Exploring Using Artificial Intelligence (AI) for Remote Sensing, NWP and Situational Awareness (SA)

S. Boukabara*, E. Maddy+, A. Neiss+, K. Garrett+, E. Jones+, K. Ide~, N. Shahroudi~ and K. Kumar+

*NOAA/NESDIS Center for Satellite Applications and Research (STAR) , College park, MD, USA

+ Riverside Technology Inc. (RTI) @ NOAA/STAR, College park, MD, USA

~ University of Maryland (UMD), College Park, MD, USA

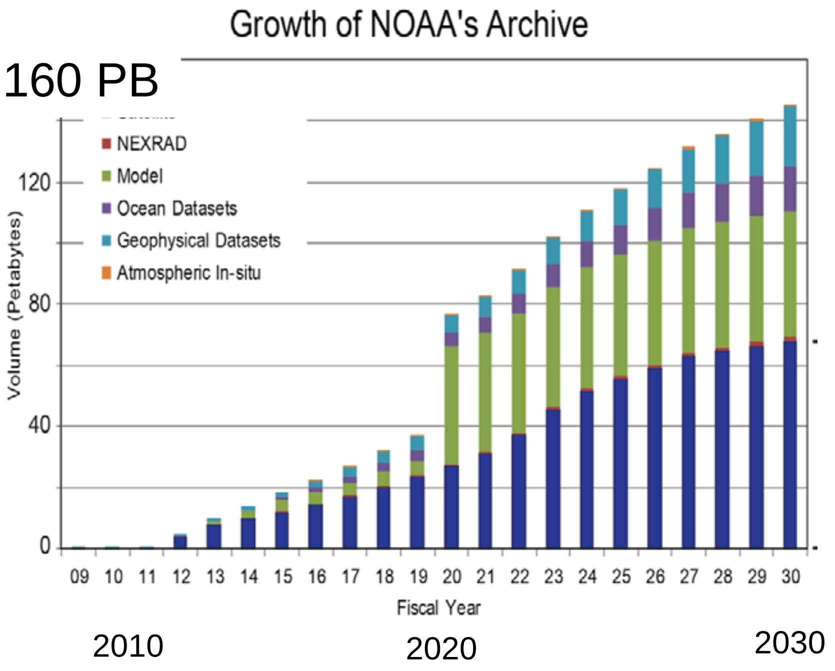
ITSC-XXI Conference, Darmstadt, Germany, Nov 29th – Dec 5th 2017

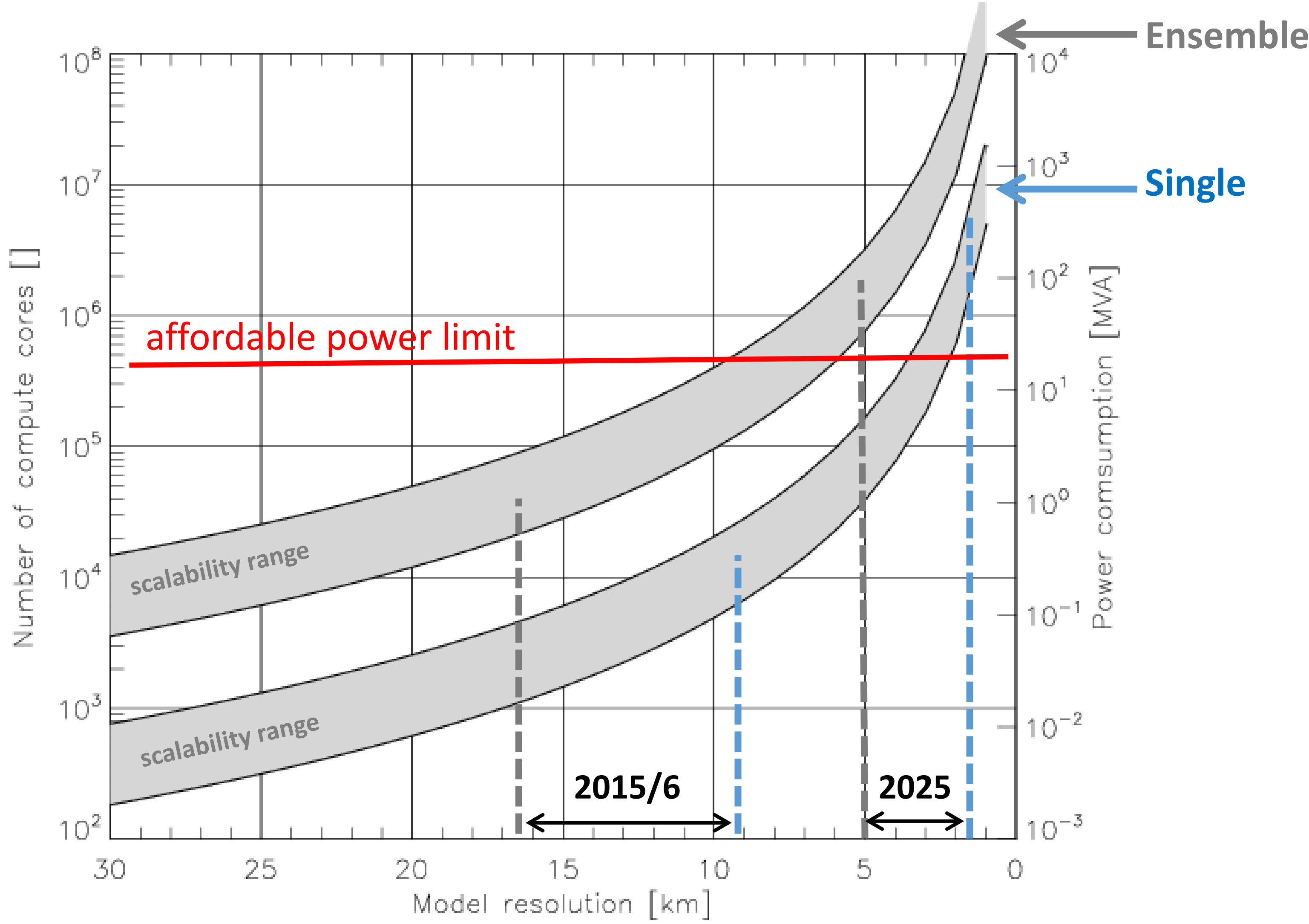
Agenda

1. Why Artificial Intelligence (AI) ? Background and Motivations
2. Methodology & Description
3. AI for Remote Sensing and Data Assimilation/Fusion/NowCasting
4. Conclusions

Expected Increase in HPC requirements and Data Volume

(for ECMWF NWP center: using currently 5-10% of satellite data)





GOS Trends:

- New Players in GOS (international, private, ..)
- New Sensors (higher resolutions,..)
- New technologies (small sats, etc)
- Emergence of New GOS (IoT, etc)
- **Significant Increase in data volume/diversity**
- Budget and HPC Constraints

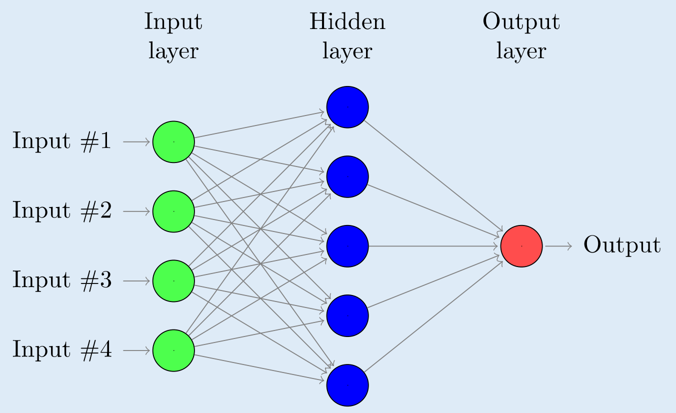
[ECMWF, Bauer et al. 2015]

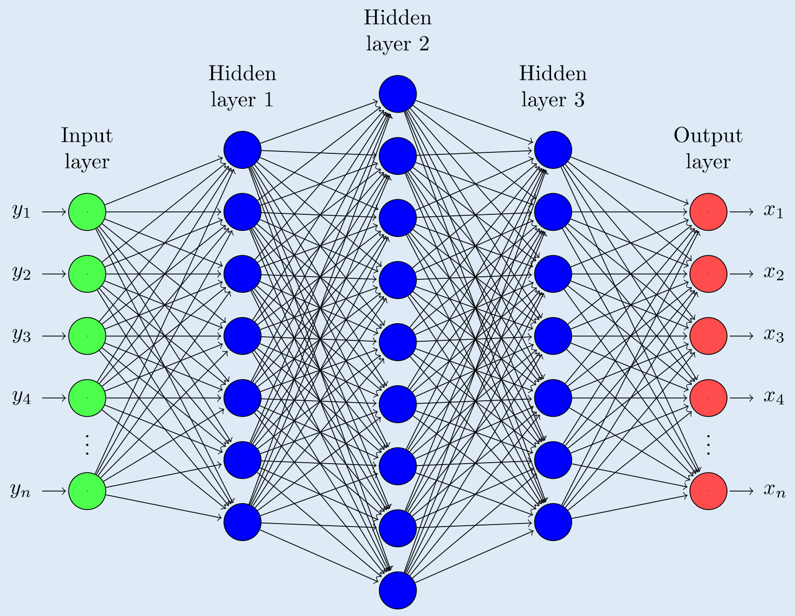
NOAA Data Volume graph, Courtesy Steve Del Greco & Ken Casey, NOAA/ NCEI (via Jeff de La Beaujardiere)

Why AI?

