Data Science and Analytics: Keynote Presentations

Key Questions:

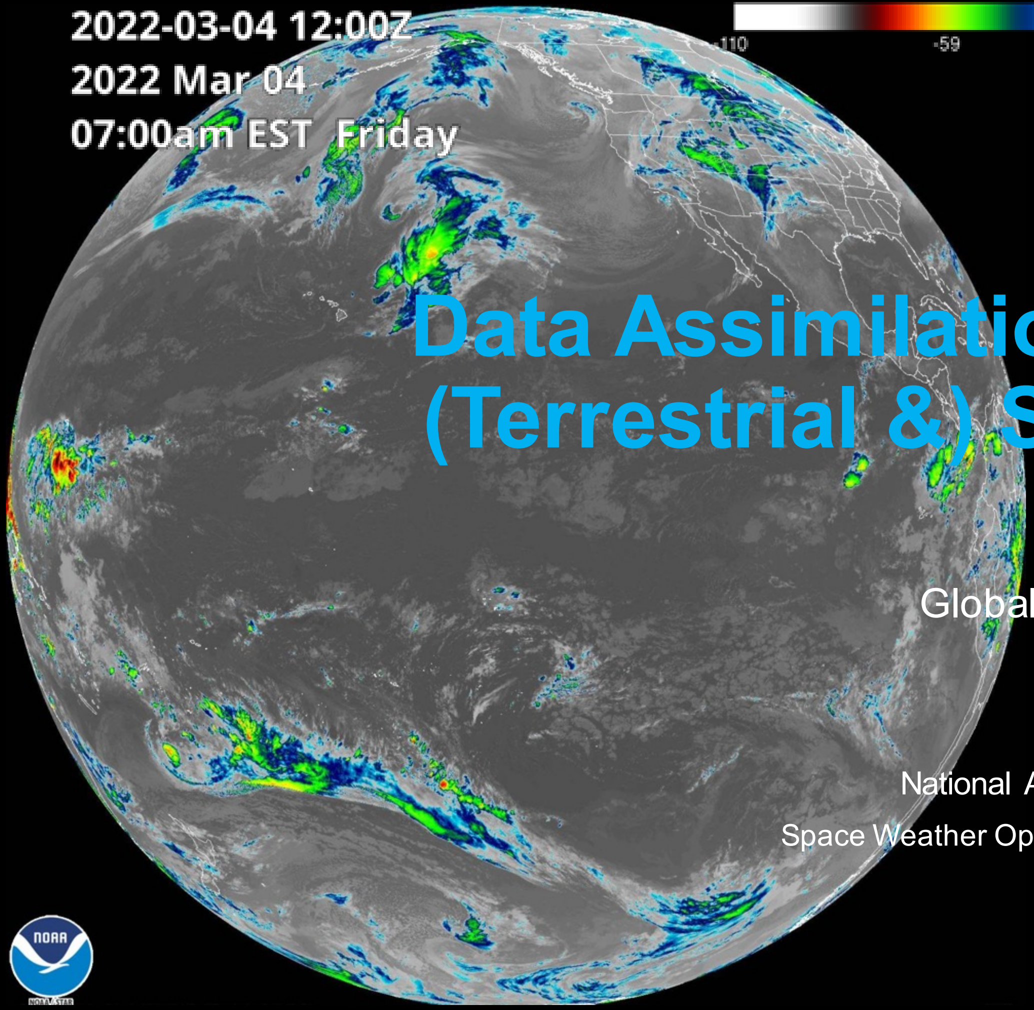
1. Compared to 'Earth System', Space Science is usually data starved. How do we leverage data assimilation and machine learning to overcome this?

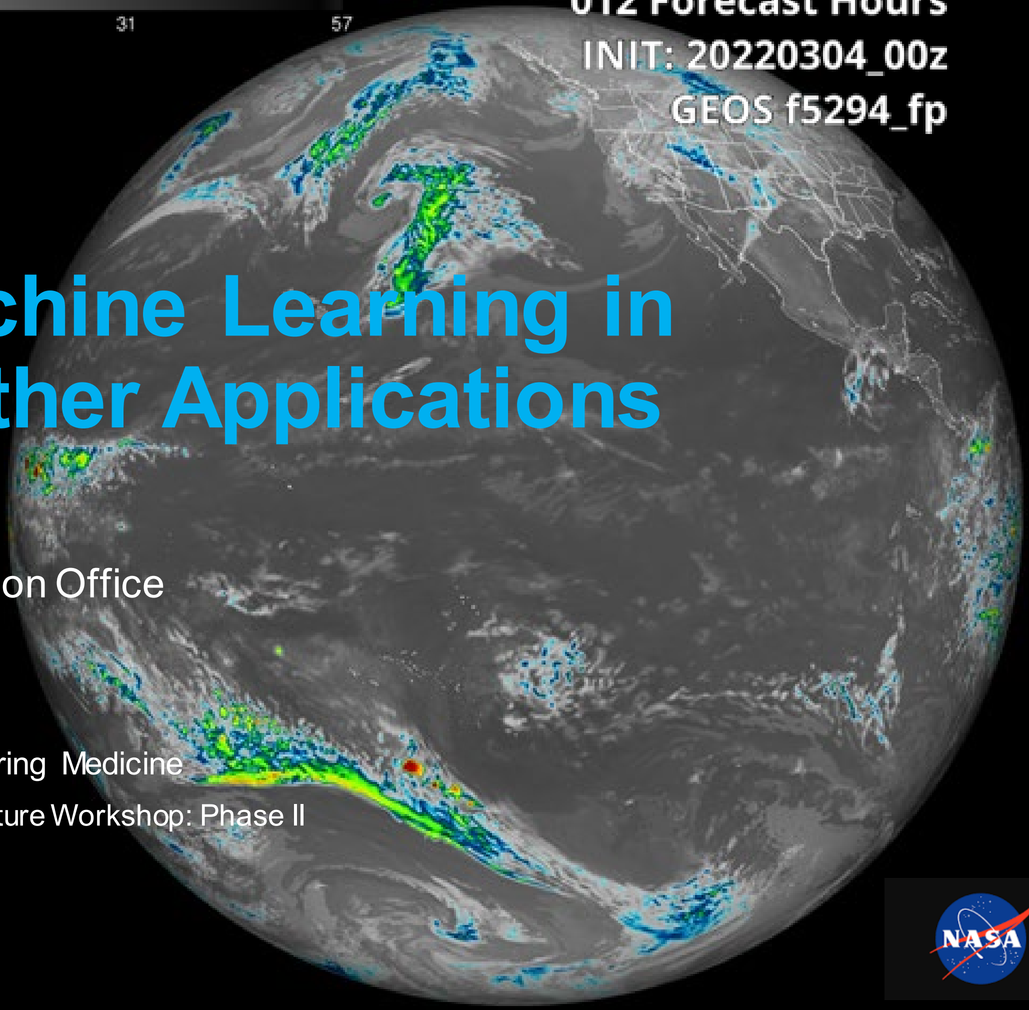
2. What is the role of testbeds in paving the way of data assimilation and machine learning from research to operations?

Moderator: Delores Knipp, Committee

Ricardo Todling,	*GSFC/NASA*	*Research Meteorologist, Goddard Space Flight Ctr, NASA*
Enrico Camporeale,	*CIRES/SWPC*	*Research Scientist CIRES, University of Colorado & SWPC*

Space Weather Operations and Research Infrastructure Workshop: Phase II, Wednesday April 13, 2022, 1100 ET





Data Assimilation and Machine Learning in (Terrestrial &) Space Weather Applications

Ricardo Todling
Global Modeling and Assimilation Office
NASA/Goddard

National Academies of Science Engineering Medicine
Space Weather Operations and Research Infrastructure Workshop: Phase II
11-14 April 2022

Disclaimer: This presentation has a strong Terrestrial Weather Applications bias; it might need some UQ to adjust it to Space Weather Applications!

