Data Science and Analytics:
Data Fusion and Assimilation Panel

Key Questions:

1. What are the new data assimilation/fusion approaches that will likely lead to improved space weather forecasting performance?
2. Do you anticipate adequate data resources for these schemes? If not, how can data buys or other investments alleviate shortcomings?
3. How can we quantify uncertainty in data assimilation schemes that use multi-source observations?

Moderator: Charles Norton, SpaceWx-II Committee

Mark Cheung	*LMSAL*	*Senior Staff Physicist, Solar and Astrophysics Lab*
Bernard Jackson	*UCSD*	*Research Scientist, Center for Astrophysics and Space Sciences*
Slava Merkin	*JHUAPL*	*Principal Scientist*
Alex Chartier	*JHUAPL*	*Ionospheric Scientist*
Tomoko Matsuo	*UC Boulder*	*Assistant Professor, Smead Aerospace Department*
Erik Blasch	*AFOSR*	*Program Officer for Dynamic Data and Information Processing*

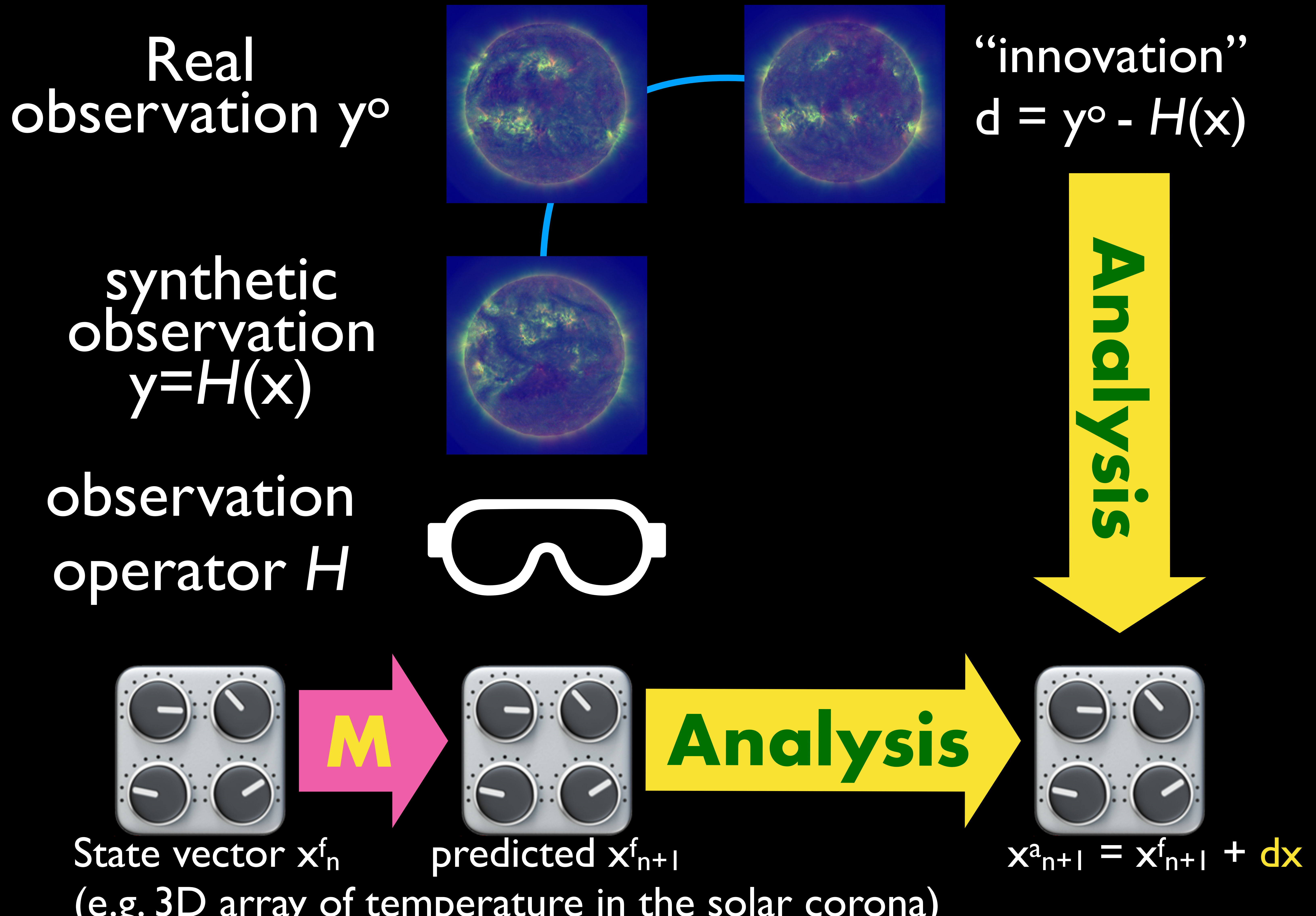
Space Weather Operations and Research Infrastructure Workshop: Phase II, Wednesday April 13, 2022, 1315 EDT

1. What are new data assimilation/fusion approaches that will likely lead to improved space weather forecasting performance?

2. Do you anticipate adequate data resources for these schemes? If not, how can data buys or other investments alleviate shortcomings?

3. How can we quantify uncertainty in data assimilation schemes that use multi-source observations?

Data Assimilation



How do we use the discrepancy **d** **("innovation")** to update **x,** and **quantify** uncertainties?

How to compute **dx** to nudge state to a physically plausible solution?

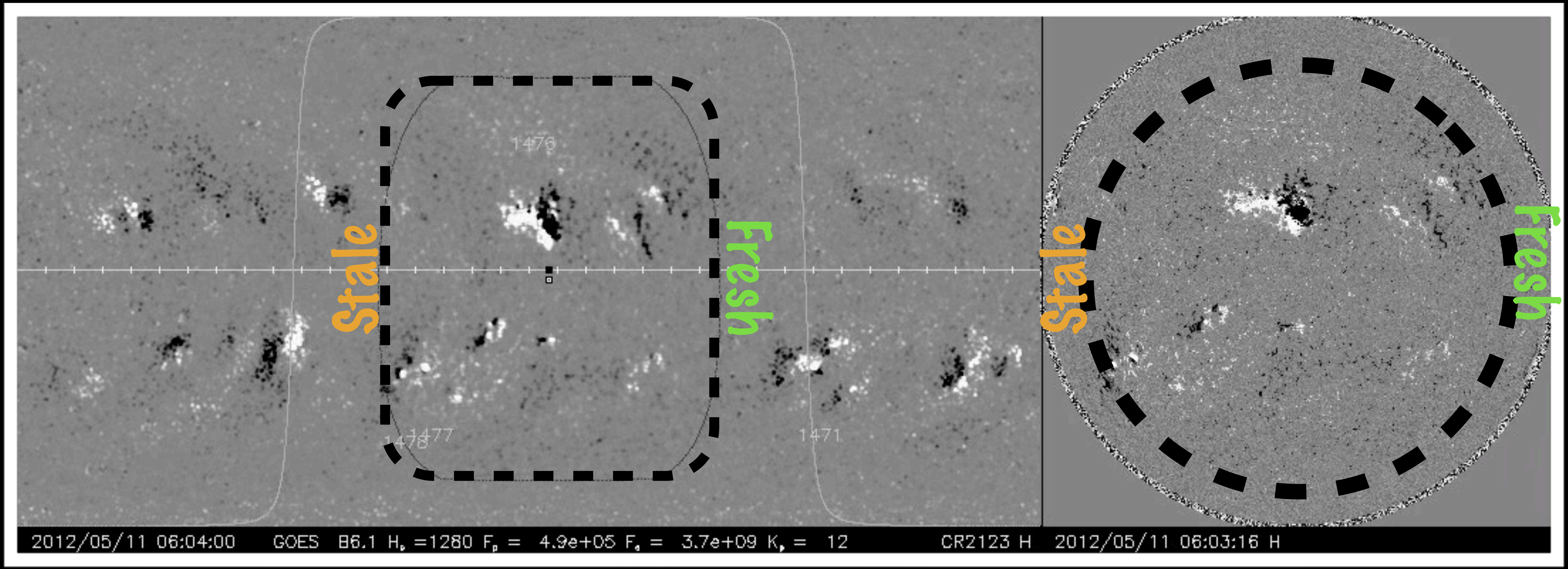
following notation of Ide, Courtier, Ghil & Lorenc (1997)

1. What are new data assimilation/fusion approaches that will likely lead to improved space weather forecasting performance?

- The large dimensionality of SpWx predictions problems require efficient, scalable DA (i.e. inference) techniques:
  - Density estimation and data generation (e.g. normalizing flows; Papamakarios+, 2021; JMLR);
  - Physics-informed Neural Networks (Raissi+ 2019, J. Comp Phys, 378, 686) ,
  - Neural Network-based Surrogate Models (e.g. Fourier Neural Operator; Li+, 2021, ICLR)
  - Simulation-based / Likelihood-free inference (Kranmer+, 2020; PNAS, 117, 48)
  - Automatic differentiation (Baydin et al., 2018, JMLR, 18, 1)