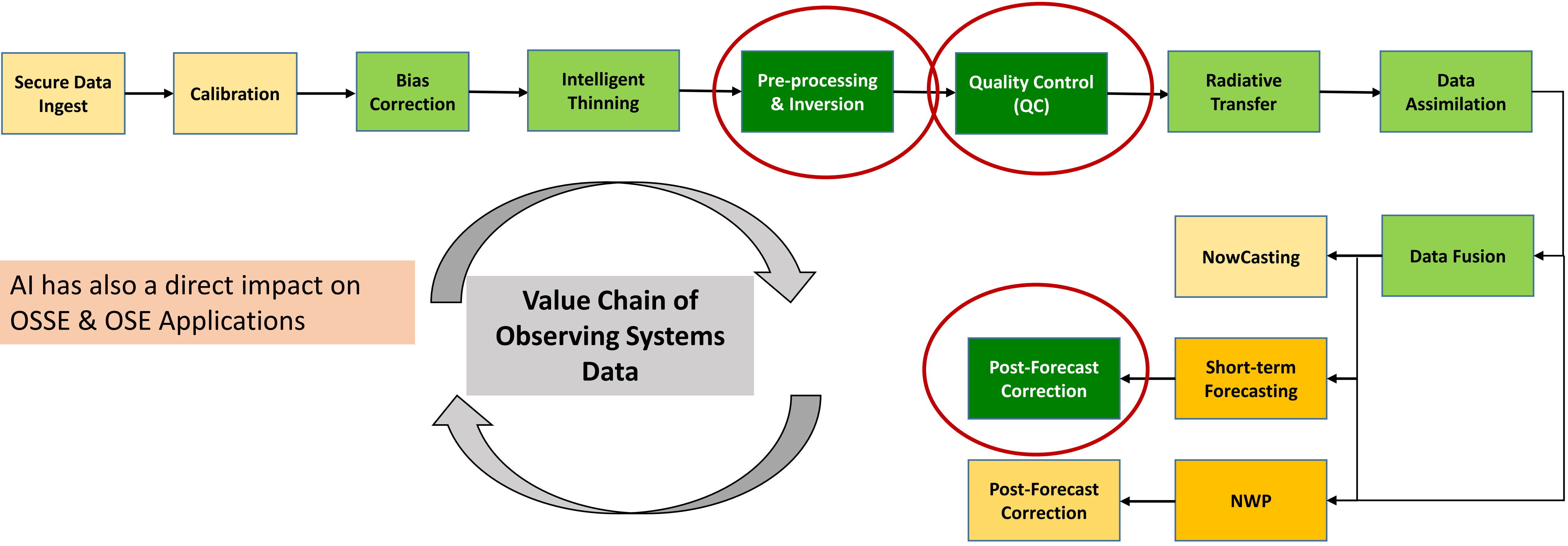
- **AI applied successfully in fields with similar traits as Environmental data & NWP/SA: (1) # obs. systems to analyze/assimilate/fuse and (2) predict behavior**
  - Medical field (Watson Project): Scan Image Analysis, Cancer detection, heart Sound analysis
  - In finance: Algorithmic Trading, market data analysis, portfolio management
  - In Music: Composing any style by learning from huge database & analyzing unique combinations.
  - Self-Driving Transportation Devices: Fusion of Multiple Observing Systems for situational awareness
  - …..
- **We believe Environmental data exploitation (remote sensing, data assimilation and perhaps forecasting), presents a viable candidate for AI application.**
- **<u>This presentation is meant to present a few examples to convey that the potential is significant.</u>**

<u>Neural Network vs Deep Learning (AI)</u>





Exploring AI for Remote Sensing, NWP & Situational Awareness (SA). Status

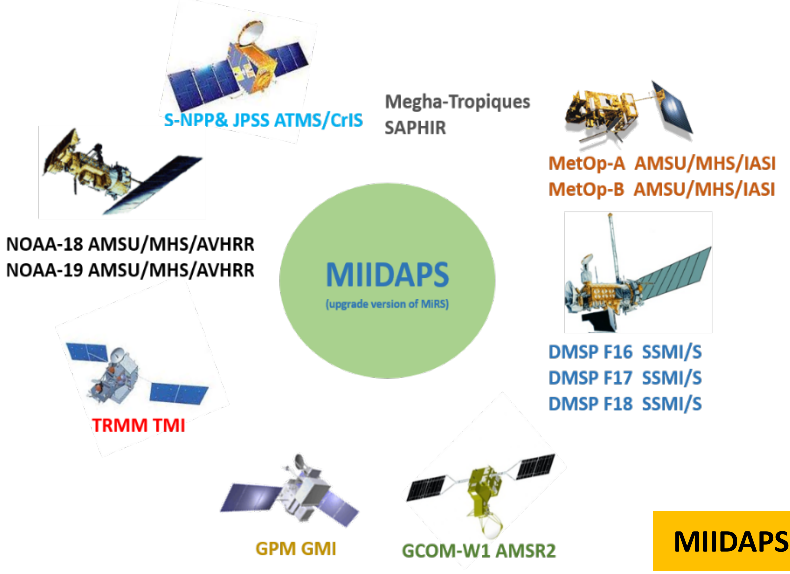


Not Tested. Unknown level of confidence | Not tested. Reasonable level of confidence | Tested or not: Moderate level of confidence | Tested. High level of confidence

Pilot Project: MIIDAPS-AI:

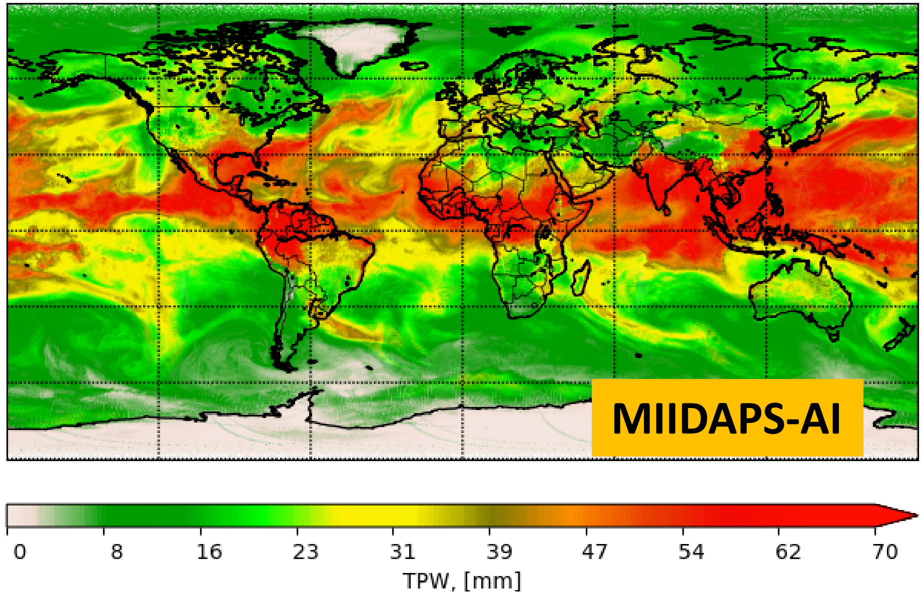
Multi-Instrument Inversion and Data Assimilation Preprocessing System

Exploring Artificial Intelligence for Remote Sensing/Data Assimilation/Fusion Applications

