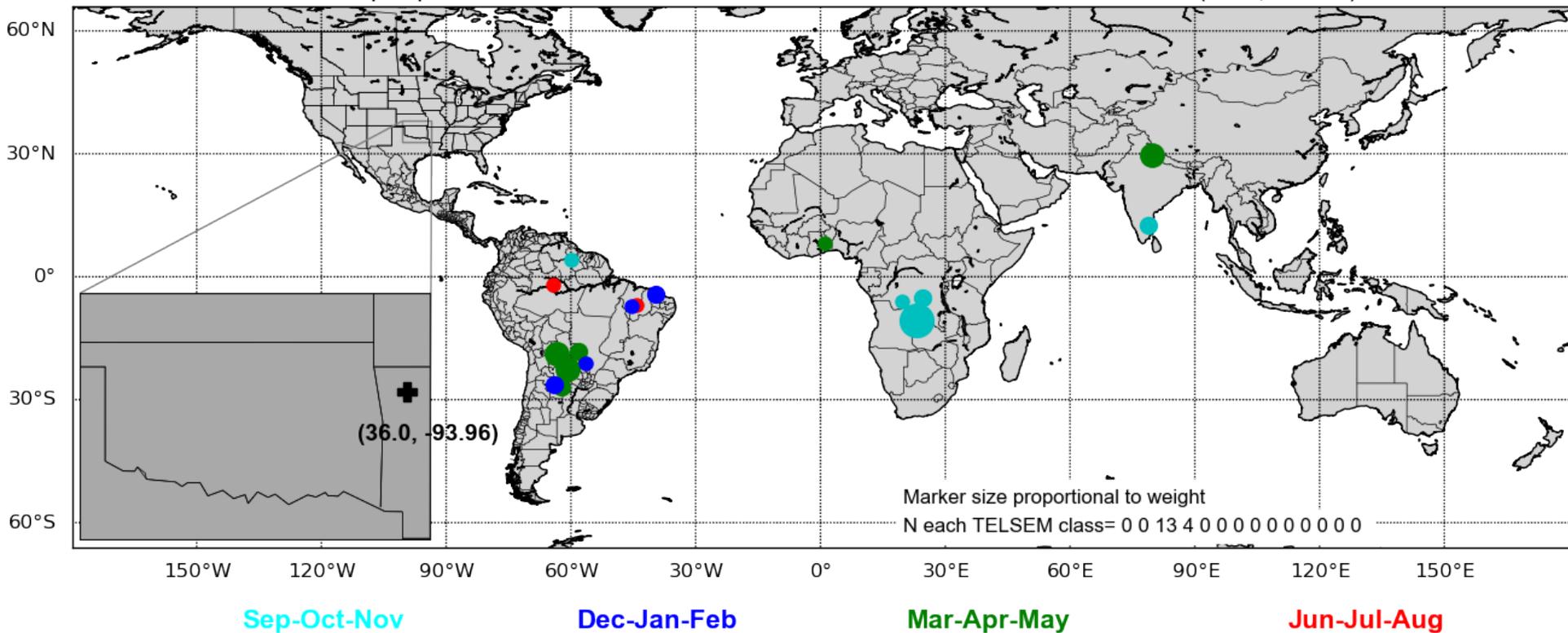
Dec-Jan-Feb

Mar-Apr-May

Jun-Jul-Aug

Geographical locations of top ranked *a-priori* dataset candidates (GMI pixel nearest 36N 94W)