OUTLINE

1. The Progress in Terrestrial Weather Prediction through DA

2. The Hierarchy of Models and DA Strategies in Terrestrial and Space Weather

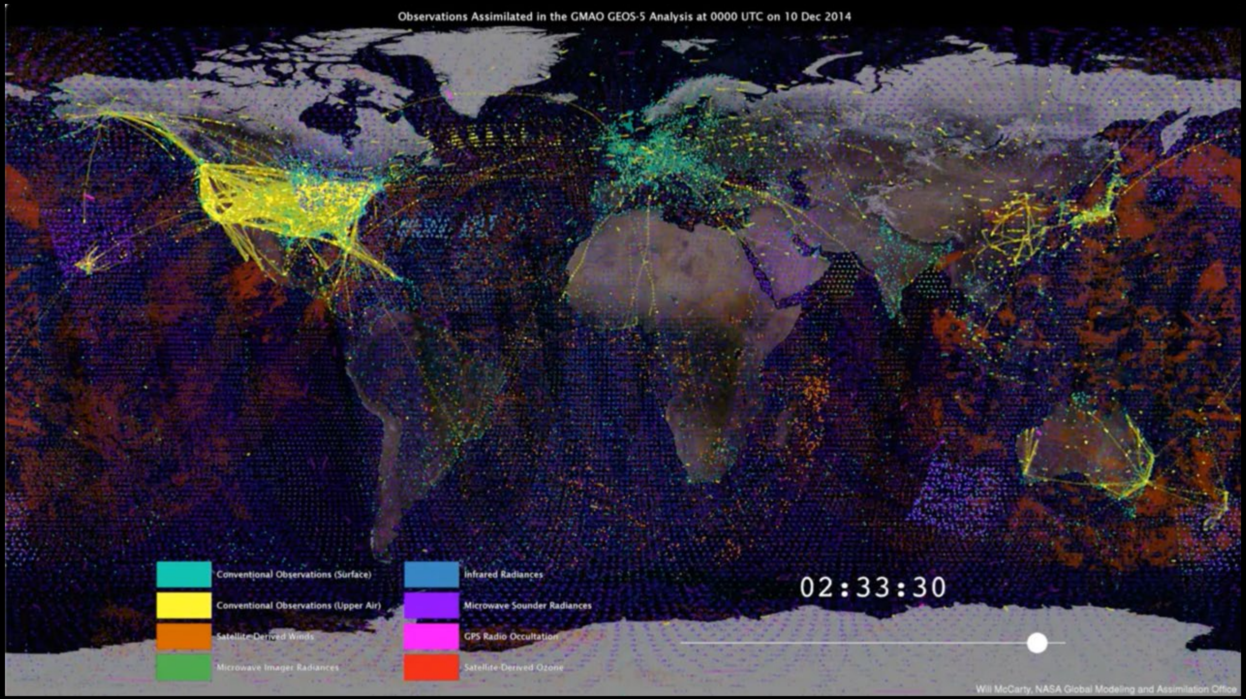
3. Hybrid Concepts

   - The Learning Aspect of DA
   - Machine Learning as Tool to Aid DA

4. A Few Words on Frameworks

5. Closing Remarks

Terrestrial Weather Prediction: Better Data, Models & Techniques

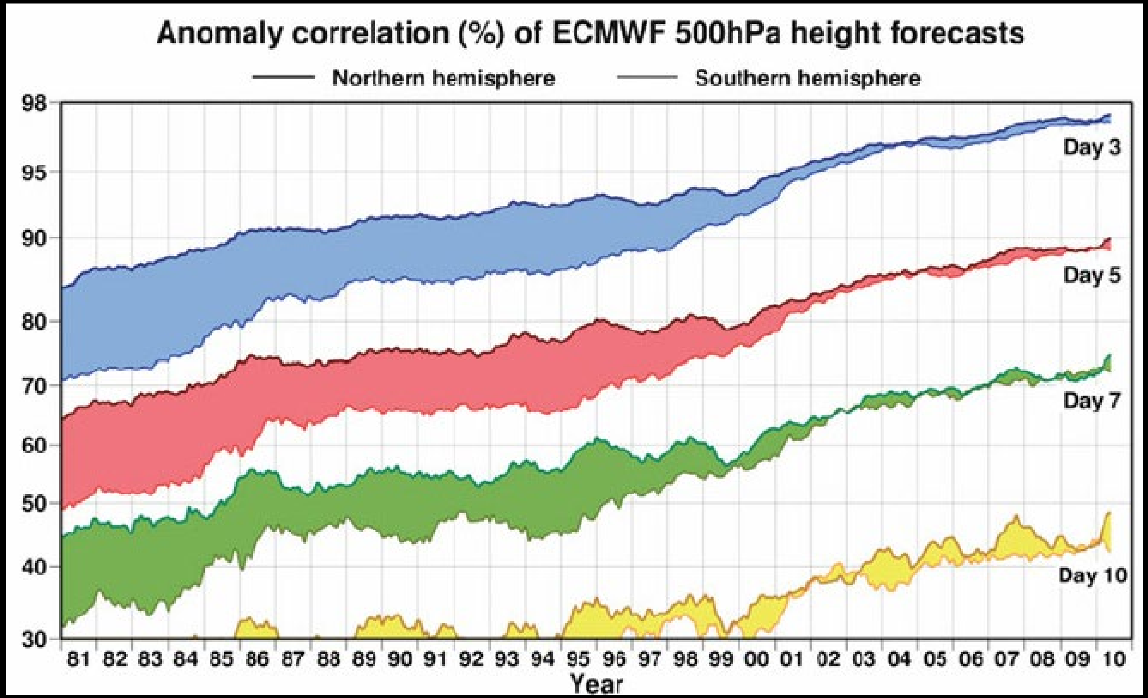


Observation assimilated in GEOS in the 6-hour period between 2100 UTC 9 Dec 2014 and 0300 UTC 10 Dec 2014 (Courtesy of Will McCarty).