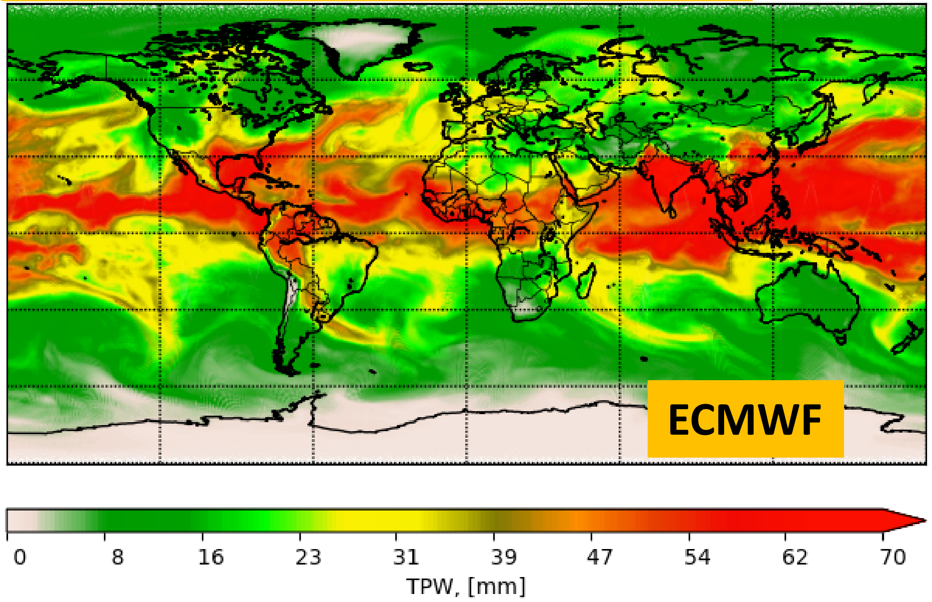


Google TensorFlow Tool used for MIIDAPS-AI

MIIDAPS-AI outputs (TPW) Using SNPP/ATMS Real Data



Reference source of TPW: ECMWF Analysis



How to assess that AI-based output (Satellite Analysis) is valid?

(1) Assessing quality by comparing against independent analyses

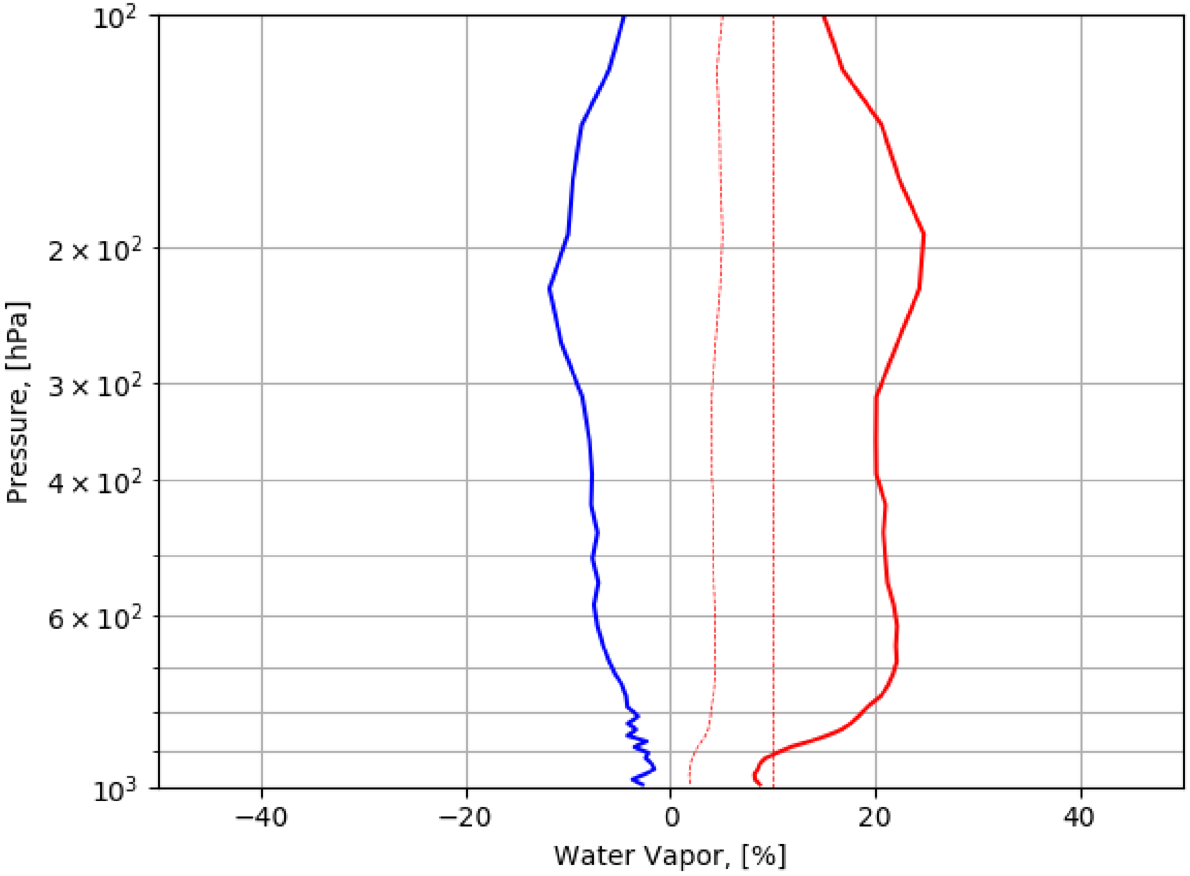
(2) Assessing Radiometric Fitting of Analysis

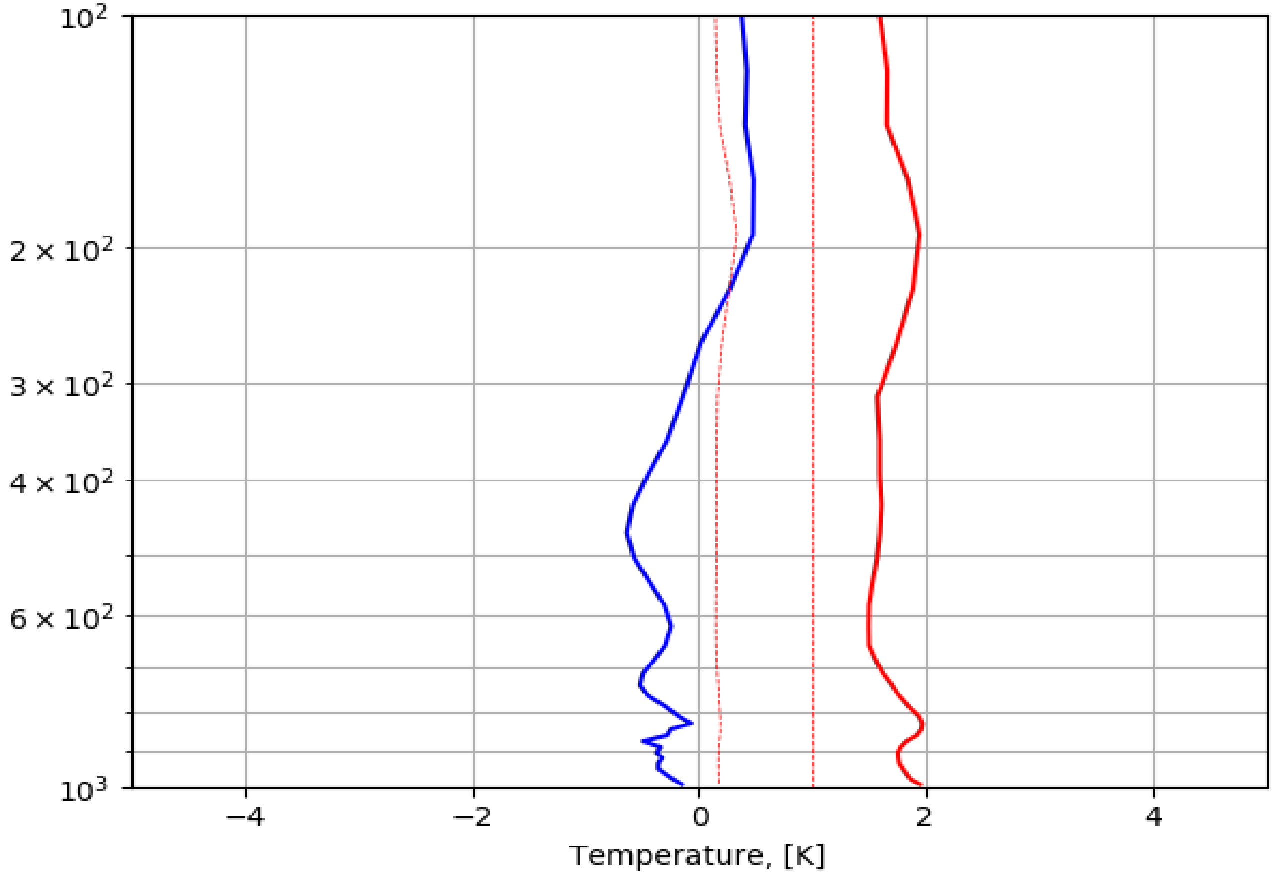
(3) Assessing analysis spatial coherence