2. Do you anticipate adequate data resources for these schemes? No

Magnetic patterns off the east limb are stale



Magnetic patterns off the west limb are fresh

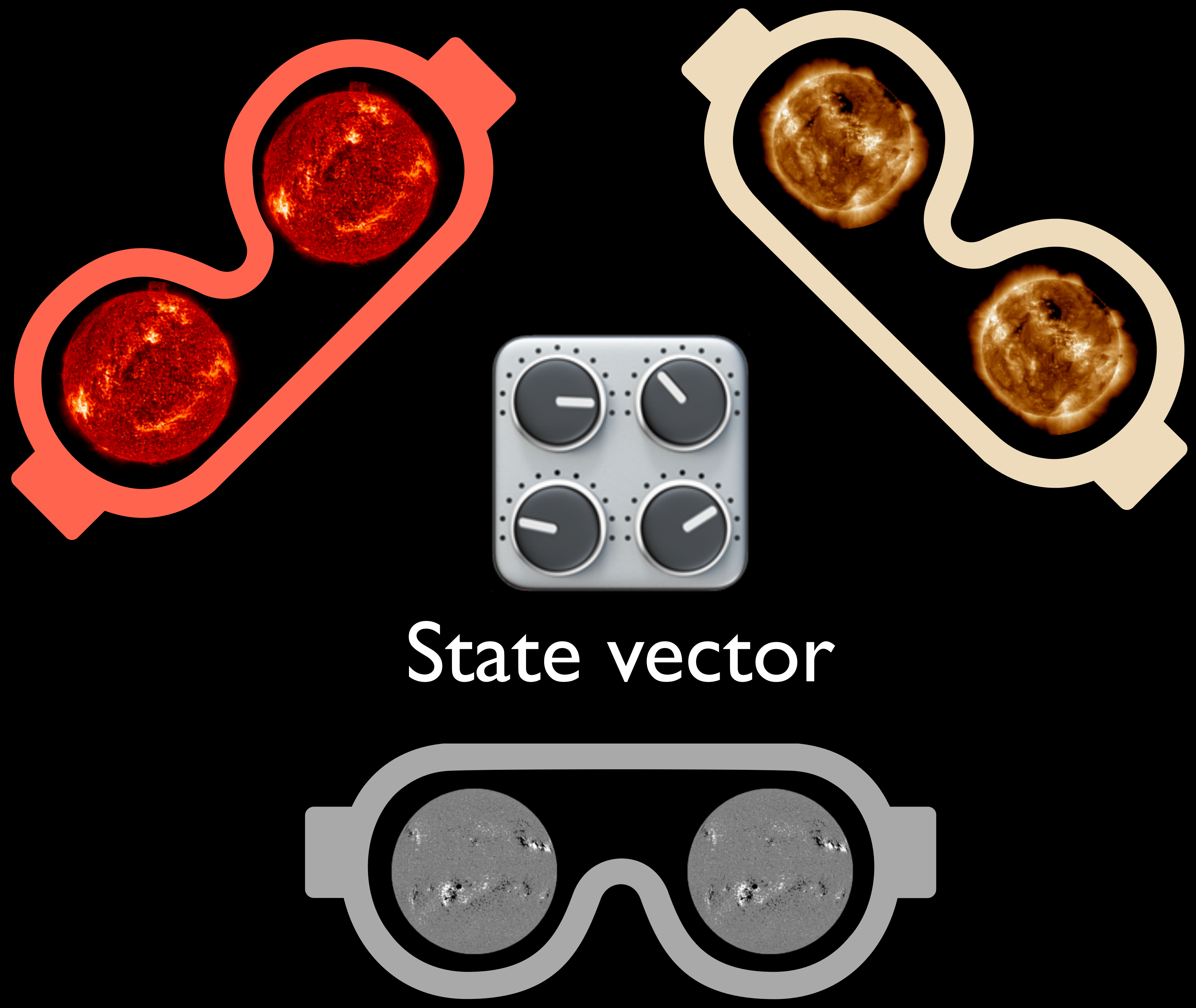
- Lack of coverage of solar farside coverage misses active region evolution
- Lack of understanding reliable polar field measurements (including cross-instrument calibration errors)

How can data buys or other investments alleviate shortcomings?

- Miniaturization; rideshares; lower cost of magnetograph data via data buy (mission dev and ops not agency-managed).
- Don't take the Solar Dynamics Observatory for granted. Invest now (ngGONG; future space borne vector magnetographs).

3. How can we quantify uncertainty in data assimilation schemes that use multi-source observations?

The DA framework is not limited to single-source measurements. Multiple *H* operators can be used to generate synthetic observations. Differentiable programming techniques mentioned previously can be used for UQ from multi-source data.



Approximate Bayesian Computation (ABC) methods using domain-relevant cost functions involving multi-source data can be used to assess inference quality and model selection.

Heliospheric Space Weather Predictions and Forecasts

Bernard V. Jackson (bvjackson@ucsd.edu)

Data Assimilation/Fusion Approaches Now and in a Few Years:

1) Provide all inner heliospheric monitors into the remote sensing mix to provide better global heliospheric analyses.
2) Add more worldwide Interplanetary Scintillation (IPS) Stations (WIPSS) to those currently existing. The UCSD analyses allow this now.

- **Longer Term (3-5 years from now and more):**

1) Add remote sensing data (Thomson-scattering brightness and speed data) from heliospheric imagers (STEREO HI works now, but has too low a latency): tests needed.
2) Provide heliospheric imagers that can view and add remote sensing data (ASHI, ESA Virgil, PUNCH, for fields maybe Faraday Rotation ? ground based or active space experiments)

- **Quantifying uncertainty:**

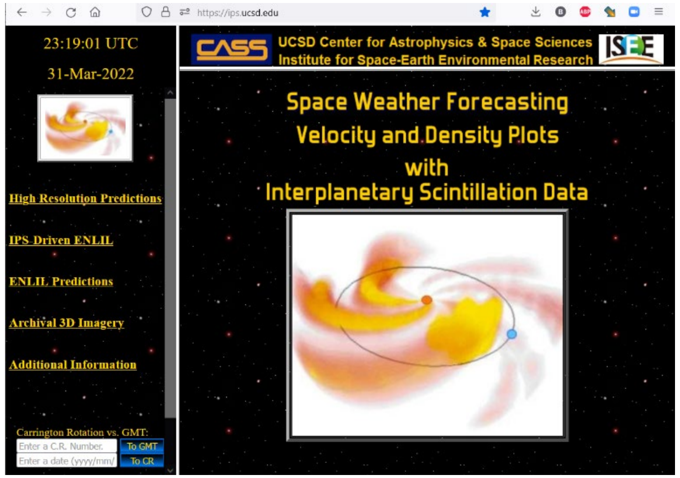
1) Pearson's "R" correlations operated both for data to the present and into the future compared with in-situ measurements works pretty well with caveats.
2) Tests provided by contingency tables that give hit rates, probabilities of detection, and false alarm rates do a pretty good job of relating different effects with enough samples.

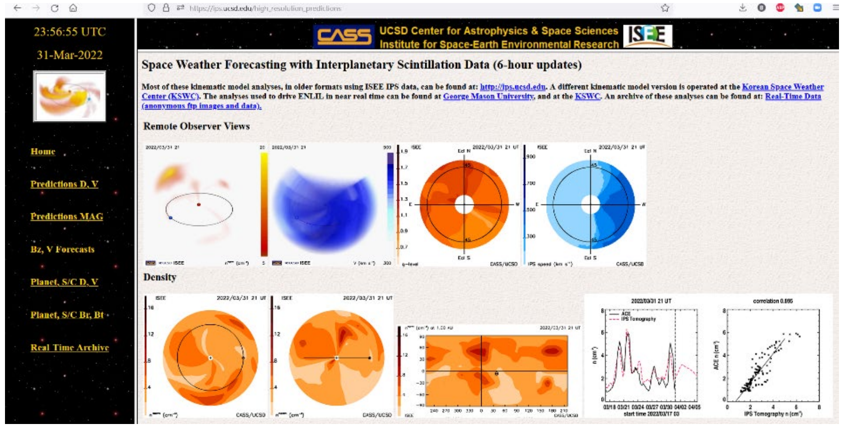
Heliospheric Solar Wind Predictions and Forecasts

Jackson, B.V., *et al.*, 2011, *Adv. in Geosciences,* 30, 93-115; Jackson *et al.*, 2013, *Solar Phys.*, 258, 151-165; Jackson et al., 2020 Frontiers, doi:10.3389/fspas.2020.568429

https://ips.ucsd.edu/   https://ips.ucsd.edu/high_resolution_predictions/   https://ips.ucsd.edu/experimentalforecasts

UCSD and other IPS Web pages (2022)





Websites:

ISEE: http://stsw1.stelab.nagoya-u.ac.jp/index-e.html

USCD: https://ips.ucsd.edu/high_resolution_predictions

CCMC: https://iswa.ccmc.gsfc.nasa.gov/IswaSystemWebApp/

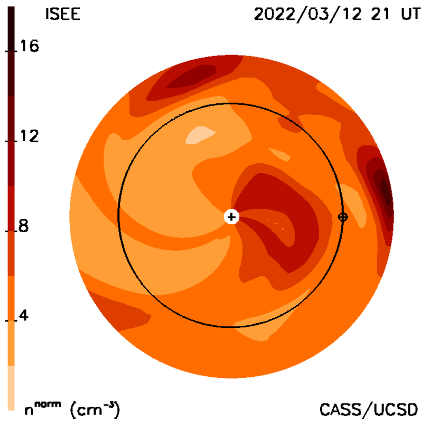
KSWC: http://www.spaceweather.go.kr/models/ips

GMU: http://spaceweather.gmu.edu/projects/enlil/

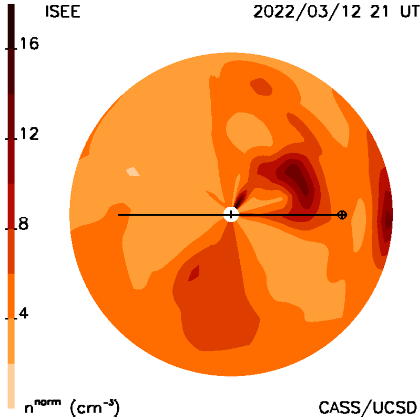
Soon at the UK Met

(Python scripting of UCSD analysis finished,
Siegfried Gonzo, UK Met)
ENLIL Python scripting underway?

CME Forecast



Ecliptic cut

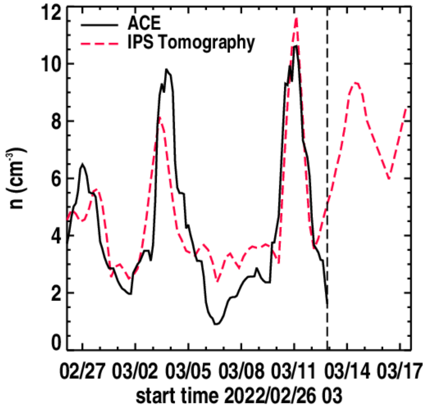


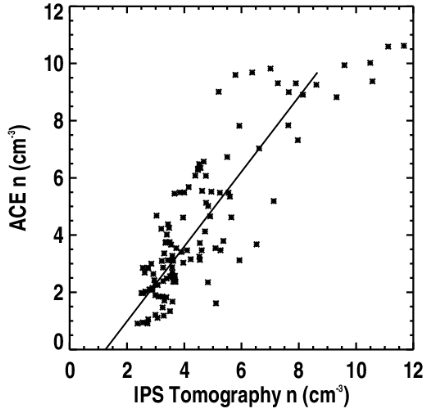
Meridional cut

Time Series Comparison

GSM Bz Forecast

Pearson's 'R' Correlation





Web analysis runs "automatically" using Linux on a P.C.