OBS/cycle: 5×10^6

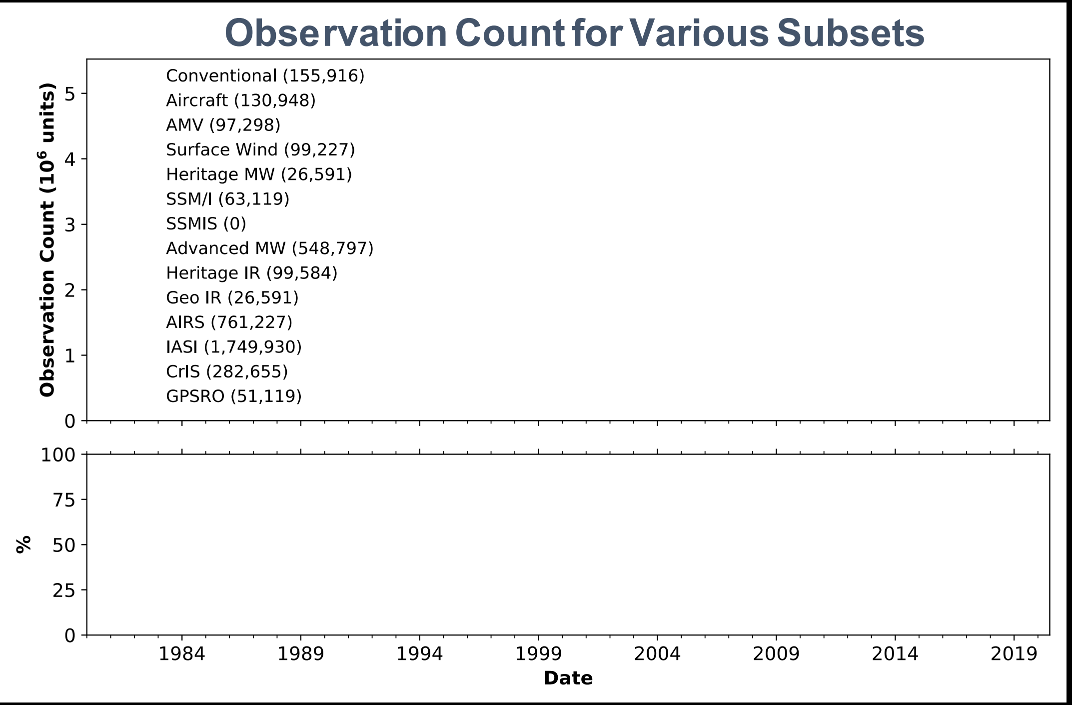
Model: 10^9-10^{10}

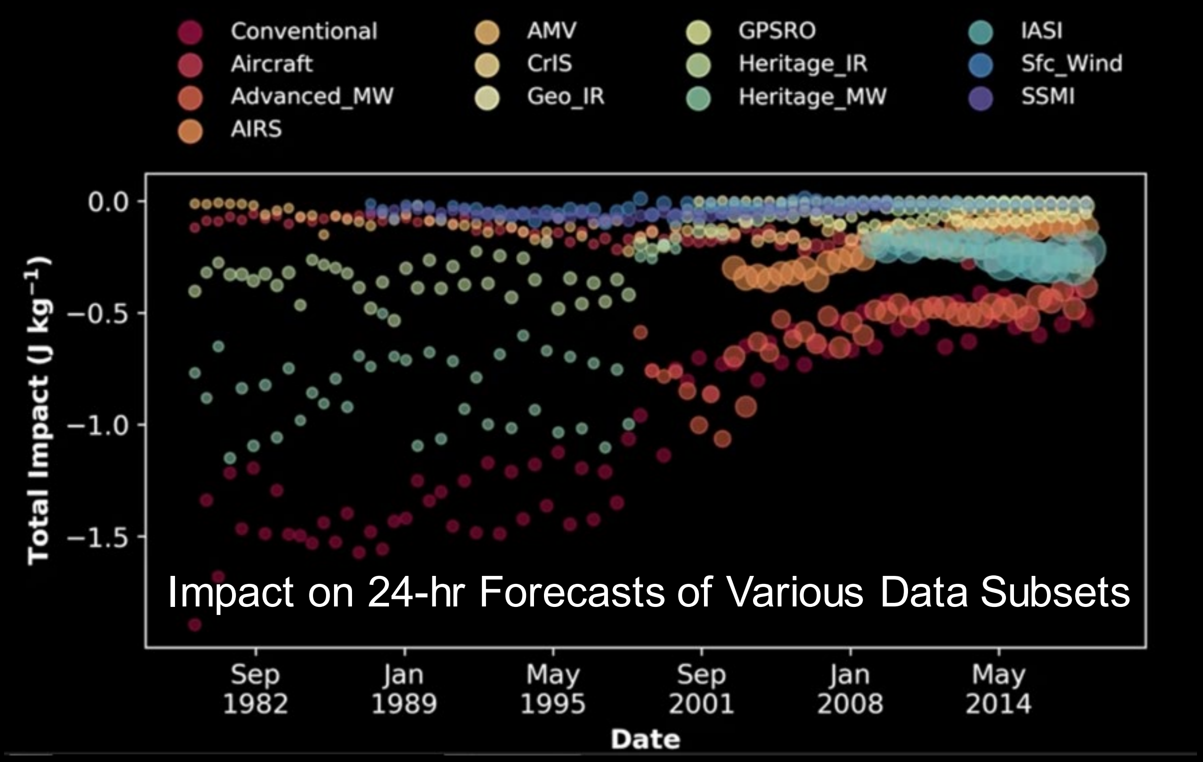
Evolution of ECMWF forecast skill for varying lead times (3 days in blue; 5 days in red ; 7 days in green; 10 days in yellow) as measured by 500-hPa height anomaly correlation. Top line corresponds to the Northern Hemisphere; bottom line corresponds to the Southern hemisphere. Large improvements have been made, including a reduction in the gap in accuracy between the hemispheres (Source: Courtesy of ECMWF. Adapted from Simmons and Hollingsworth (2002).



Terrestrial DA: Impact of 40 Years of Assimilation

Illustration of the increase in data count in MERRA-2 over the past 40-plus years. The impression of a settling data count toward present day is simply a reflection of limitations in M2 to add newly available sensors; a look at the near-real-time, high resolution, GEOS DA system would reveal a continued rise in data count.





Impact of different types of assimilated observations along the course of MERRA-2. The reduced impact in absolute terms is a consequence of the improved quality in the state of the model due to the assimilation of an increased number of high-quality sensors. (size of dots is obs count ([Diniz & Todling 2020](#))

Terrestrial DA-based Predictions: Range of Scales

The accuracy of weather forecasts is a result of increased model resolution, physical processes representation and the large volume of observations assimilated through advanced DA techniques.

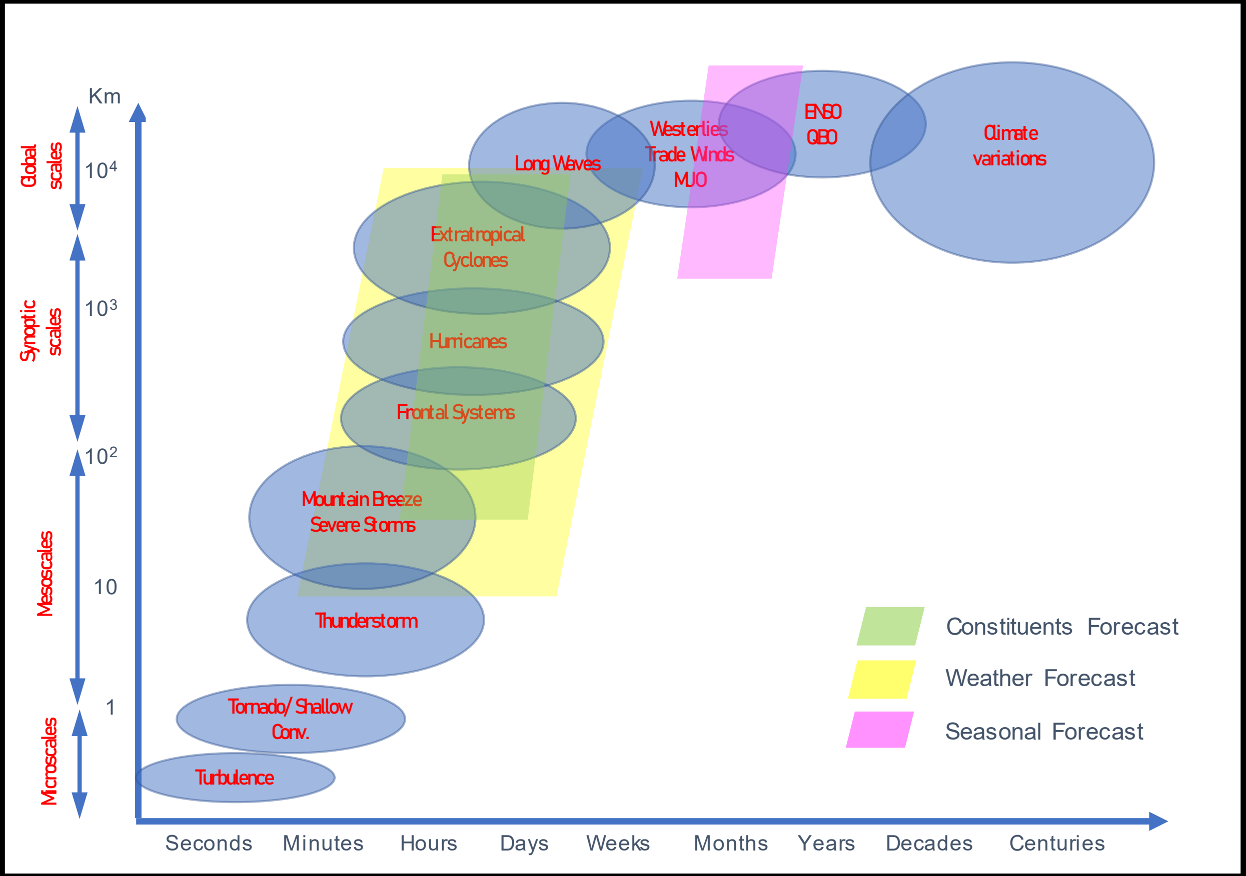
The diagram means to illustrate the range of applicability of DA to Global Terrestrial applications.

Global NWP is now entering the low range of the mesoscales.

Global NWP ranges from hours up to 10 days.

Global Constituent Forecast ranges from hours to 5 days.

Seasonal Prediction extends NWP capabilities in time, with added model complexities, but at the cost of reduced resolution.



Adapted from Tavakolifar et al. (2017; J. Climate)

Terrestrial DA & Prediction: A hierarchy of Components & Strategies
Three Examples from GEOS Forecasting Systems

Different applications invoke different level of model coupling.

Not a one-fits-all approach: Each Forecast System typically includes more than one DA approach.

The Replay strategy roughly nudges one system to results from another.

What's BC today tends to turn into a full modeled component tomorrow.