(4) Assessing inter-parameters correlations

	MIIDAPS-AI	MIIDAPS
Processing Time for a full day data. A single sensor (ATMS). Excluding I/O	~5 seconds	~ 2 hours

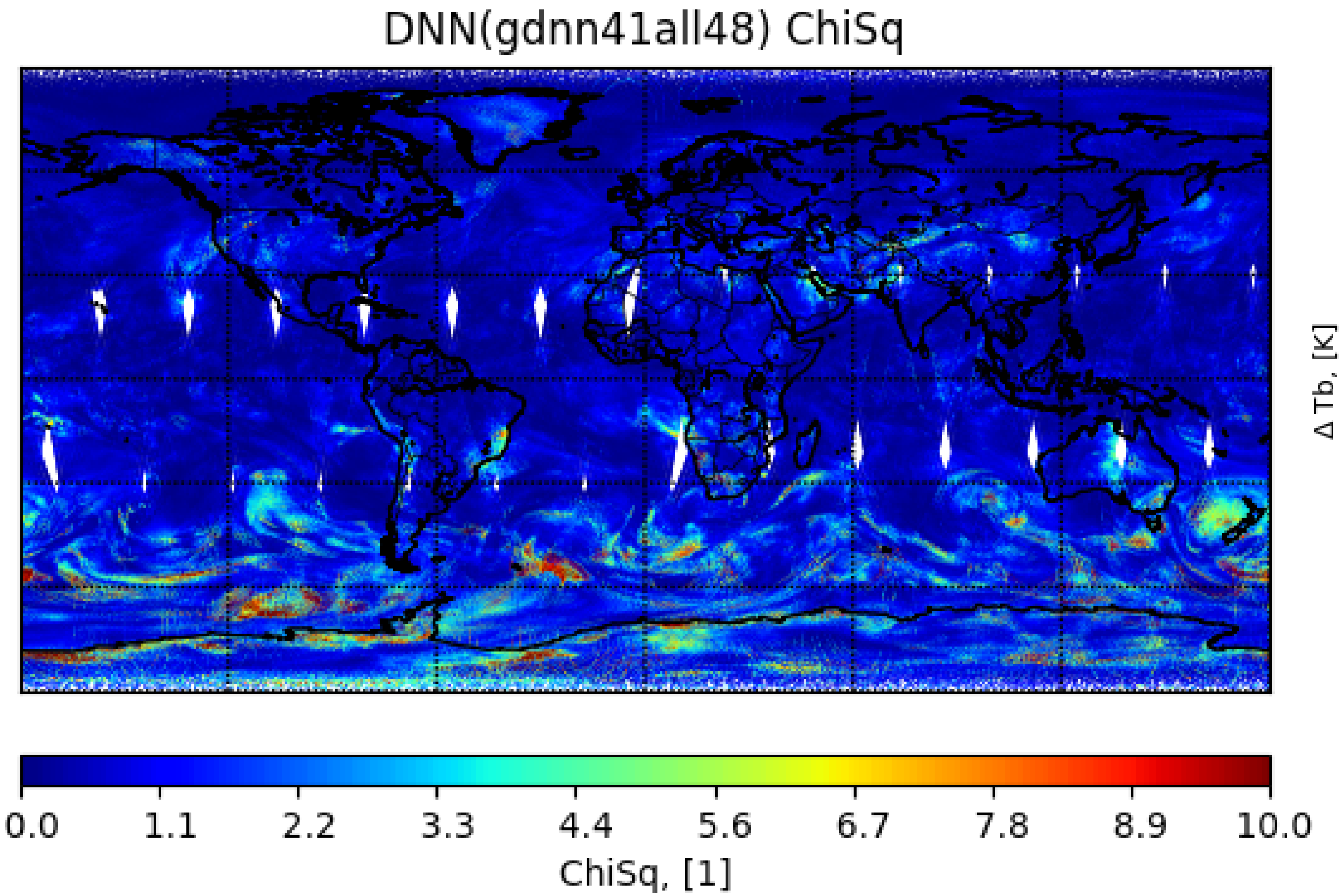
(1) Performance Assessment (T, Q)

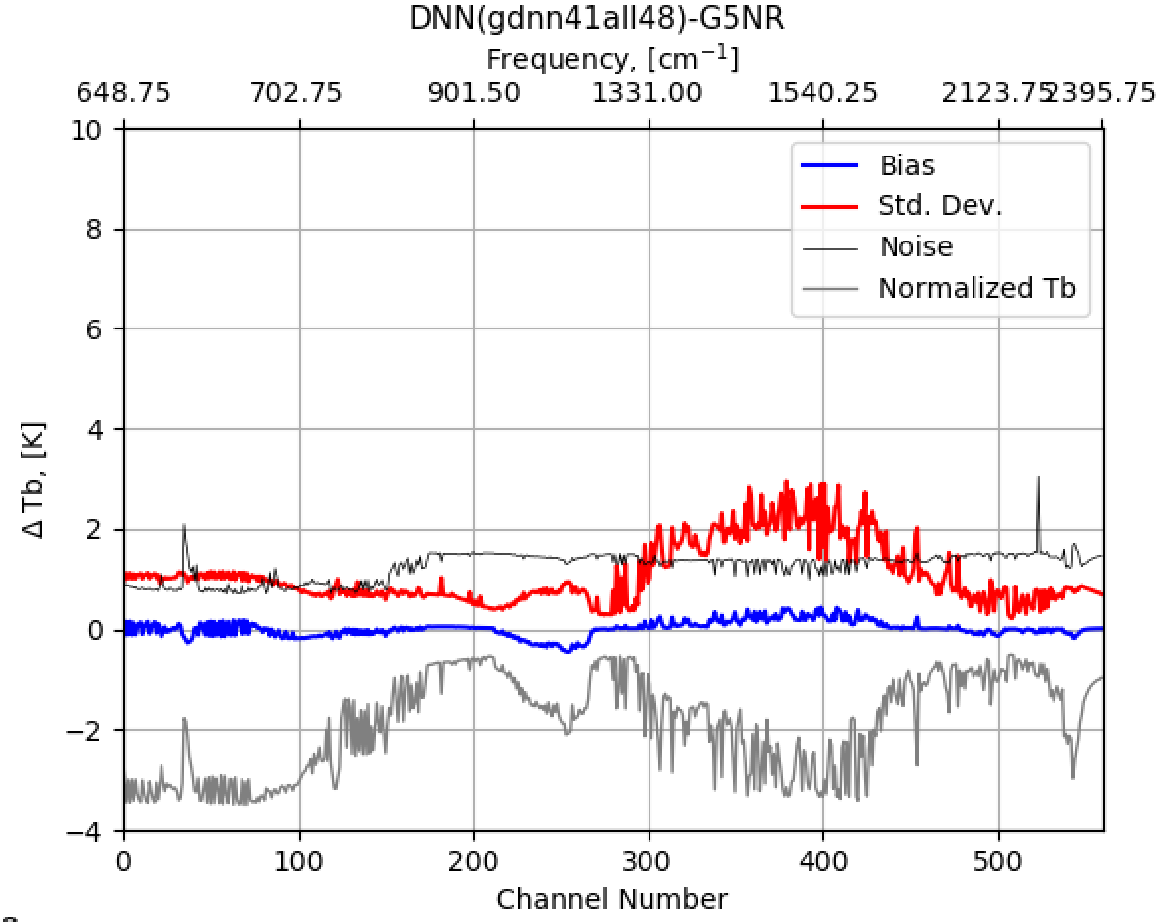




ECMWF used as independent reference set. Clear and cloudy points. All surfaces included.