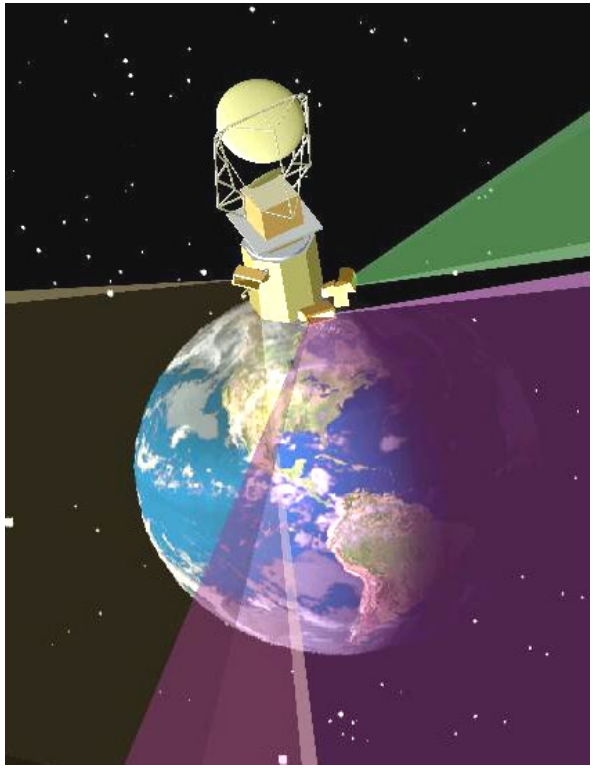
Solar Wind Prediction Analyses

Jackson et al., 2020 doi: 10.3389/fspas.2020.568429

Time Series at Earth at a 1.5 -Hour Cadence Resolution

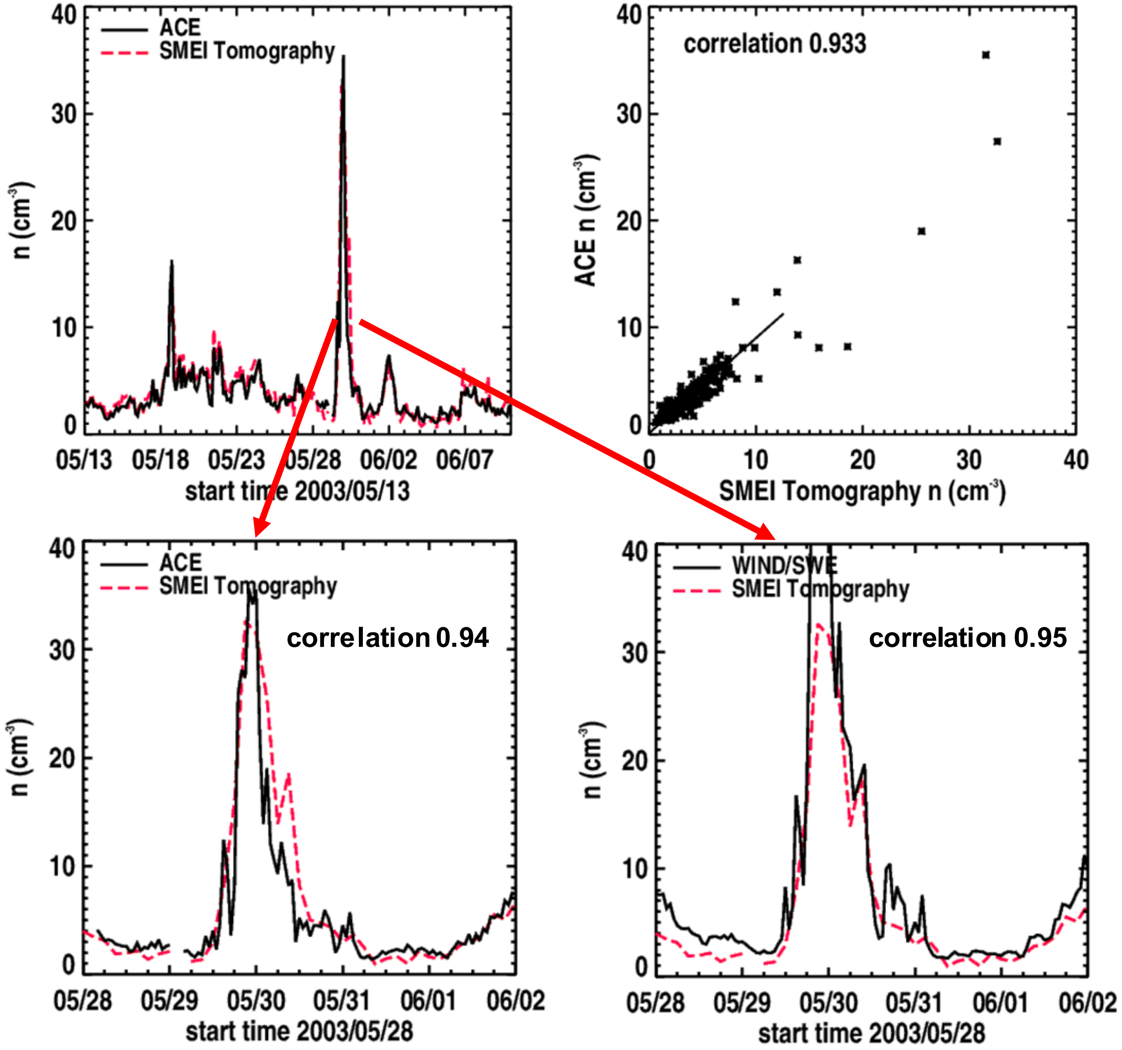
The Solar Mass Ejection Imager (SMEI)

US Air Force - NASA Project



SMEI Analysis

New SMEI Analysis ~ 1.5 hour cadence



Solar Wind Prediction Analyses

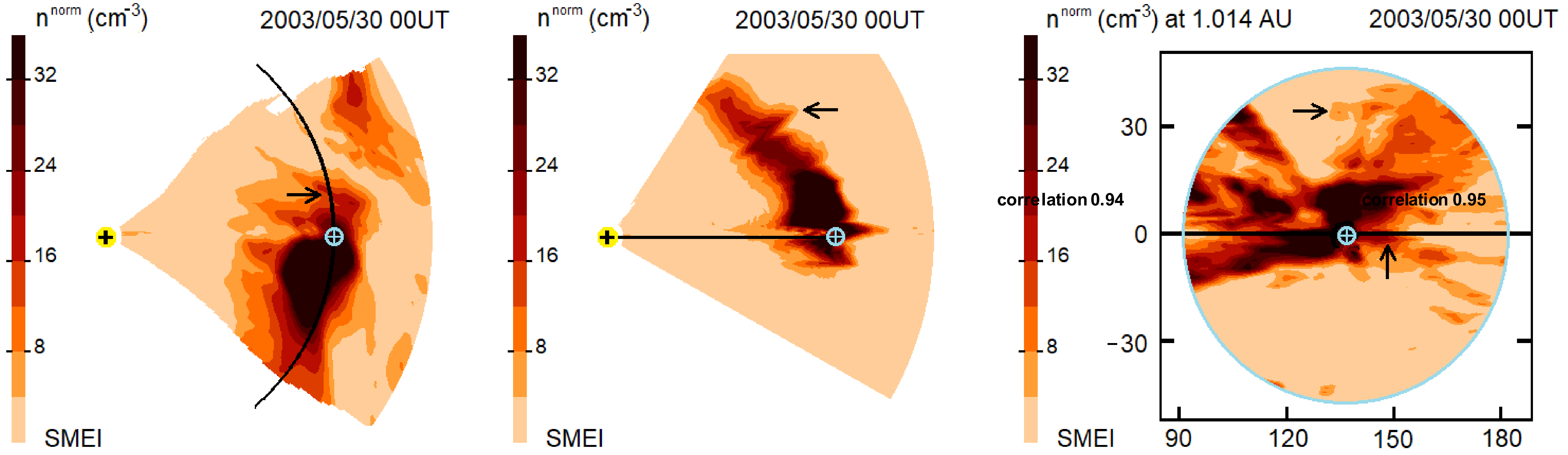
Jackson et al., 2020 doi: 10.3389/fspas.2020.568429

Ecliptic,Earth Meridional, and Synoptic Cuts at 1.5-Hour Cadence Resolution

SMEI Analysis

New SMEI Analysis ~ 1.5 hour cadence

Analyses show CMEs are corrugated and spotty!



Heliospheric Space Weather Predictions and Forecasts

- **Motivation:** to provide best solar space weather predictions and forecasts throughout the global heliosphere
- **Research:** remote heliospheric sensing that provides global models of the heliosphere from Sun to Earth, the inner planets, and outward from there.
- **Projects:** SPWx predictions and forecasts that work using heliospheric data from SMEI (UCSD), IPS (ISEE, Japan), STEREO HI Images (RAL-Space, UK), Worldwide IPS Stations (WIPSS) Network (Includes LOFAR ASTRON, NL).
- **Planned Projects:** All Sky Heliospheric Imager (ASHI - UCSD), the NASA SMEX PUNCH (SWRI), the Vigil HIs (ESA, UK)

http://smei.ucsd.edu https://ips.ucsd.edu/ https://ips.ucsd.edu/stereo https://ashi.ucsd.edu

Combining first-principles and data-derived approaches

Perspective of a global (geospace) modeler

V. G. Merkin (JHUAPL)
and the CGS team

What are new data assimilation/fusion approaches that will likely lead to improved space weather forecasting performance?