	Forecasting Systems		
Model Coupled Components	**Weather 12.5 Km**	**Seasonal 50 Km**	**Chemical Composition 50 Km**
Meteorology	Hybrid 4DEnVar	3D Replay	3D Replay
Ozone	Hybrid 4DEnVar	3D Replay	3D Replay
Aerosols	3DVar	3D Replay	3D Replay
Land	None (Soon EnKF)	None	None
Sea-Ice	BC	None	BC
Ocean	BC	3D-EnOI	BC
Chemical Constituents			None (Soon 3D-Var)
Emissions	BC	BC	BC

Included / Prescribed (BC) / Parameterized

Space Weather Prediction: A hierarchy of models

Larger (shorter) range of spatial (temporal) scales

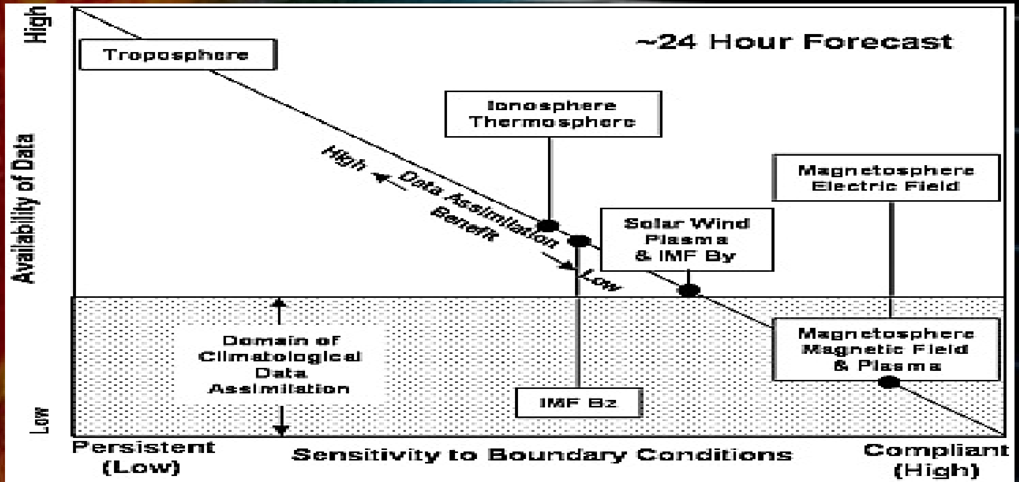
Coronal Model
(e.g., MAS, WSA)

Solar Wind Model
(e.g., Enlil, EUHFORIA)

Magnetosphere Model
(e.g., LFM)

Ionosphere Model
(e.g., Ridley)

Thermosphere/
Atmosphere
(WAM-IPE)



From Siscoe & Solomon (2006)

- Data Sparsity
- Directly vs Instability Driven Dynamics
- Short Timescales
- Range of Forecast Validity
- Intervening Turbulence vs Sensitivity to Initial Conditions

Background pic by: K. Endo

DA Invades Space Weather

Photosphere:
- ETKF, LETKF (Hickmann et al., 2015)
- Scale-Dep EnKF (Hickmann et al. 2016)

Corona:
- EnKF (Butala et al. 2010)

Flares:
- 4D-Var (Bélanger et al. 2007)

CME & Solar Wind:
- LETKF(Lang et al. 2017)
- VarDA (Lang et al. 2018, 2021)

Magnetosphere:
- EnKF (Doxas et al. 2007)
- EnKF (Koller et al. 2007)
- Particle Filter (Nakano et al. 2008)
- OI (Merkin et al. 2016)
- EnKF-based (Godinez et al. 2016)
- SplitOp KF (Cervantes et al. 2020)

Thermosphere-Ionosphere & WAM:
- 3D-Var (Wang et al. 2011)
- EAKF (Morozov et al. 2013)
- EnKF (Chartier et al. 2016)
- EnKF (Cheng et al. 2017)
- ROM-POD-KF (Mehta & Linares 2018)
- EnSRF (Cantrall et al. 2019)
- EAKF (Pedatella et al. 2020)
- EAKF (Hsu et al. 2021)
- 4D-LETKF (Koshin et al. 2022)

Ionosphere:
- SGM-KF & EnKF (Scherliess et al. 2011)
- Nudging (Petry et al. 2014)
- EnKF (Chen et al. 2016)
- LETKF (Durazo et al. 2017)

Earlier SWDA works can be found in Siscoe & Solomon (2006)

Most works above are proof of concept done at coarse resolution and using simplified assumptions; enty other attempts are cited in the reference lists of the works above.

The Data Assimilation Soup

Bayesian View of DA

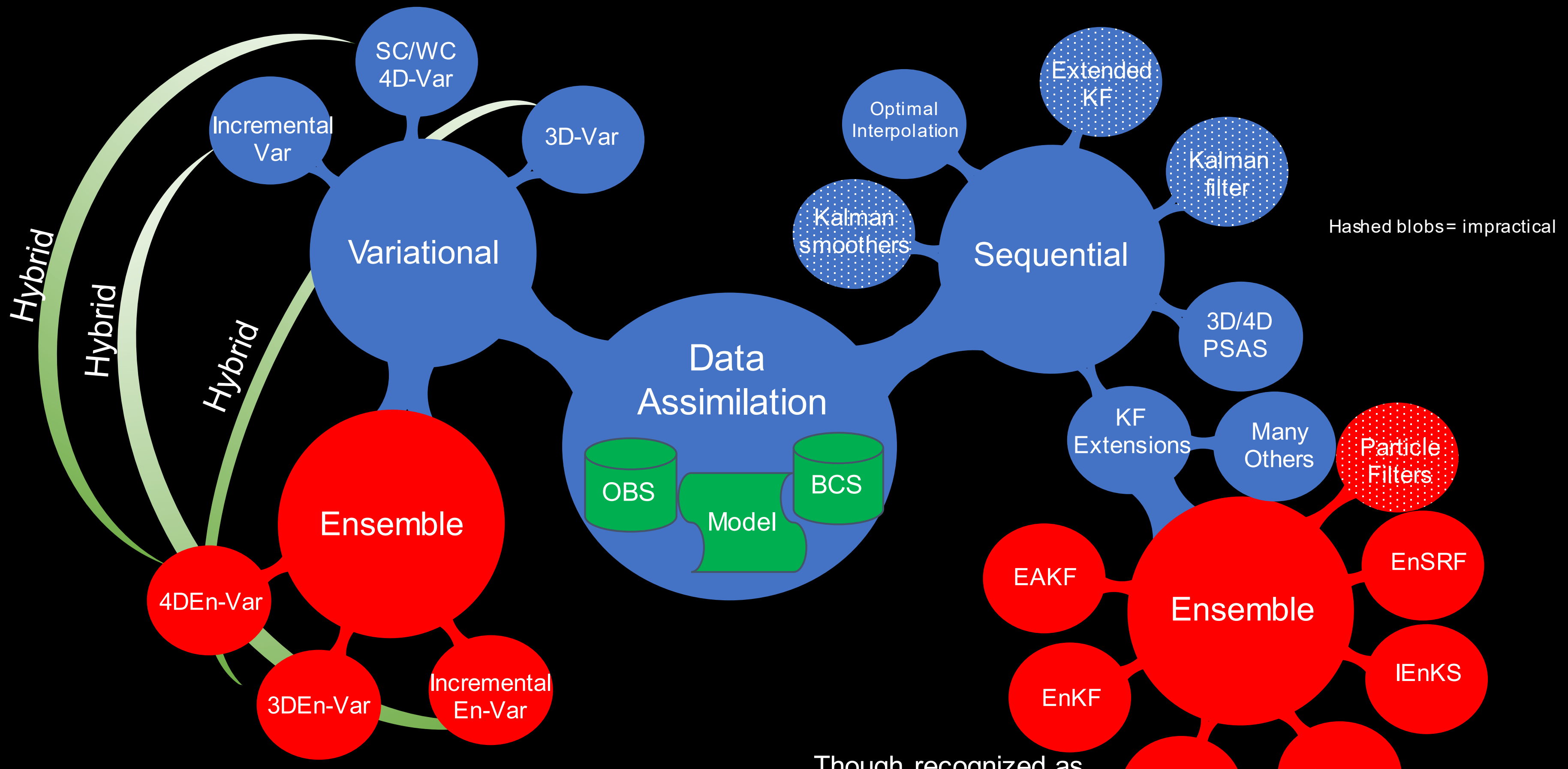
$$p(\mathbf{X}|\mathbf{O}) = \frac{p(\mathbf{O}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{O})}$$

with X and O being a time history of model states and observations over a given time interval.