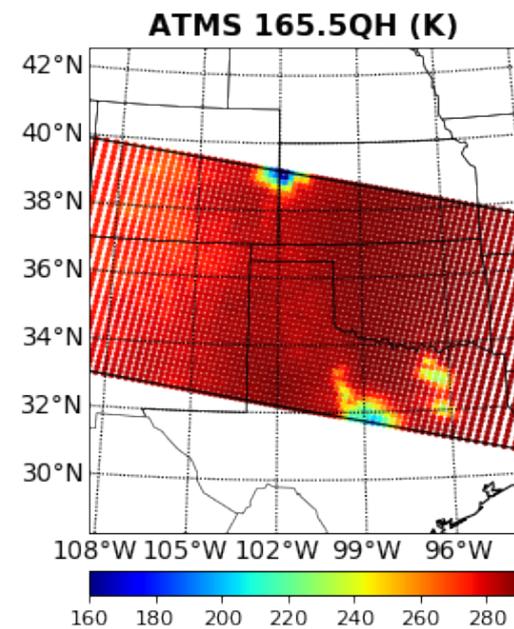
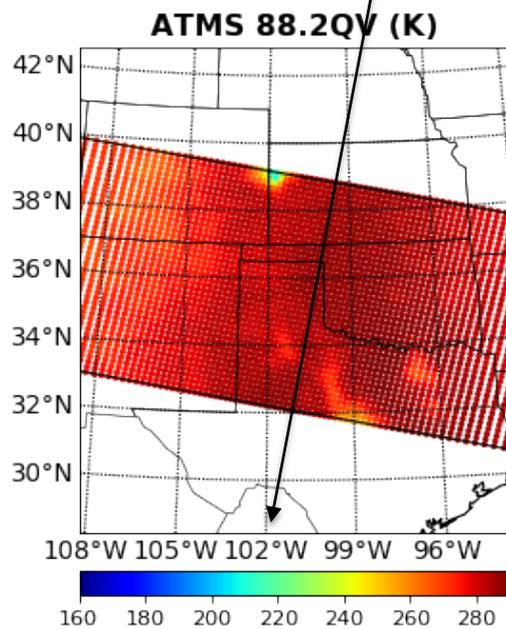
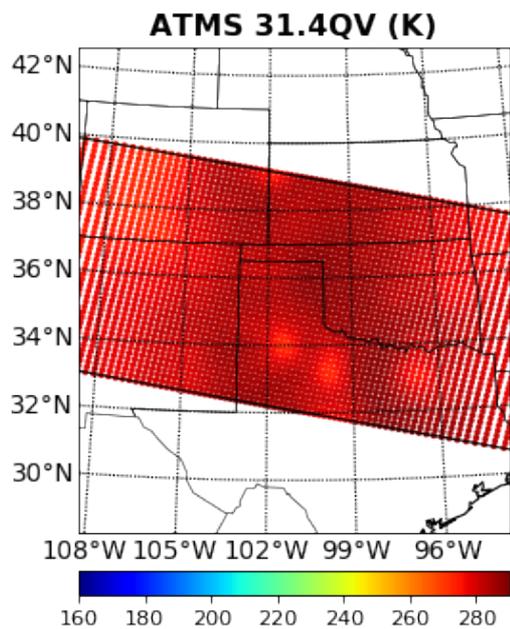
Locations of top a-priori DB entries GPM GMI Rev=024290 2018/06/08 0043 UTC (36.0, -93.96) R=0.0



NPP-ATMS Overpass

6 June 2018 0833 UTC

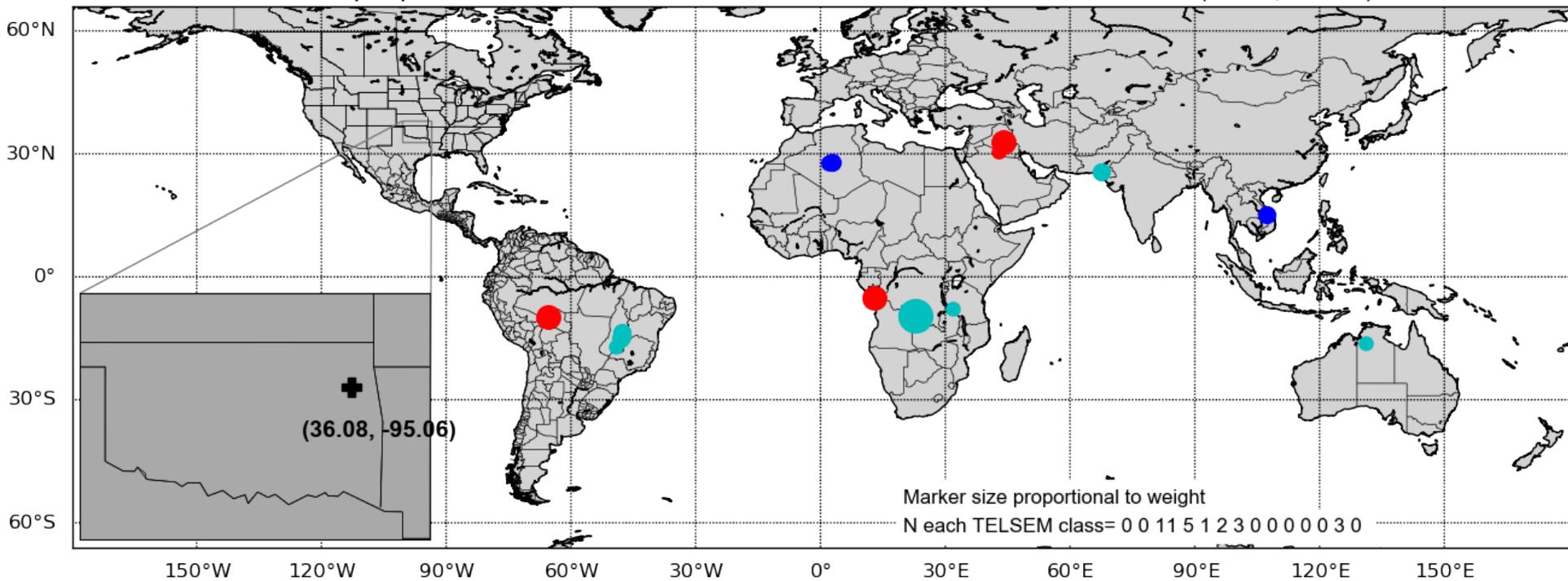
Nadir track



**Only 100 ATMS scanlines
were processed over this
area of interest**

Geographical locations of top ranked *a-priori* dataset candidates (ATMS pixel nearest 36N 95W)