The answer to this question depends on:

1) Types and availability of data
2) Types of physics-based models

Challenges

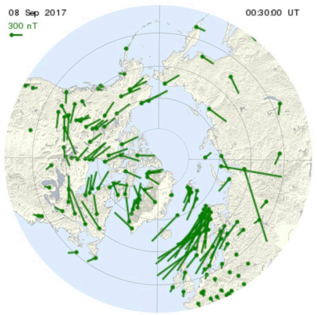
- Very sparse in situ data
- But also very unevenly sampled
  - Much better coverage near Earth (ionosphere, ground)
  - Remote sensing is possible in some regions and for some variables but not for others
- Unique features of the geospace system:
  - Driven system (memory, internal time scales, disparate domains...)
  - Low dissipation (e.g., in the magnetosphere) leads to difficulty in generating physically consistent analysis increments
  - Uncertainty is dominated by model incompleteness (i.e., missing physics)

What are new data assimilation/fusion approaches that will likely lead to improved space weather forecasting performance?

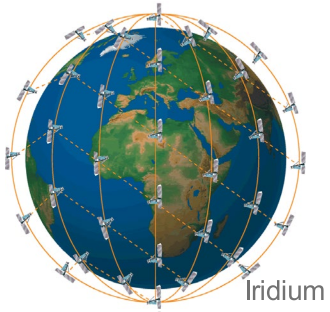
Possible solutions

- Spacecraft constellations
- Leverage better near-Earth coverage
- Leverage historical data
- Use all available data to:
  - Rectify model incompleteness (i.e., supply missing physics)
  - Develop data ingestion/assimilation methods that nudge models by supplying missing physics (i.e., gray-box models)

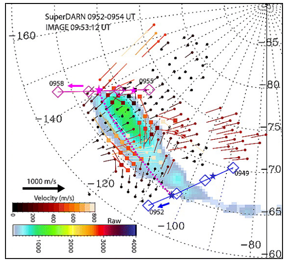
SuperMAG



AMPERE



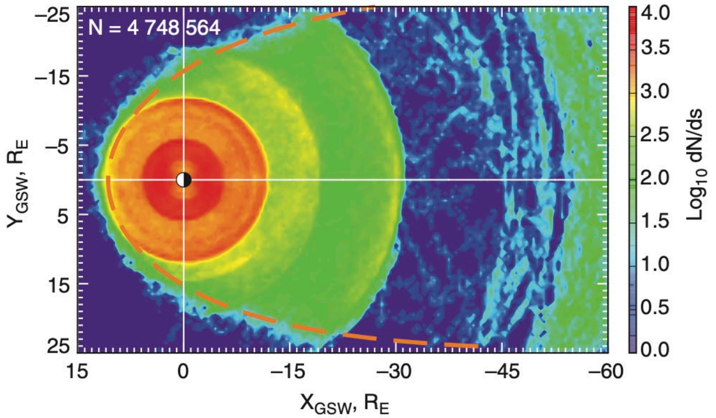
Iridium Communications Inc.



SuperDARN+
DMSP+
IMAGE

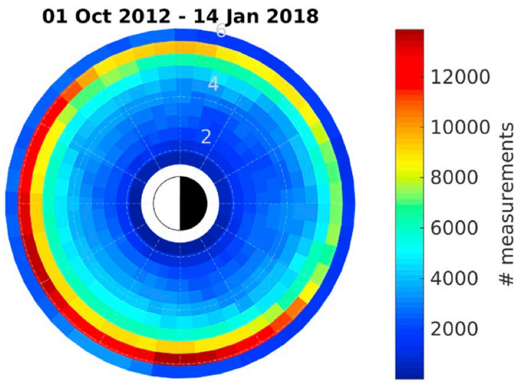
Zou et al. (2009) 10.1029/2008JA013449

Historical magnetometer data (Tsyganenko et al. 2021)



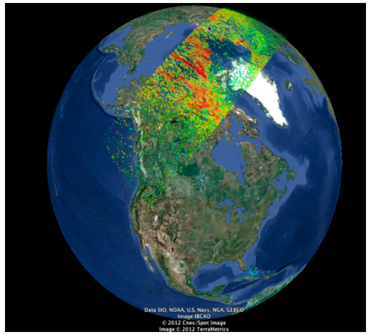
10.1002/9781119815624.ch39

Historical Van Allen Probes data (Wang et al. 2019)



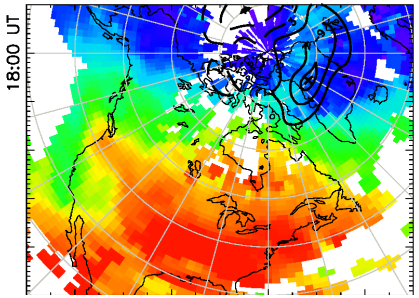
10.1029/2018JA026183

DMSP/SUSSI



sussi.jhuapl.edu

GPS TEC



vt.superdarn.org

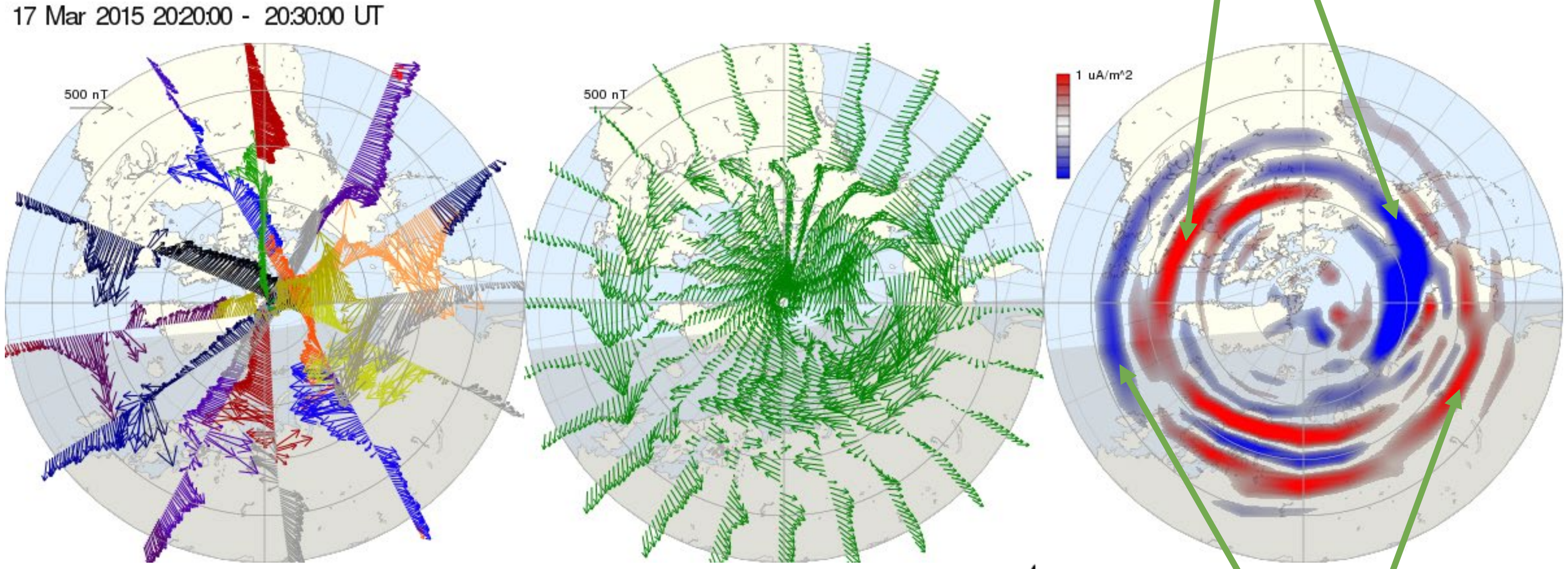
How can we quantify uncertainty in data assimilation schemes that use multi-source observations?

- Develop rich, multi-component cost or evaluation functions.
- Fold together agreement not only with direct in situ measurements but also:
  - composite indices, distributed datasets, remote sensing, and data-mining/empirical reconstructions
- The cost function should reflect data-model consistency over a time window, not a snapshot in time (*shadowing*)
- Explore different component weightings:
  - Require (and quantify) general agreement between the simulation and observations (avoid getting stuck in local minima)
  - Weight reduced-dimensional (carefully selected) global indices and "science metrics" strongly. It will increase the physical relevance of the region and the minimum identified.

Assimilation of low-altitude magnetic field

AMPERE: Measurements by Iridium constellation

R1, directly driven by SW



$$\mathbf{j}_{\parallel} = \frac{4\pi}{c}\nabla \times \delta\mathbf{B}.$$

- Assume quasi-static approximation
- Vasyliunas eq-n:

$$\frac{j_{\parallel}}{B_{\mathrm{i}}} = c\frac{\hat{\mathbf{b}}}{2B_{\mathrm{m}}} \cdot \nabla V \times \nabla p,$$

measured

modeled

R2, driven by inner magnetosphere pressure

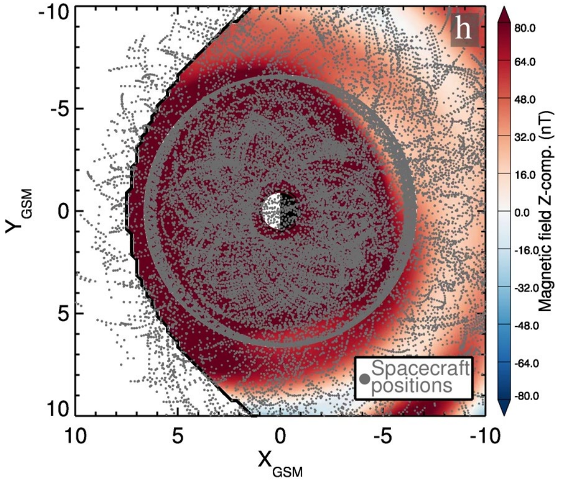
Tweak pressure to optimally match low-altitude mag. perturbation