Provides foundation for both sequential and variational DA frameworks.

Provides insight for hybrid DA.

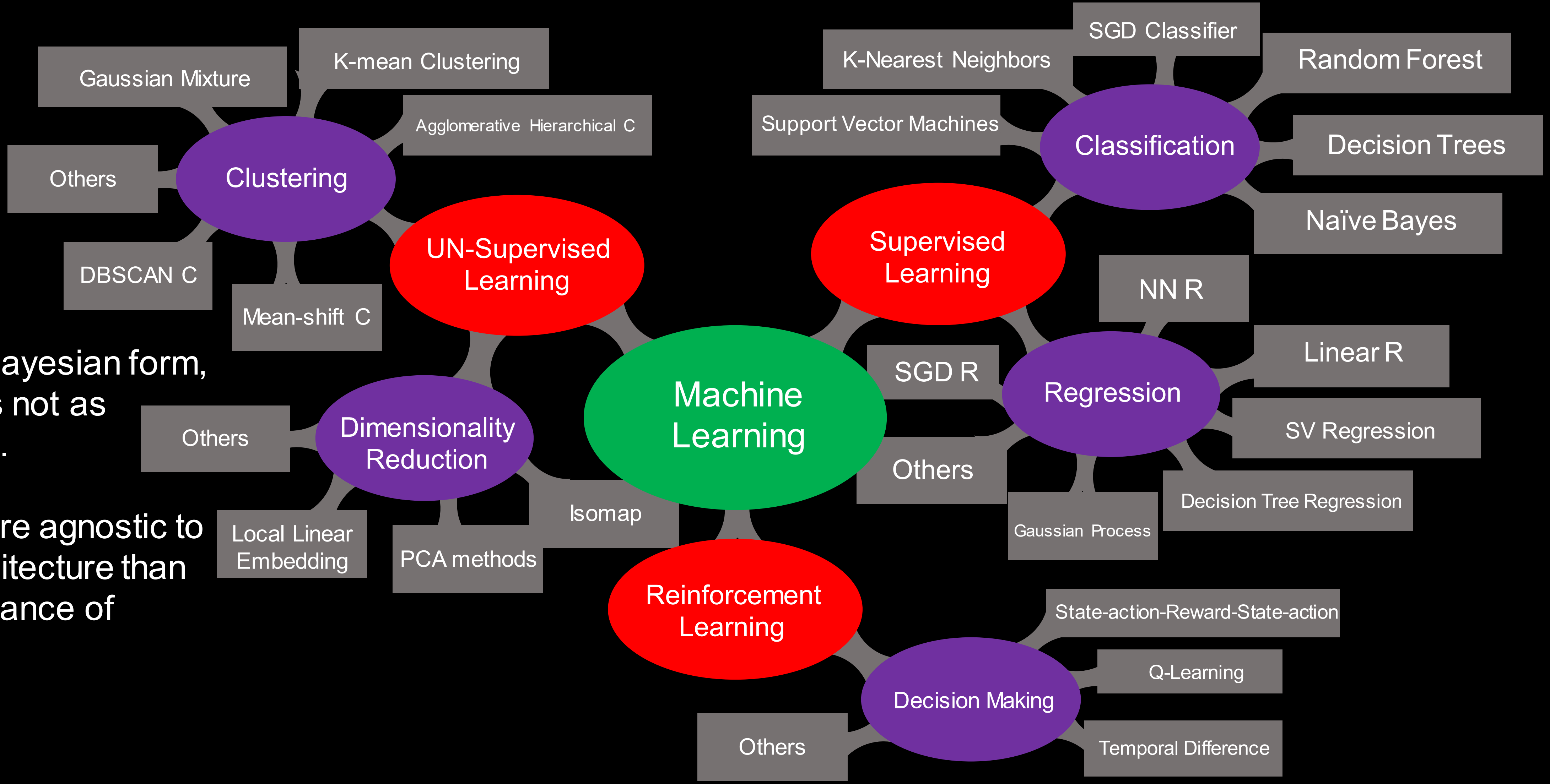
Hybrid DA combines traditional Var (or Seq) with Ensembles.



Inspired by Fig. 1.5 in Asch, Bocquet & Nodet (2016)

Though recognized as needed, there hasn't been much *hybrid* activity on this side of the picture.

The Machine Learning Soup



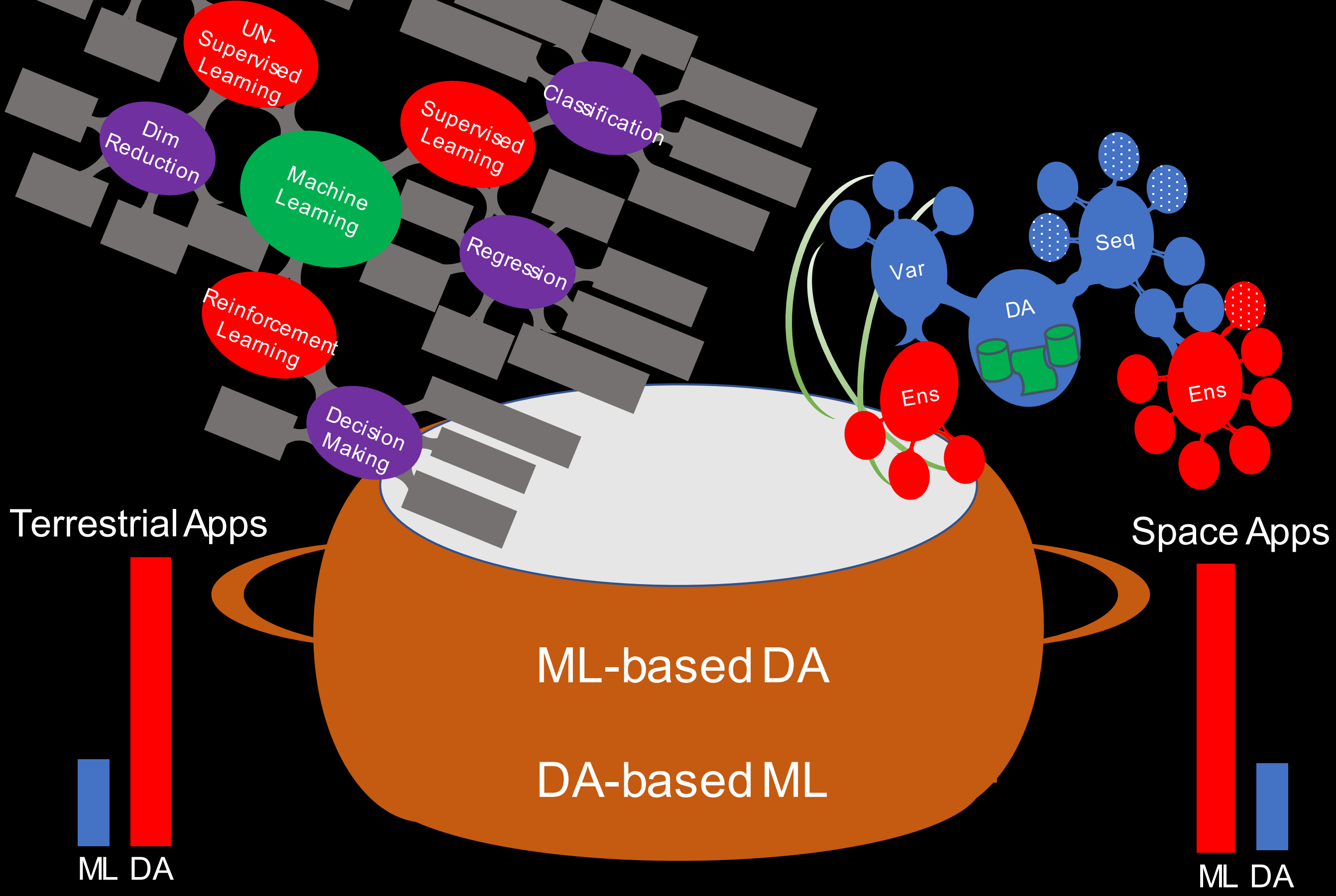
Can be put in Bayesian form, though perhaps not as clear-cut as DA.

Tends to be more agnostic to computing architecture than DA, thus importance of frameworks.