Locations of top a-priori DB entries NPP ATMS Rev=034233 2018/06/06 0834 UTC (36.08, -95.06) R=0.0



Sep-Oct-Nov

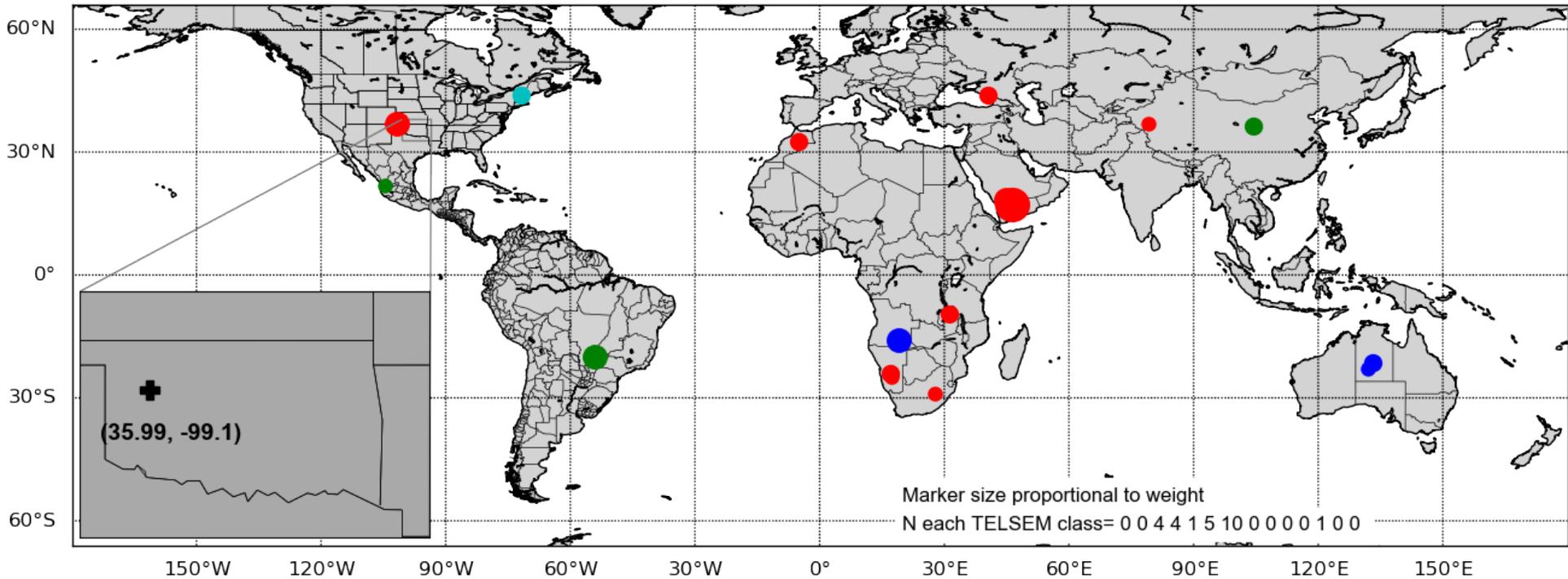
Dec-Jan-Feb

Mar-Apr-May

Jun-Jul-Aug

Geographical locations of top ranked *a-priori* dataset candidates (ATMS pixel nearest 36N 99W)

Locations of top a-priori DB entries NPP ATMS Rev=034233 2018/06/06 0834 UTC (35.99, -99.1) R=0.0



Sep-Oct-Nov

Dec-Jan-Feb

Mar-Apr-May

Jun-Jul-Aug

Summary

Only two examples were shown, from the favored central US location, where vegetation increases (decreases) eastward (westward) of 100W longitude, under non-raining conditions.

Did not have time to go into more details of precipitation characteristics, including vertical water content structure, over various surface types (Utsumi et. al. 2020 paper in review).

In future Part 2, I will show examples and precipitation profiles over other surface conditions, e.g., all water, mixed water/land, snow of different characteristics, etc.