Merkin, V. G. et a. (2016) 10.1002/2015SW001330

Assimilation of magnetic field measurements in the magnetosphere

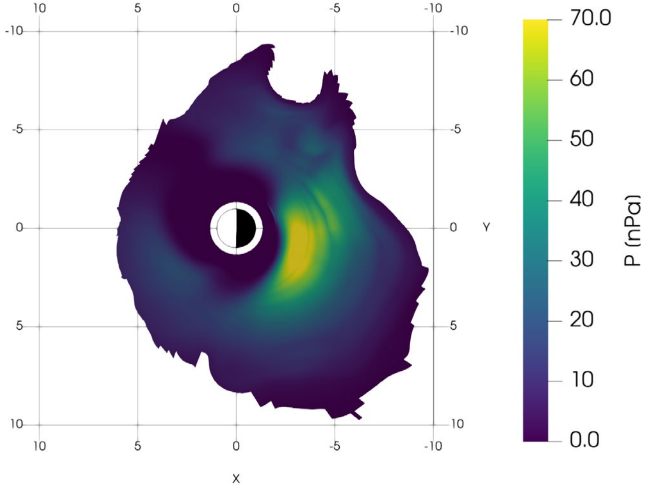
Empirical pressure ingestion*

Mining of historical magnetometer data



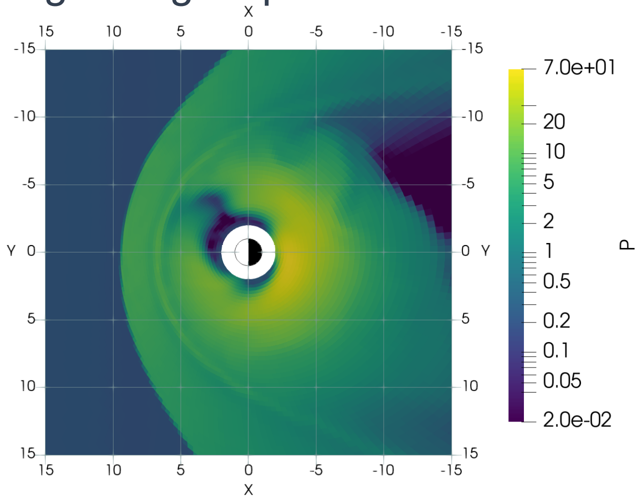
(Sitnov et al., 2020)
10.1029/2020SW002561

Plasma pressure reconstruction



(Stephens et al., 2020)
10.1029/2020SW002561

Pressure ingestion in global geospace model



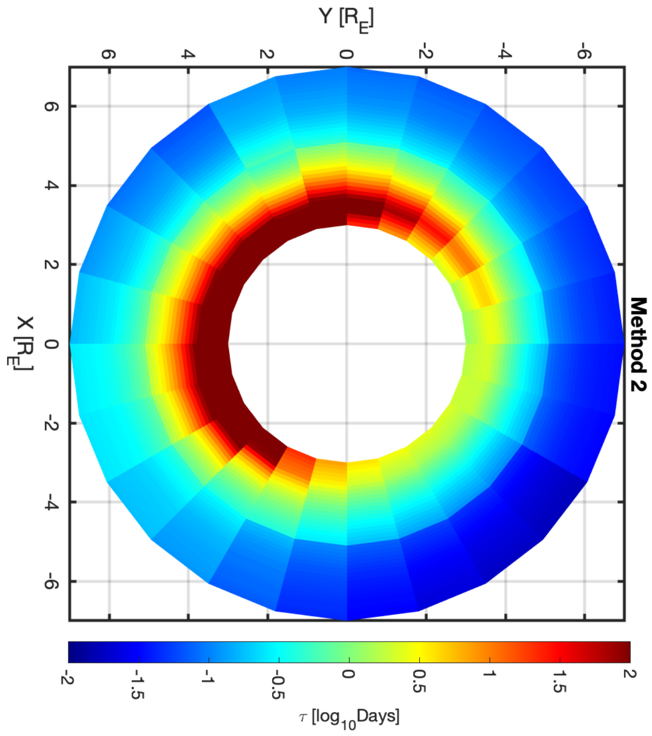
(Sciola et al., ML-HELIO, 2022)

* Similar approach for finding tail X-lines/ adjusting resistivity (H. Arnold)

Data-derived models of the inner magnetosphere plasma waves

Particle precipitation from data-derived wave/lifetime models*

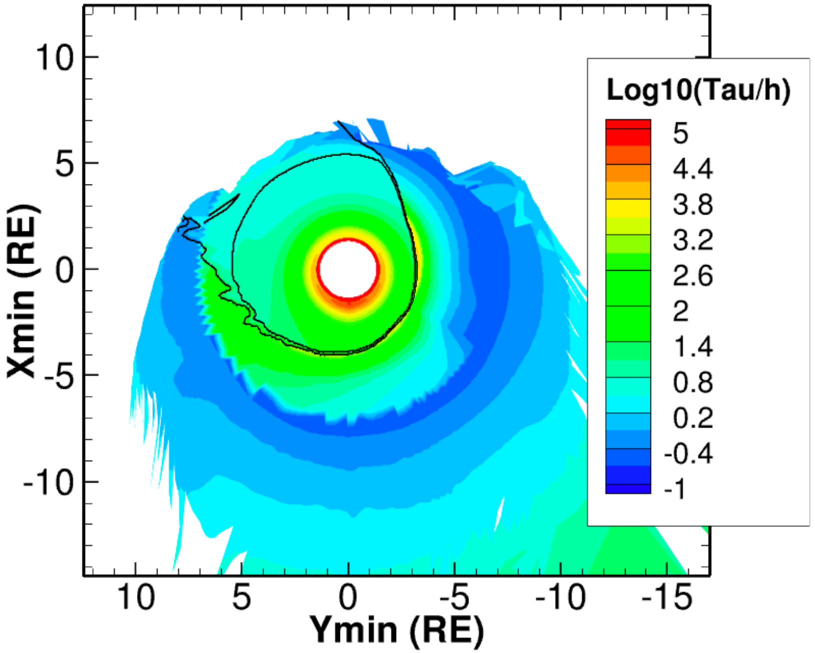
Van Allen Probes historical wave data



(Wang et al., 2019) 10.1029/2018JA026183

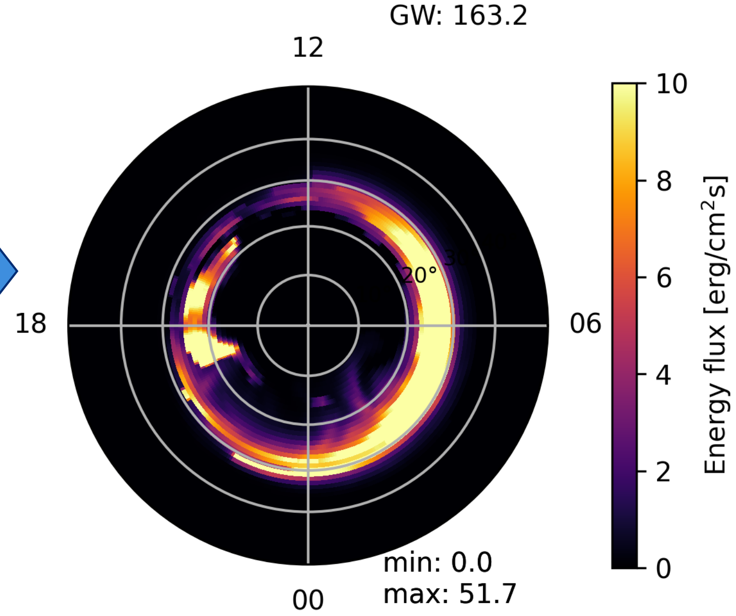
also (Wang et al., 2022) in review

Electron lifetime in inner mag. model



(Bao et al., ML-HELIO, 2022)

Precipitating electron energy flux in global geospace model



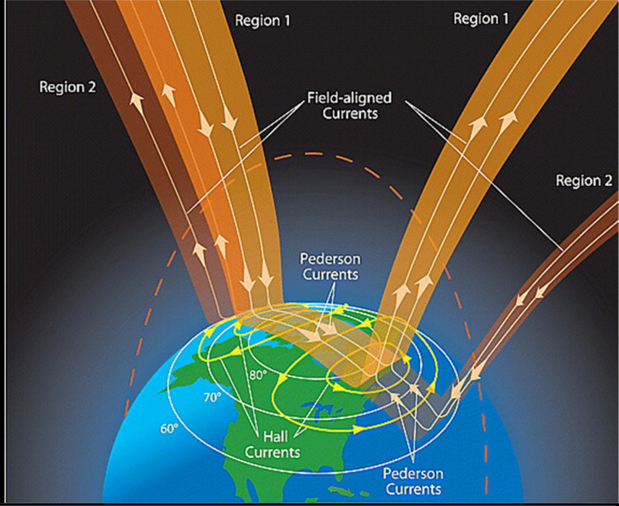
(Bao et al., ML-HELIO, 2022)

* Similar approach for radiation belt losses (A. Michael)

APL

JOHNS HOPKINS

APPLIED PHYSICS LABORATORY



Electromagnetic and kinetic energy represent the largest unknown inputs to the upper atmosphere (up to ~700GW)

These inputs are highly variable
and temporally, so they cannot l
observed from a single vantage