Hybrid Concept: Hybridizing the Hybrid

The past few years has seen a substantial rise in the number of proposals to get ML to aid DA.

In the process, it is also being discovered that DA procedures can in turn aid ML strategies.



See Geer (2021) for broader aspects of this symbiosis.

DA for ML

Handing of space/time sparse incomplete observations; obs operators.

Handling of noisy data.

Inference of processes indirectly related to observations.

Incorporation of prior knowledge, along with Bayesian approach.

Availability of ensembles.

Quantitative representation of uncertainties.

Uncertainty propagation.

Normalization based on physical principles (viz. background errors)

ML for DA*

Process Emulator:

Surrogates

- TL/AD modeling
- Transport/Dynamics
- Chemical integrator
- Obs Error Cov. Construct

Physical Parametrizations

Observations Retrievals

Data Homogenization

Post-processing enhancement

*DA's had its Gray-Box for some time; the Gray-Box of ML (Camporeale) exacerbates it (DA's) further.

The *Learning* Aspect of DA …

Traditional DA is a learning machine …

Adaptive DA is a self-correcting robust machine …

But an inefficient machine in many ways …

The *Learning* Aspect of DA: Terrestrial Applications Adaptive Estimation

DA procedures have incorporated *learning* mechanisms to correct biases and uncertainties for quite a while: ADAPTIVE schemes.

- Variational Bias Correction (VarBC)

From Dee & Uppala (2009)

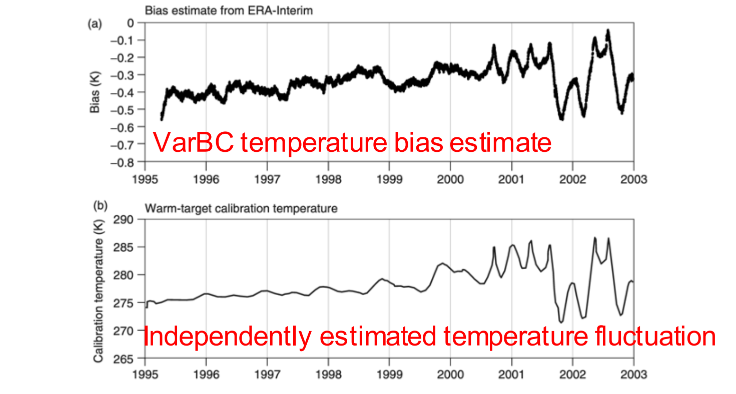
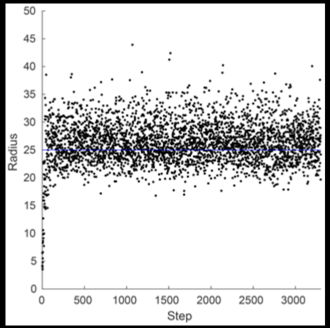


Figure 4. (a) shows global mean bias estimates from ERA-Interim for NOAA-14 MSU channel 2 radiances, as shown in Figure 3. (b) shows recorded variations of the warm-target calibration temperature on board NOAA-14 from Grody *et al.* (2004).

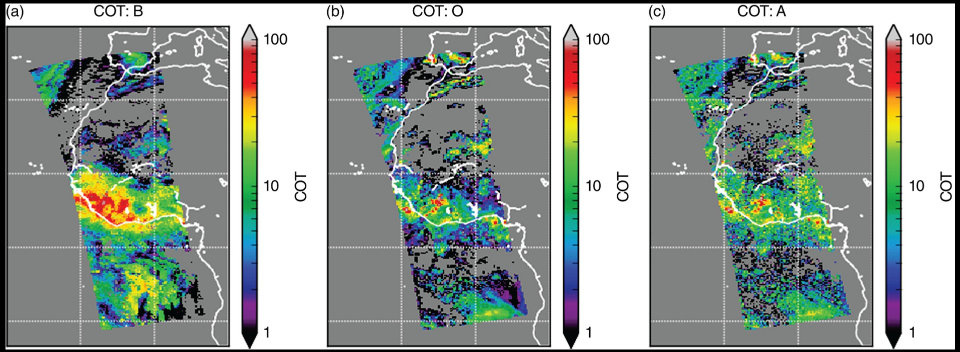
- Adaptive Error Covariance Localization



Left: Localization radii adaptively estimated for a QG model error covariance; Popov & Sandu (2019)

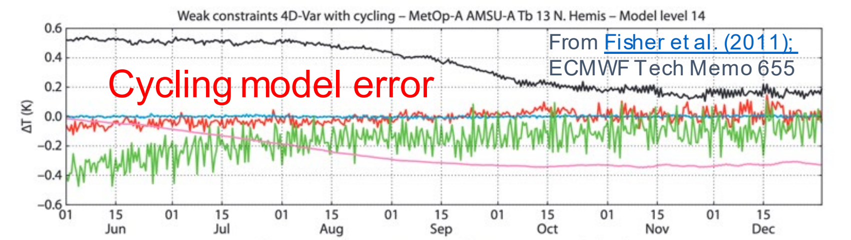
Other works have also explored procedure for adaptive inflation in ensemble DA schemes. Some have made it into SWDA literature, e.g., Godinez & Koller (2012).

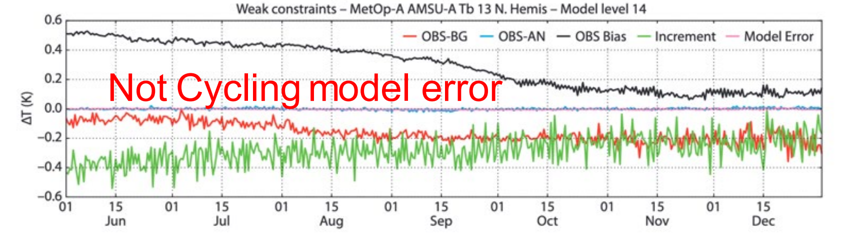
- Model parameter estimation



E.g., Cloud Optical Thickness over MODIS pass: background, observations, analysis; Norris & da Silva (2016; Part II)

- Weak Constraint 4D-Var





The *Learning* Aspect of DA: Terrestrial Applications Adaptive Estimation

DA procedures have incorporated *learning* mechanisms to correct biases and uncertainties for quite a while: ADAPTIVE schemes.

- Variational Bias Correction (VarBC)

From Dee & Uppala (2009)

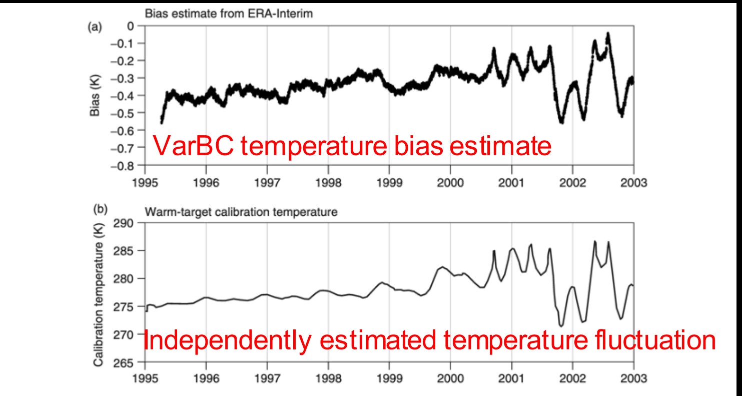
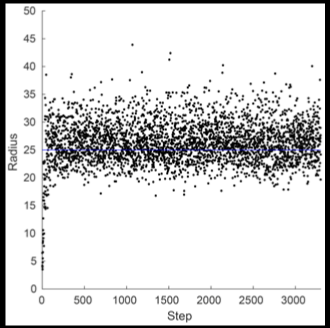


Figure 4. (a) shows global mean bias estimates from ERA-Interim for NOAA-14 MSU channel 2 radiances, as shown in Figure 3. (b) shows recorded variations of the warm-target calibration temperature on board NOAA-14 from Grody *et al.* (2004).