(2) Convergence Assessment (CrIS Case)

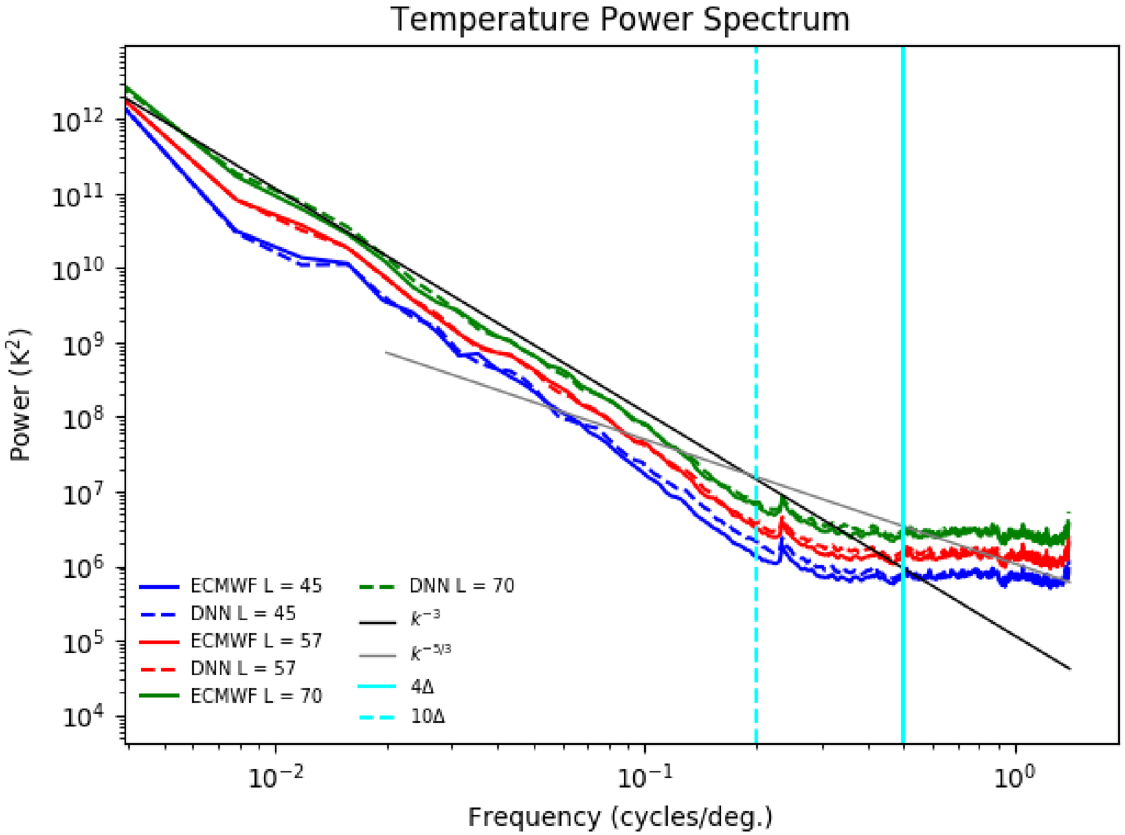




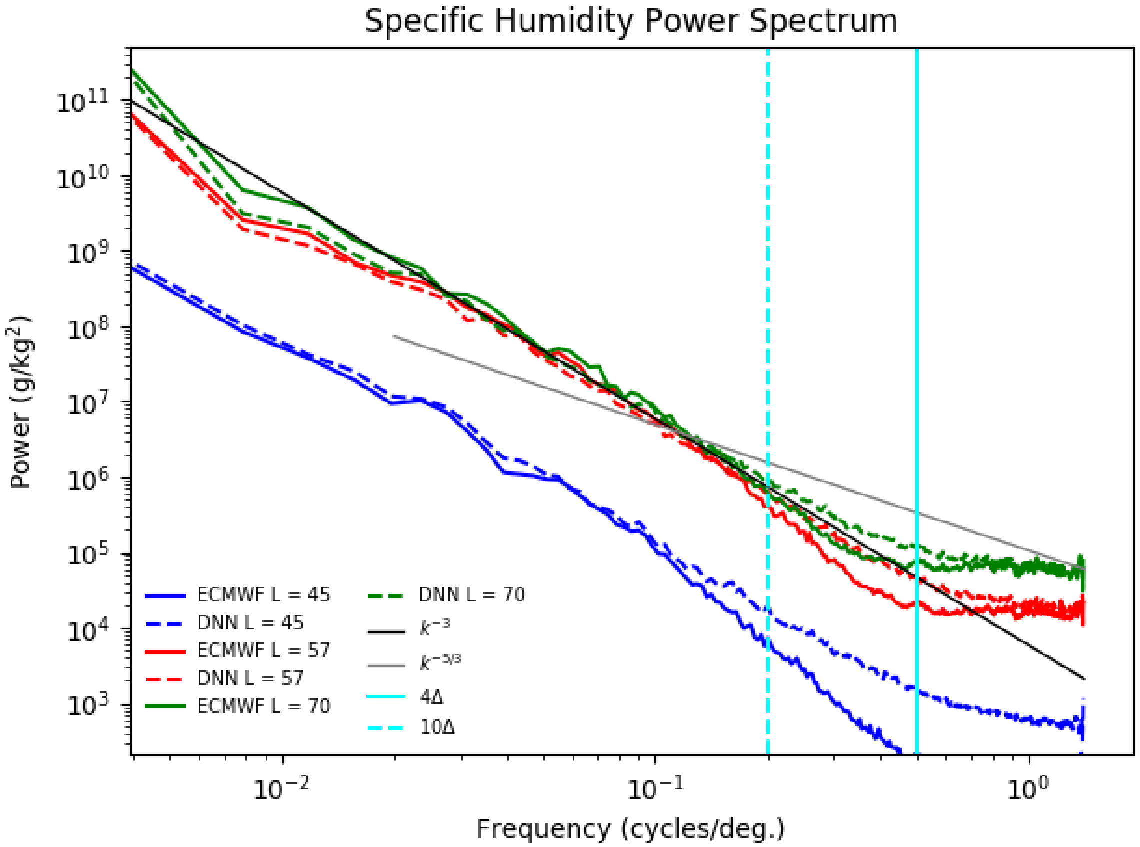
AI-based analysis is fed to CRTM and then simulation is compared to CrIS radiances