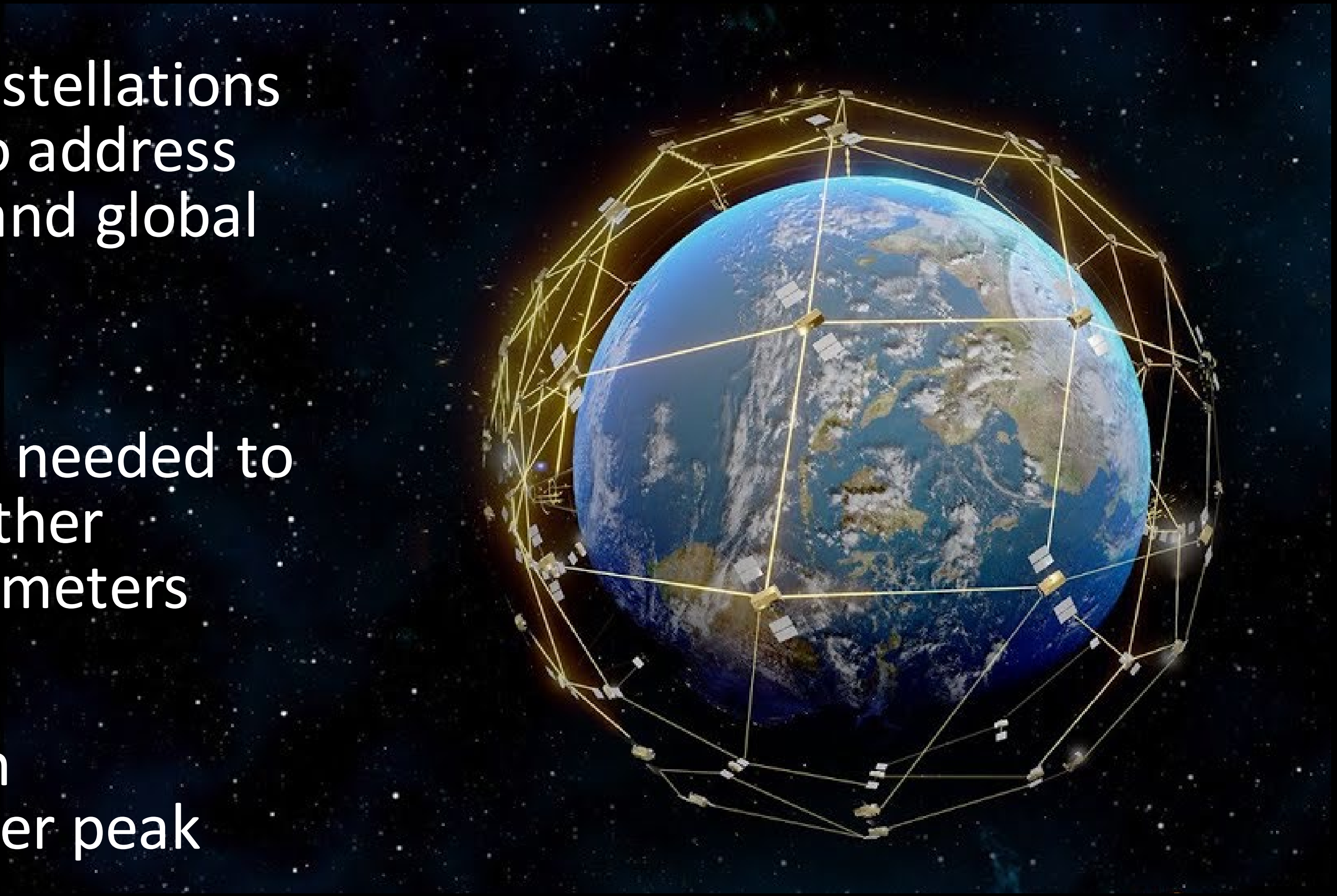
Important space weather plasm
phenomena are driven by these
especially at high latitudes

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Proliferated low-Earth orbit constellations provide the coverage needed to address major unknown energy inputs and global system response

Careful treatment of the data is needed to remove biases, combine with other datasets and infer physical parameters

Diverse datasets needed at high spatial/temporal resolution: layer peak densities, E-field, particles

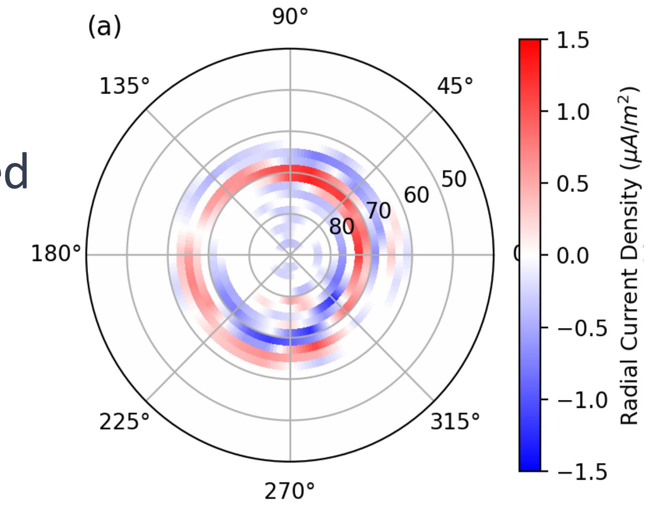


66-sat Iridium constellation provides AMPERE magnetometer data

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

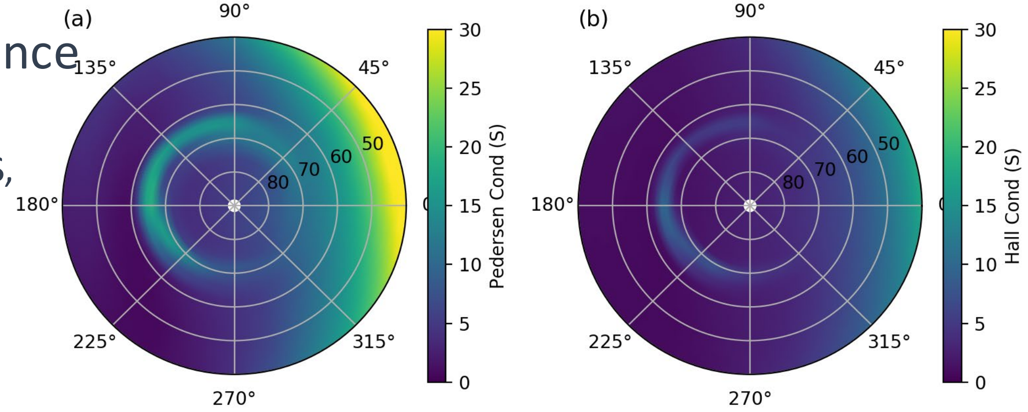
Solving for the high latitude potential

Observed field-aligned currents $\boldsymbol{j}$
(harmonic fit to constellation d**B** measurements)

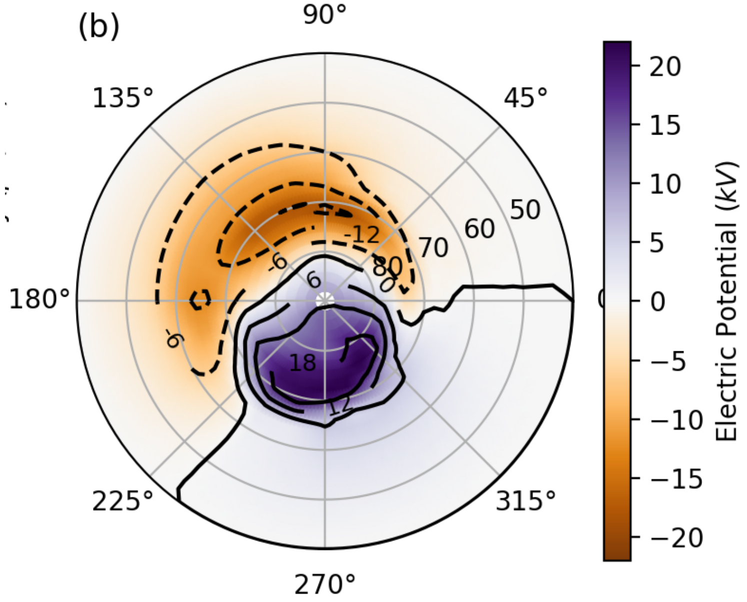


+

Modeled conductance
Σ_P Σ_H
(SAMI3, Hardy, FISM/GOES, HWM/MSIS)



MIX high-latitude potential **Ψ**
(see Merkin and Lyon, 2010)



The Data Fusion and Assimilation Panel

(National Academy of Science, Engineering and Medicine, Space Weather II Workshop)