- Adaptive Error Covariance Localization



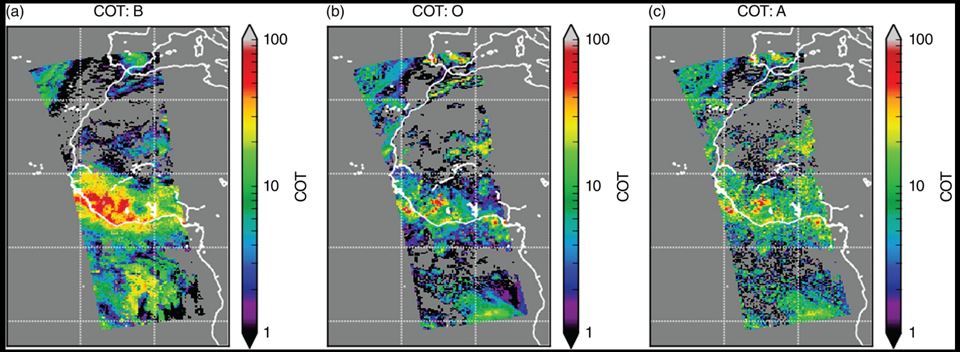
Left: Localization radii adaptively estimated for a QG model error covariance; Popov & Sandu (2019)

Other works have also explored procedure for adaptive inflation in ensemble DA schemes. Some have made it into SWDA literature, e.g., Godinez & Koller (2012).

All these procedure Error Estimates

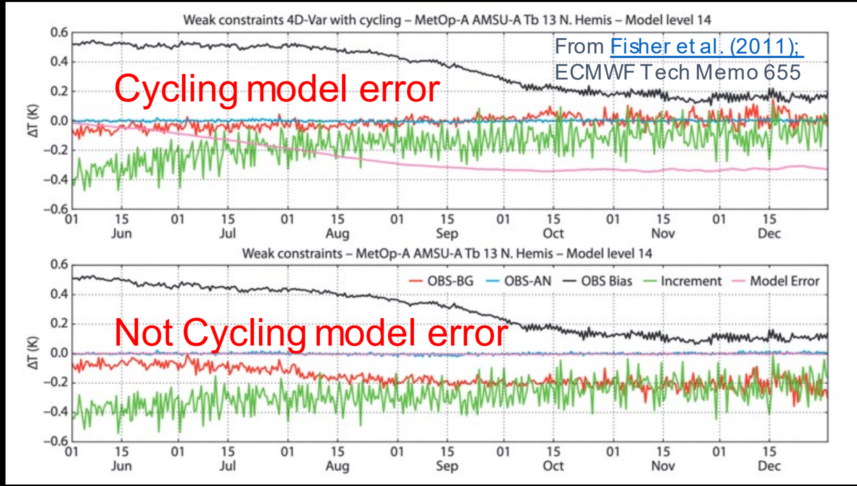
AKA: Uncertainty Quantification

- Model parameter estimation



E.g., Cloud Optical Thickness over MODIS pass: background, observations, analysis; Norris & da Silva (2016; Part II)

- Weak Constraint 4D-Var



Ensemble-derived for DA & UQ …

Introduced to address nonlinearities & efficiency …

Also good for UQ …

Uncertainty Quantification (UQ): Ensemble DA

Observations (O)

Model

Constrained (X)
Unconstrained (Z)

Bayesian Applications

$$p(\mathbf{X}|\mathbf{O}) = \frac{p(\mathbf{O}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{O})}$$

In this perspective there are two levels of uncertainty required to make DA work, those being in the:

Observations: $\mathbf{O} \rightarrow \mathbf{R}$

Model: $\mathbf{X} \rightarrow \mathbf{B}, \mathbf{Q}$

Straightforward DA and ML do not provide UQ on Z, but MC-based do.

Ensemble Model

Hybrid DA

Ensemble of DA

Monte Carlo-like DA

- U from Adaptive DA
- U Propagation (X,Z)

- Probabilities (Fair) Scores
- MC / Ensemble Learning
- ML for obs error covariance enhancement (LSTM) RNN.
- ML/WC for model error estimation.

- ML to improve Predictions (offline)
- ML for Downscaling.

Here understood as:

Constrained Vars: variables directly affected by DA/ML;
can include model parameters.
Unconstrained Vars: everything else derived from a model.

UQ of U-Var are desirable in many areas: climate research, instrumentation, risk analysis, improved DA/ML methodologies, validation, etc.

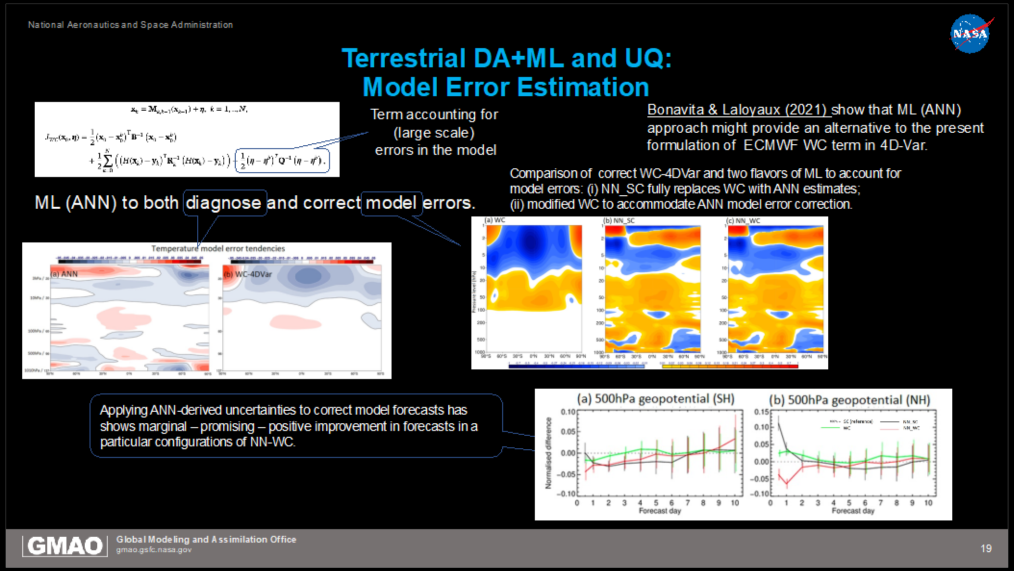
Did anybody say Digital Twin?

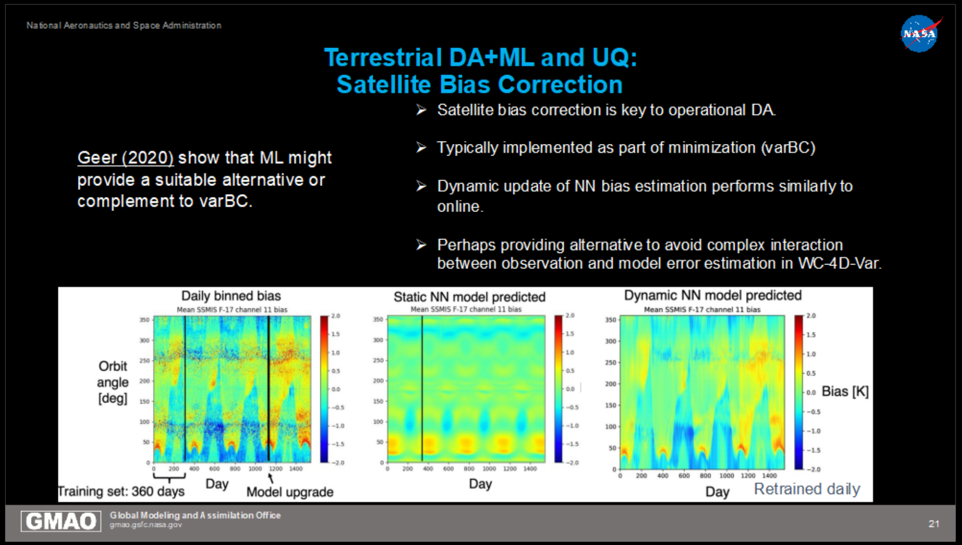
ML as Aiding Device for DA …

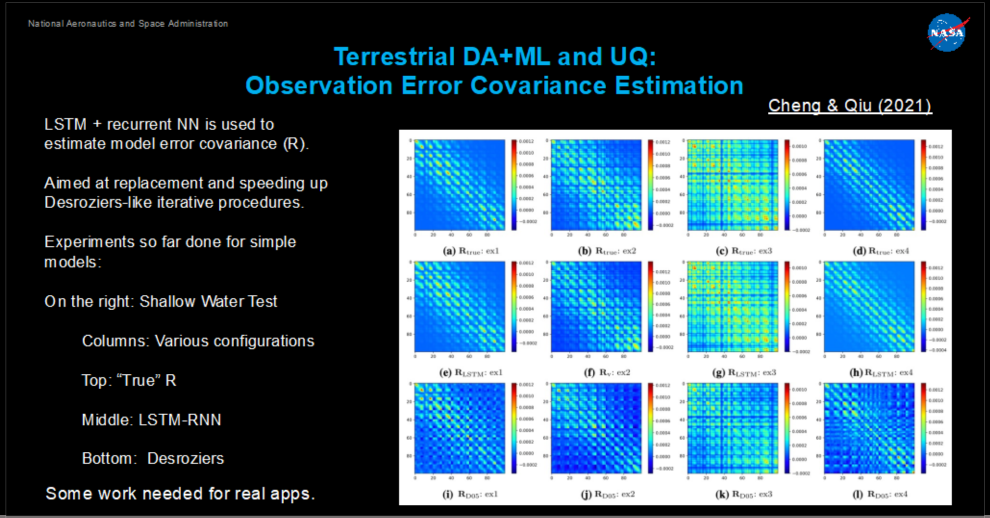
Alternative ways to derive uncertainties …