(3) Spatial Coherence Assessment

Temperature



Water Vapor

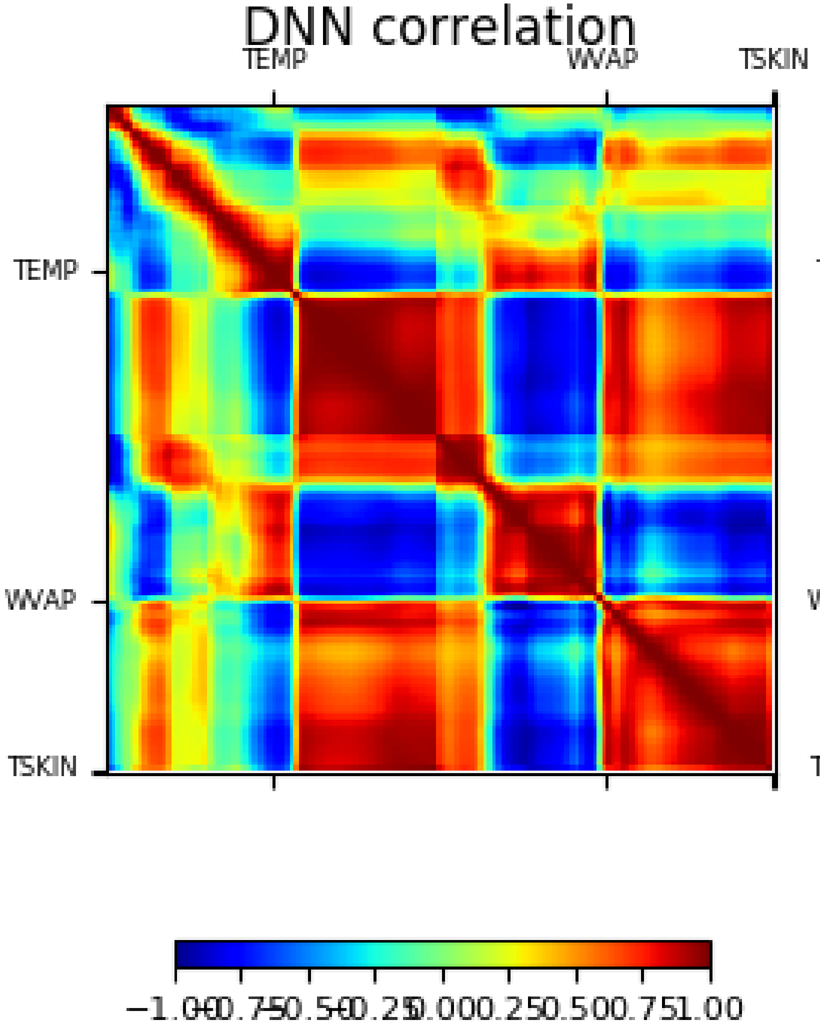


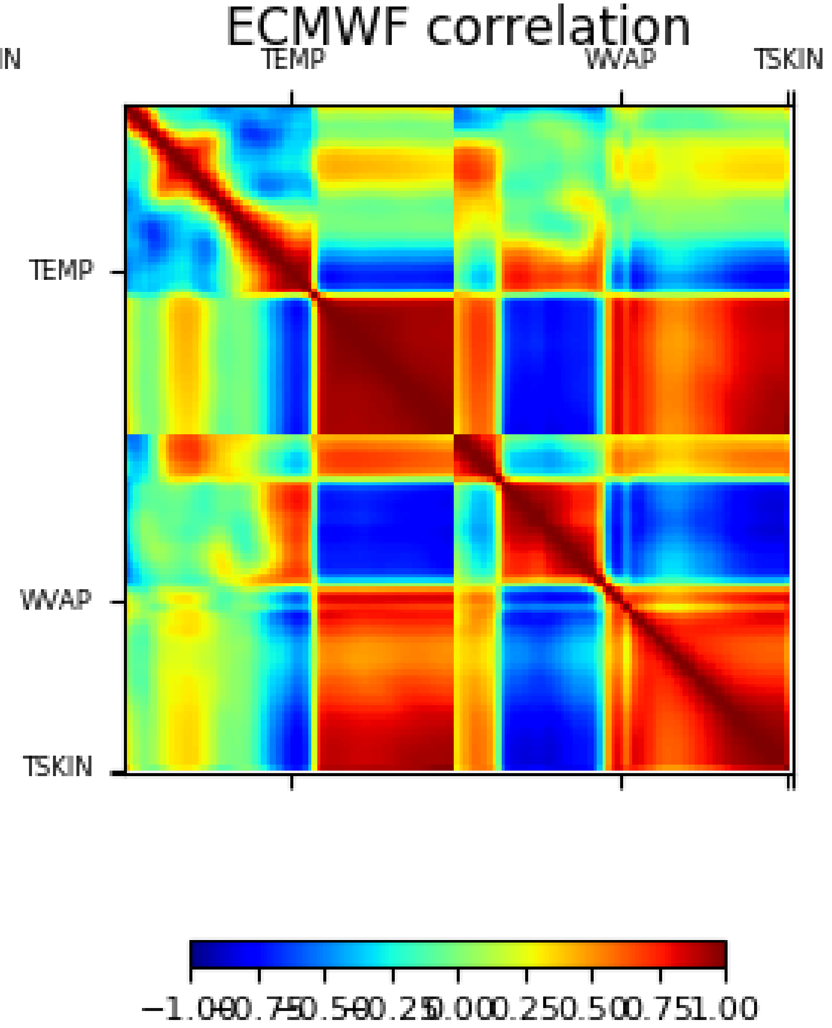
Water vapor fields and Temperature fields generated by AI (and satellite data) are consistent with those from ECMWF, except for high variability scales (as expected)

Spatial coherence – Global Temperature and Water Vapor 1D power spectrum from ATMS and ECMWF

(4) Inter-Parameters Correlation Assessment

AI-Based Algorithm vs ECMWF – ocean





Water vapor, temperature and Skin temperature generated by AI applied to ATMS are correlated with each other in a similar way that those same parameters obtained from an NWP analysis, are.

Can AI Be Used as Forward Operator?

Status:
- EOF of Geoph Data Used as Inputs
- Only clear sky was tested
- Only surface-blind channel tested
- ATMS tested. All channels together
- ~million points used:
- Jacobians need to be
- Quick test: CRTM use

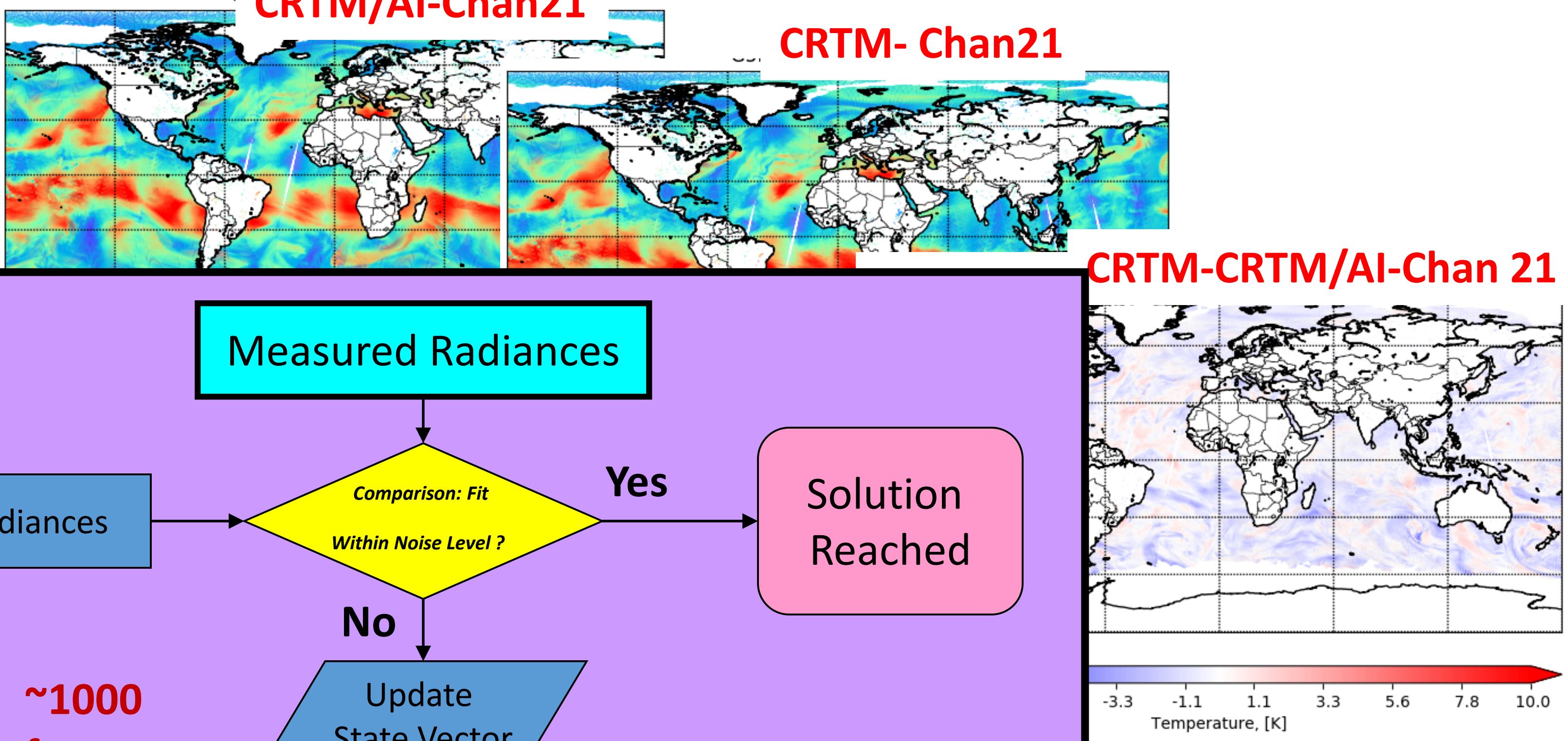
Potential Advantages:
- Multiple Orders of ma
- Allows using this in a setting (inversion, DA
- Is just an extension of implementation of tru (Line-By-Line Models)
- Does not Replace LBL training just like CRTM