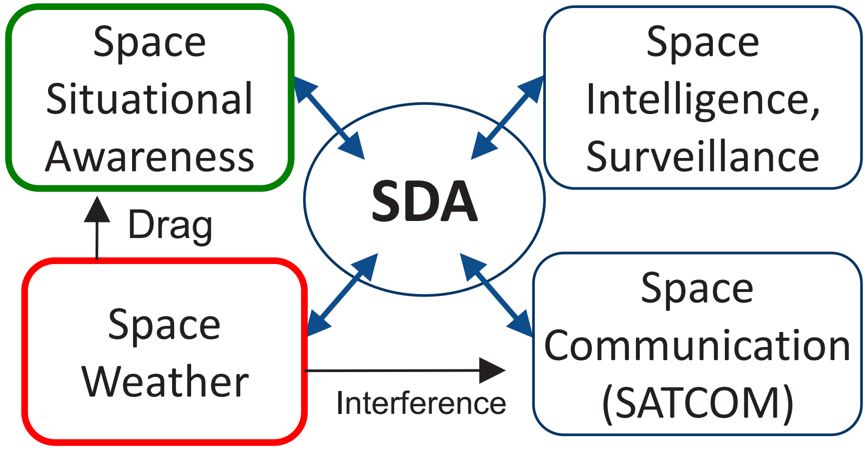
Dr. Erik Blasch

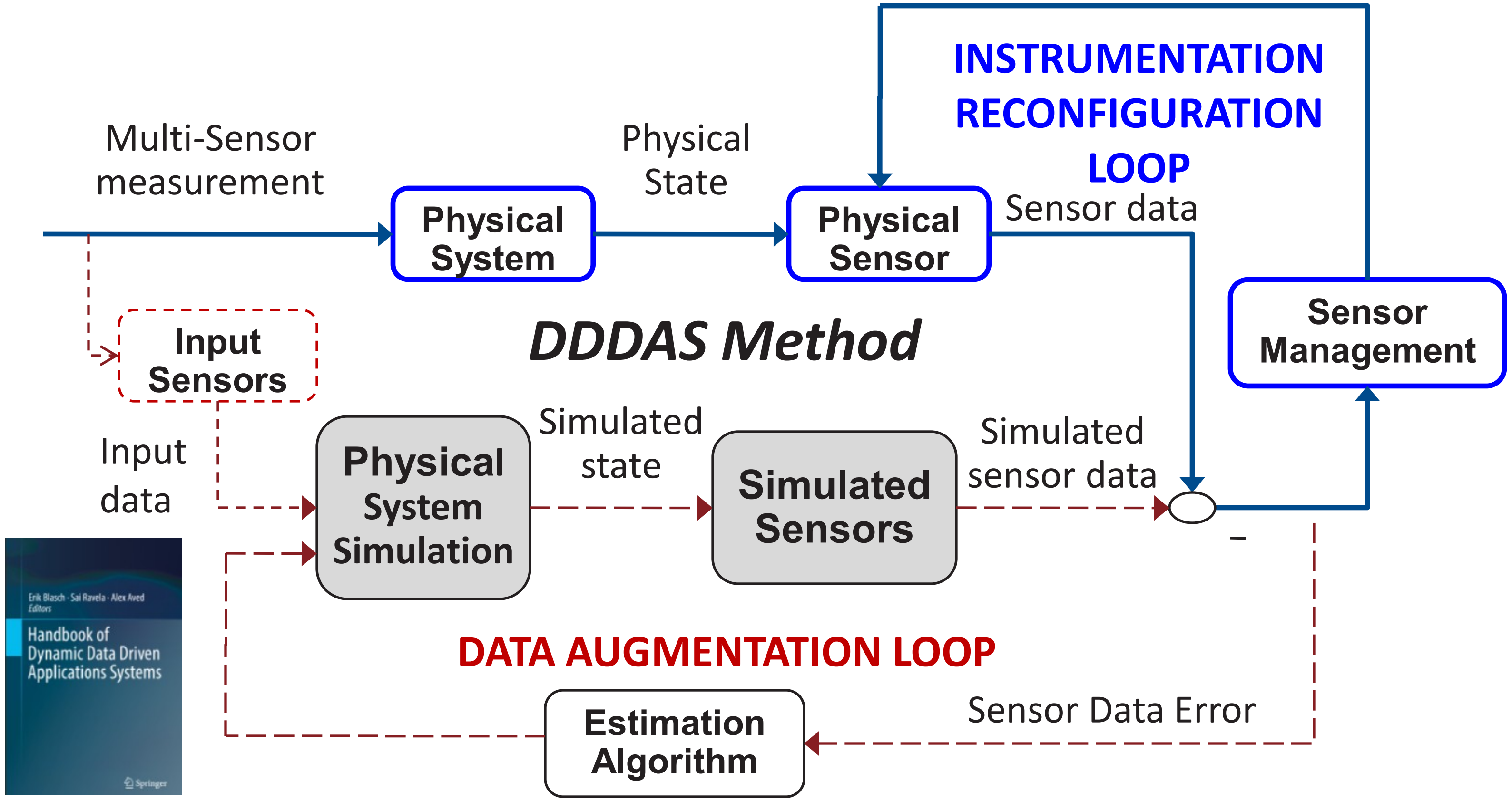
Program Officer, Air Force Office of Scientific Research

13 April 2022

Challenge – Space Domain Awareness (SDA)

- Approaches (for SSA)
  - **Physics-Based and Human-Derived Information Fusion (PHIF)**
  - **Context-Enhanced Information Fusion**
  - **Dynamic Data Driven Applications Systems (DDDAS)**





Data Fusion
- Multimodal sensing to reduce error
- Leverage contextual knowledge
- State assessment supports awareness
- Challenge: Non-constant sensor error

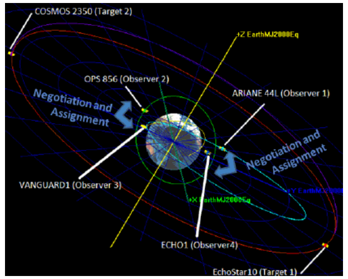
Data Assimilation
- First-principles physics modeling and simulation.. (data augmentation)
- Reduced order modeling (ROM)
- Ensemble Filtering/Learning (EnKF, Machine Learning) … **Deep Learning**

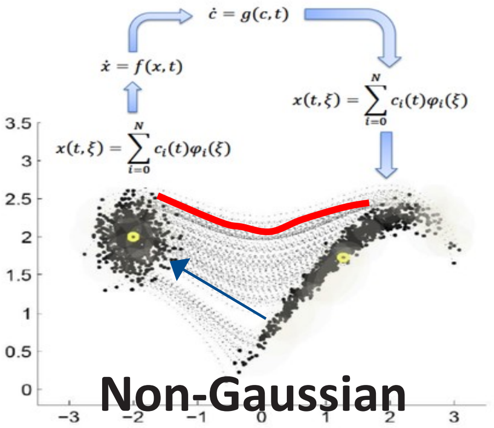
E. Blasch, S. Ravela, A. Aved (eds.), *Handbook of Dynamic Data Driven Applications Systems*, Vol 1, 2nd Ed, Springer, 2021.

Assertion: Non-Linear, Non-Gaussian Evidential Reasoning

Evidence: Temporal Decision Analysis

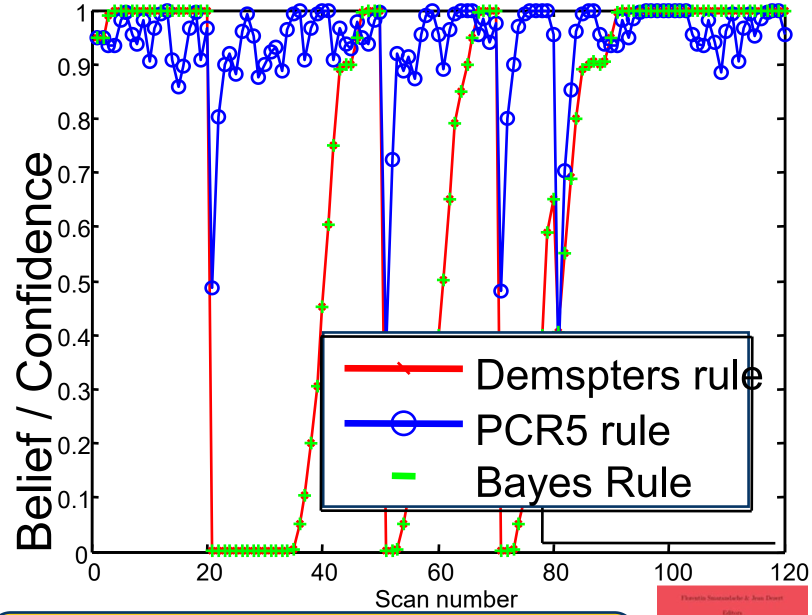
Blasch 2000-



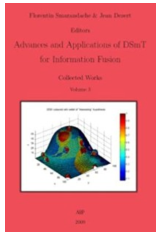


Data Fusion: filtering, estimation, and prediction

- **Program**: Multi-domain correlation and fusion (association)
- **Challenge** Associate data from two sensors (ground & space)
  Conflict between sensor readings (true or not)
- **Requires**: non-linear, non-Gaussian approaches
- **Leverage**: advances in distributed edge processing
- **Utilize**: Evidential (non-Bayesian) reasoning (**Evidential NN)
  Example: Proportional Conflict Redistribution (PCR)



Bayes can't change beliefs quickly with sensor conflicts

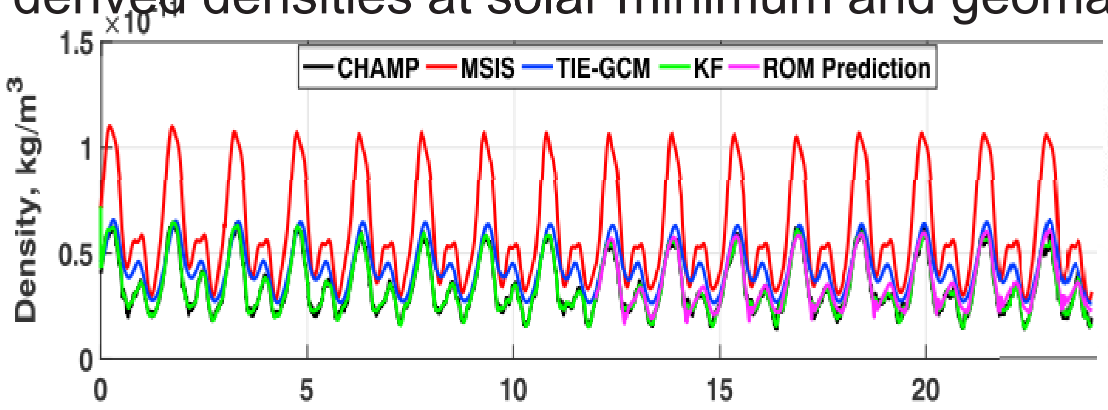


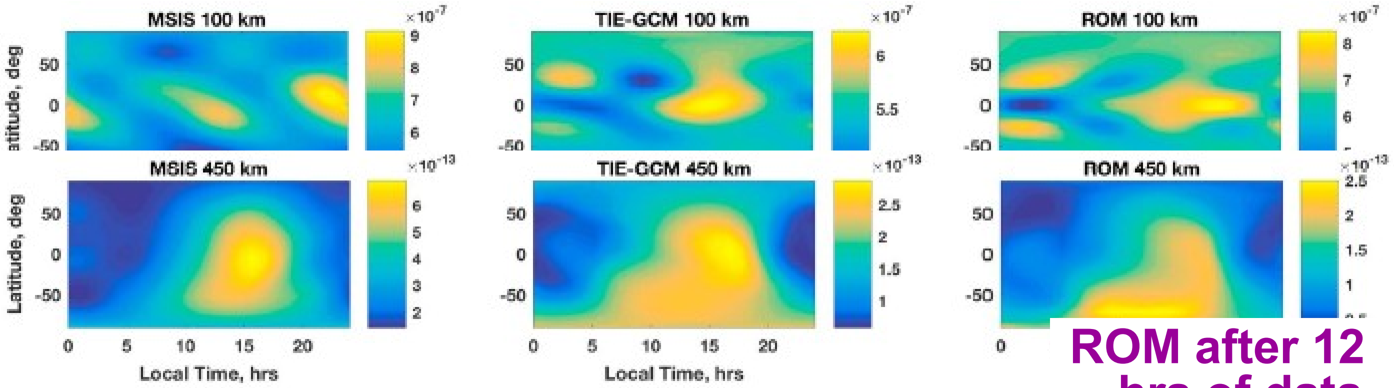
PM Mehta, **R Linares,** A new transformative framework for data assimilation and calibration of physical ionosphere-thermosphere models, Space Weather, 2018 (Data Fusion, Data Assimilation)

DA Marsillach, **MJ Holzinger**, Telescope Tasking for Maneuver Detection and Custody Maintenance using **Evidential** Reasoning and Reachability Theory, - 2020 - mostech.com

(#1) What new data assimilation/fusion approaches improving space weather forecasting performance?

- **Thermosphere-Ionosphere-Electrodynamics General Circulation Model (TIE-GCM) ROM (Reduced-order model)** - restricted to altitudes between 100 and 450 km
- **Forecasts, making drag the largest source of uncertainty** in our ability to accurately predict the state of the objects in LEO.
- Proper Orthogonal Decomposition (POD) or Empirical Orthogonal Functions (EOFs)) with dynamic systems (EKF) for simulation for prediction … assumes Bayesian and EKF
- Because existing **empirical and physical models** have the largest bias/difference with accelerometer-derived densities at solar minimum and geomagnetically active conditions, demo is 2009 instance





ROM after 12 hrs of data

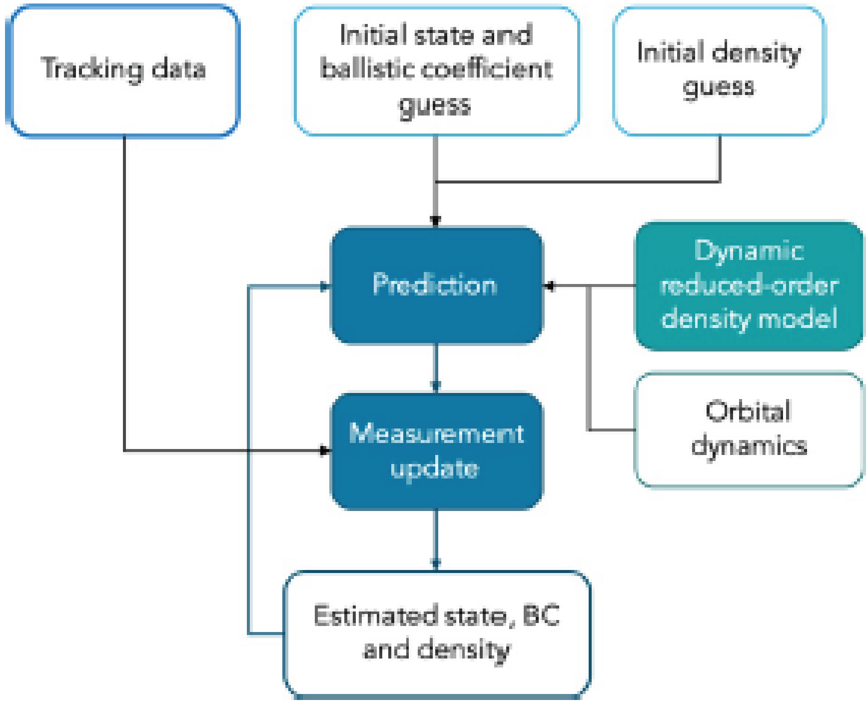
- Black: CHAMP accelerometer-derived density estimates. Red: MSIS model output along CHAMP orbit.
- Blue: TIE-GCM model output along CHAMP orbit. Green: CHAMP assimilated ROM densities on day 320 for year 2009. **Magenta**: prediction with ROM after 12 hrs of data assimilation (Error Reduction 50%)

PM Mehta, **R Linares**, A new transformative framework for data assimilation and calibration of physical ionosphere-thermosphere models, - **Space Weather**, 2018

(#1) What new data assimilation/fusion approaches improving space weather forecasting performance?

- **Data Fusion w/ NL/NG Data Assimilation**

Simultaneously estimate the orbits and global density with Unscented Kalman Filter (UKF).

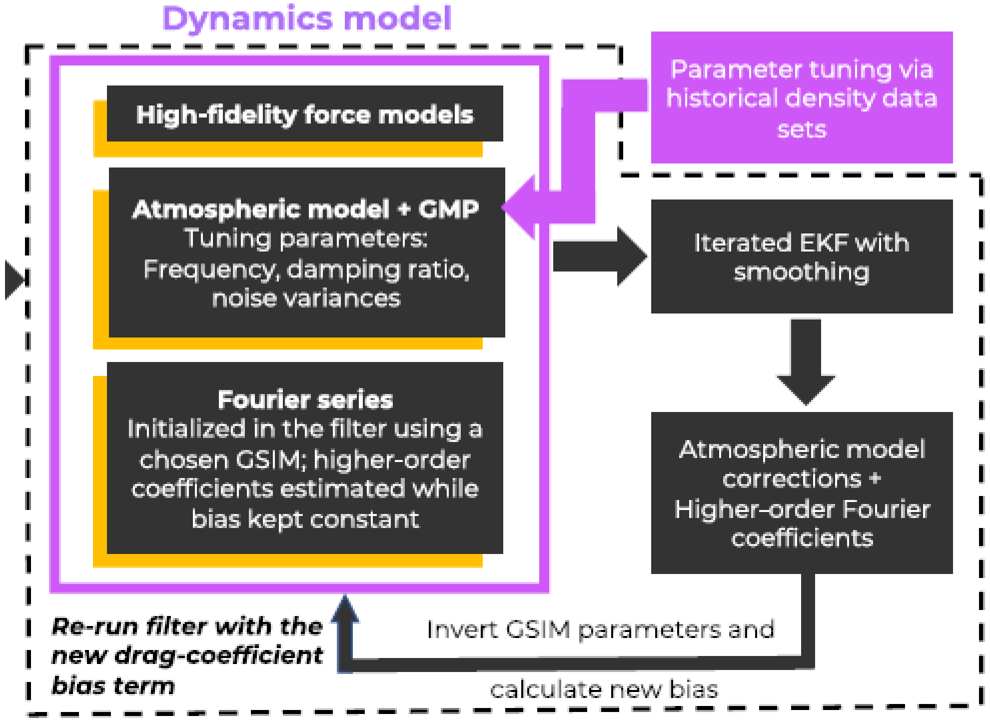


DJ Gondelach, R **Linares** Real-Time Thermospheric Density Estimation via Radar and GPS Tracking Data Assimilation - **Space Weather**, 2021

UKF is better for NL-NG systems

- **Data Fusion w/ NL/NG Data Assimilation**

Simultaneously estimate the density and drag-coefficient for satellites with a time-varying attitude.



V Ray, DJ Scheeres, S Alnaqbi, WK Tobiska, **S. Hesar**, A Framework to Estimate Local Atmospheric Densities With Reduced Drag-Coefficient Biases,... - Space Weather **2022**

49% over JB2008 compared to the High Accuracy Satellite Drag Model densities

(#3) How can we quantify uncertainty in data assimilation schemes that use multi-source observations?

Uncertainty Quantification (UQ)

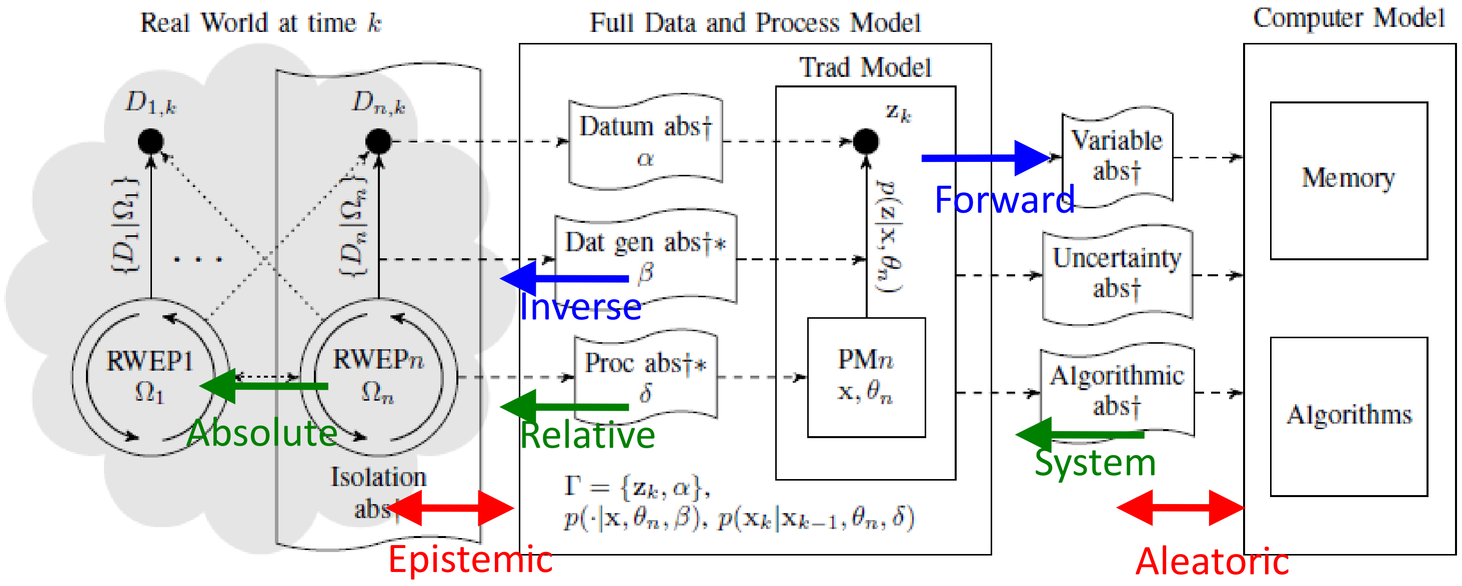
Assertion: Not that easy as uncertainty is everywhere

Evidence: Absolute Uncertainty (RSME), Relative Uncertainly (Processing)

Data Fusion (Reduce Uncertainty), Assimilation (EnKF) filtering

Evaluation of Techniques for Uncertainty Representation Working Group (**ETURWG**), https://eturwg.c4i.gmu.edu/

Uncertainty Representation and Reasoning Evaluation Framework Ontology (URREF ontology), 50+ Sources of Uncertainty



J. P. de Villiers, K. Laskey, A.-L. Jousselme, E. Blasch, A. de Waal, G. Pavlink, P. Costa, "Uncertainty representation, quantification and evaluation for data and information fusion," *International Conf. on Information Fusion*, 2015.

Absolute Uncertainty – real world performance (systems analysis), **end result**

Relative Uncertainty – algorithm performance bound (process), **conditional** f(unc. representation)

Forward uncertainty –**propagation** of uncertainty in model parameters / variables

Inverse uncertainty - **generalization** of parameter estimation error analysis

Epistemic uncertainty - owing to a lack of knowledge or ignorance about the modeled process (**outside** of process)

Aleatoric uncertainty -random events ***within** the entity* or process being modeled.

Same as Deep Learning: Explainability, Interpretability

real world entities and processes (RWEPs)