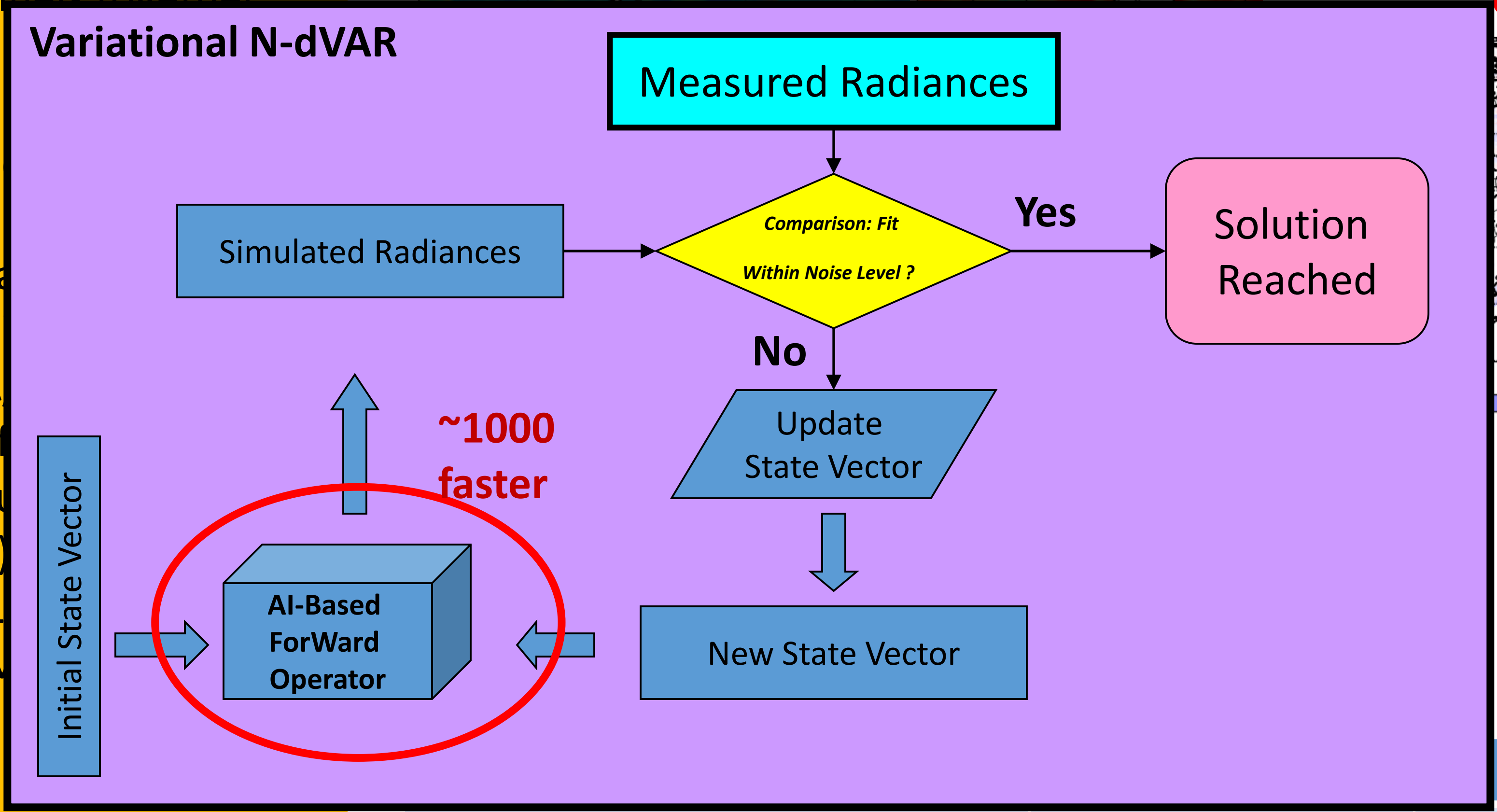
Next Steps:
- Use LBL as training
- Assess in variational setting
- Extend (cloudy, surface, IR, Jacob., etc)

CRTM/AI-Chan21

CRTM- Chan21

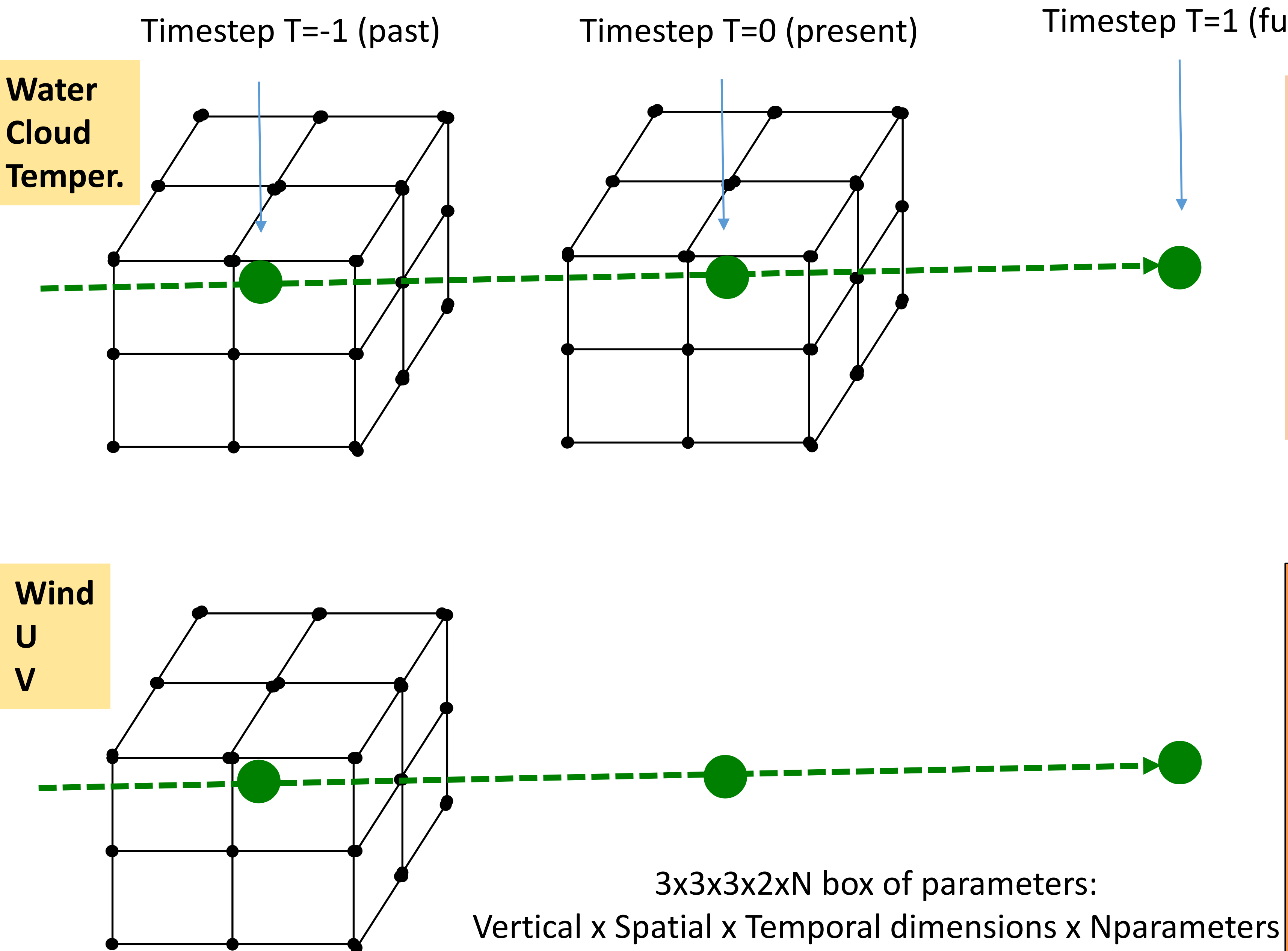
CRTM-CRTM/AI-Chan 21





		CRTM
Processing Time for a full day data. A single sensor channel(ATMS). Excluding I/O	<1 second	~ 1.3 hours

Does AI Have Predictive Applications?



3x3x3x2xN box of parameters:
Vertical x Spatial x Temporal dimensions x Nparameters

This simple model has potential to:

(1) Compute AMV from tracers (at t=0) based on spatial AND vertical tracing
(2) Correcting short-term forecast to adjust systematic errors and displacements (t=1 or 2, 3,…)
(3) NWP (t=N)

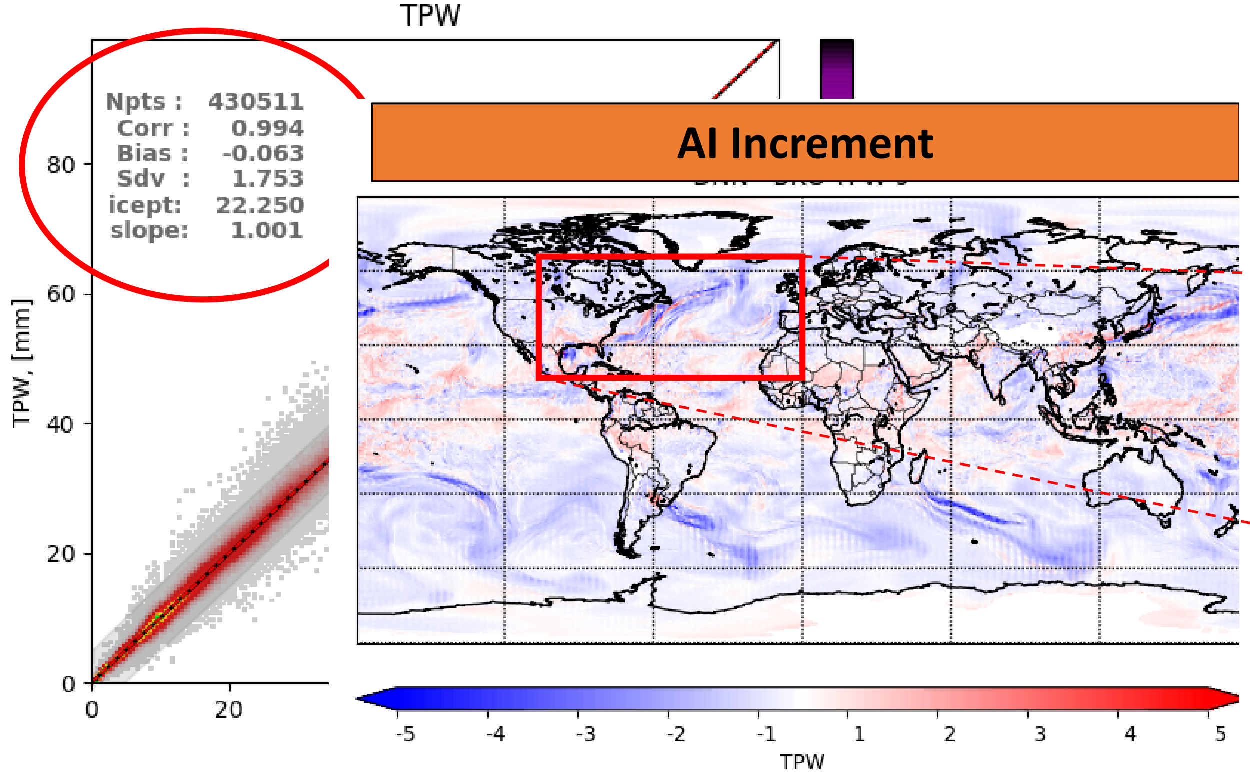
Questions:

Can we predict AMV center of box at T=0 timestep using the ~ 100 inputs parameters?

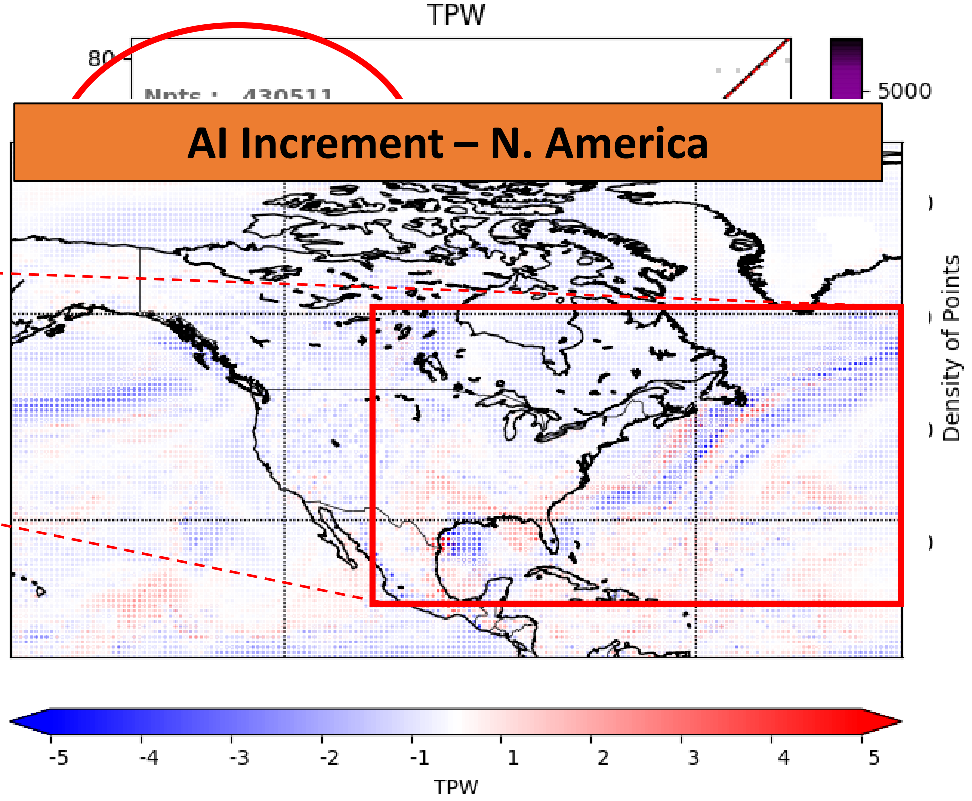
Can we improve prediction at Time step 1 if we set a target to match?

Correcting TPW Forecasting with AI?

ECMWF vs AI-corrected 6h fcst valid @ECMWF analysis time



ECMWF vs 6 hr frcst valid @ECMWF analysis time.



AI increment shows some dipoles indicating that the correction is adjusting the position of some features – Most notably the position of Harvey (Texas) and off the Eastern coast of N.America

Conclusions

- Increase in number, diversity and sources of global observing systems (GOS) including private sector. This presents unprecedented (and welcome) added resiliency and quality of the GOS. However this presents challenges: Cost and infrastructure to leverage/exploit them.

- Computing constraints, perhaps require us to explore new approaches for the future (not so distant). AI-Based Analyses (satellite-exlusive) are found to be radiometrically, spatially and geophysically consistent with traditional analyses.

- Goal of this study is not to show AI can do better, but that it can provide at least similar quality, much faster. It appears to be doing that.

- Different components can benefit from AI (Inversion, Data Assimilation, RT, QC, Data Fusion,..) for NWP and Situational Awareness SA.

- Encouraging results so far were found when assessing derivation of AMV using AI (not shown) and when assessing the feasibility of correcting GFS forecasts (using ECMWF as a target). Pointing to the potential for using AI for actual forecasting (at least short-term).

- Training is key for AI. Nature Run Datasets presents a good source for this.

- Pursuing AI applications, we believe, will allow us to :
    - (1) Reduce pressure on Infrastructure (ground systems), HPC and cost
    - (2) benefit from new environmental data (and face increased volume), including satellite data from all partners, including IoT
    - (3) Improve Latency
    - (4) Reduce cost of running legacy systems (remote sensing and data assimilation/fusion systems)
    - (5) Increase percentage of satellite data being assimilated (improved thinning, QC-ing, faster processing, etc)
    - (6) Reduce time to run OSE/OSSE and in general data assimilation/fusion systems, for decision making purposes
    - (7) Perhaps Improve forecast as a result of above and because AI can be exploited for forecast improvement

Methodology and Description

(baby steps)

- **Scope of the effort: RS and Forecasting Adjustment**
  - focus on satellite-based analyses (RS), focusing on an enterprise algorithm used for inversion and assimilation pre-processing
  - but also assess capability of short term forecast correction
  - focus on atmosphere (T, Q, Wind) but highlight surface parameters and hydrometeors capability as well
- **Tools:** Google TensorFlow
- **Real data**
  - Focus on SNPP/ATMS and SNPP/CrIS

Training & Verification:

- Sets: ECMWF Analyses, G5NR fields, GDAS Analyses
- Noise addition: uncorrelated, Gaussian distributed noise with spread of (instrument noise*2) is added to simulations
- Sampling: Training data is randomly selected from a fixed set of ~5% of a days worth of data in each training epoch
- Simple training (sample over a day generally
- Independent sets used for verification, but still the same period