Another tool to address efficiency …

ML to Aid DA: Terrestrial UQ in DA







Top Left: ANN for WC-4D-Var model error estimation

Top Right: Dynamic NN for satellite bias estimation

Bottom Right: LSTM-Recurrent NN for obs error cov

Full slides in Appendix

Frameworks …

The way for community collaboration …

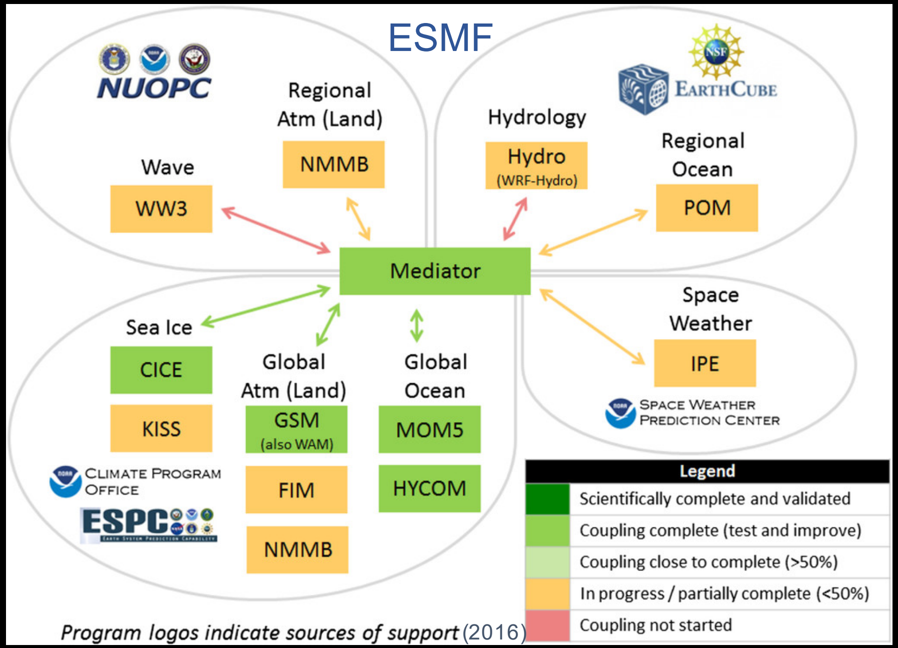
Facilitating R2O & O2R …

Facilitating rapid deployment of science …

Communities Modeling Frameworks

Terrestrial Weather Modeling

ESPS – Earth System Prediction Suite



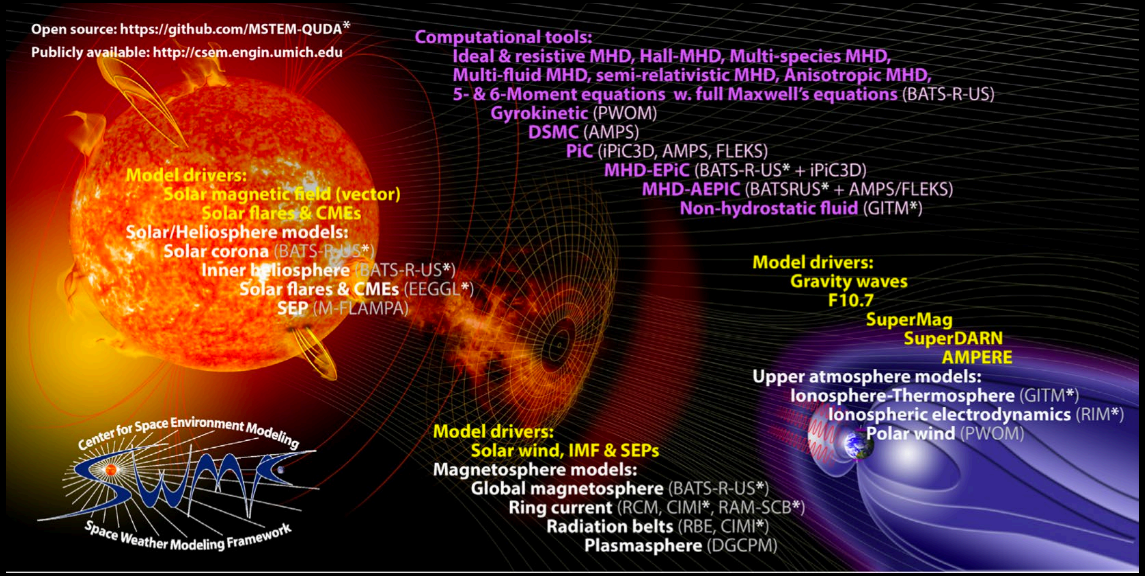
From Theurich et al. (2016)

ESMF – Earth System Modeling Framework

NUOPC - National Unified Operational Prediction Capability

Space Weather Modeling

SWMF - Space Weather Modeling Framework



From Gombosi et al. (2021)

Is a modeling framework adequate for DA?

The Terrestrial community has had this discussion and it has largely decided to answer NO.

Remarks on Frameworks for DA & ML

ML applications have condensed into general, portable, Phyton/C++-based libraries that have been globally-embraced by the user community and has facilitated continual development and enhancement data analytics, e.g.: Keras, SciKit-Learn, TensorFlow, PyTorch.

DA applications have been slow to condense into a "globally-embraced" framework. The past few years has seen the community reach a consensus on the *need* for a framework; competing frameworks exist at present:

Terrestrial DA Truly Opensource Frameworks

- EMPIRE (Reading, UK)
- DART (NCAR, USA)
- PDAF (AWI, DE)
- OpenDA (TU Delft, NL)
- JEDI (JCSDA, USA)

Space DA Opensource? Frameworks?

- ADAPT (Air Force, LANL, USA)
- DREAM (LANL, USA)

- Is there need for a specific SWDAF? Probably NOT!
- NSF/NASA on Frameworks 2020

It has been acknowledged that these frameworks must interface with ML software (e.g., JEDI).

JEDI is the framework for DA development adopted by NOAA, NASA, US Navy, US Air Force, and others. JEDI is not yet operational, but schedules are set on that. The U.K. Met Office is also committed to JEDI.

It might be helpful for the SW-DA-ML community to embrace existing DA frameworks; in the USA, JEDI.

Closing Thoughts

- Data Assimilation:
  - Can be viewed as a *traditional* machine learning device.
  - Adaptive procedures render DA self correcting & robust.

But …traditional DA is hard to implement, maintain, and inefficient:

- Ensemble techniques are fundamental to address part of such issues & provide path to UQ.

- Modern ML techniques allow for further improvement of DA through:
  - Surrogate modeling.
  - Covariance estimation.
  - Characterization of uncertainties.

- DA Frameworks should allow for agile R2O2R & to keep up with Exascale endeavors.

Thank you

Current status and future trends in ML-enhanced Space Weather predictions

Enrico Camporeale
(enrico.camporeale@noaa.gov)

CIRES / CU Boulder & NOAA Space Weather Prediction Center

Thanks to:
H. Singer, M. Cash, C. Balch, E. Adamson, G. Toth, Z. Huang, J. Bortnik, G. Wilkie, A. Drozdov, M. Gruet, M. Chandorkar, A. Care’, J. Borovsky, G. Lapenta, X. Chu, R. McGranaghan, …, and probably others…

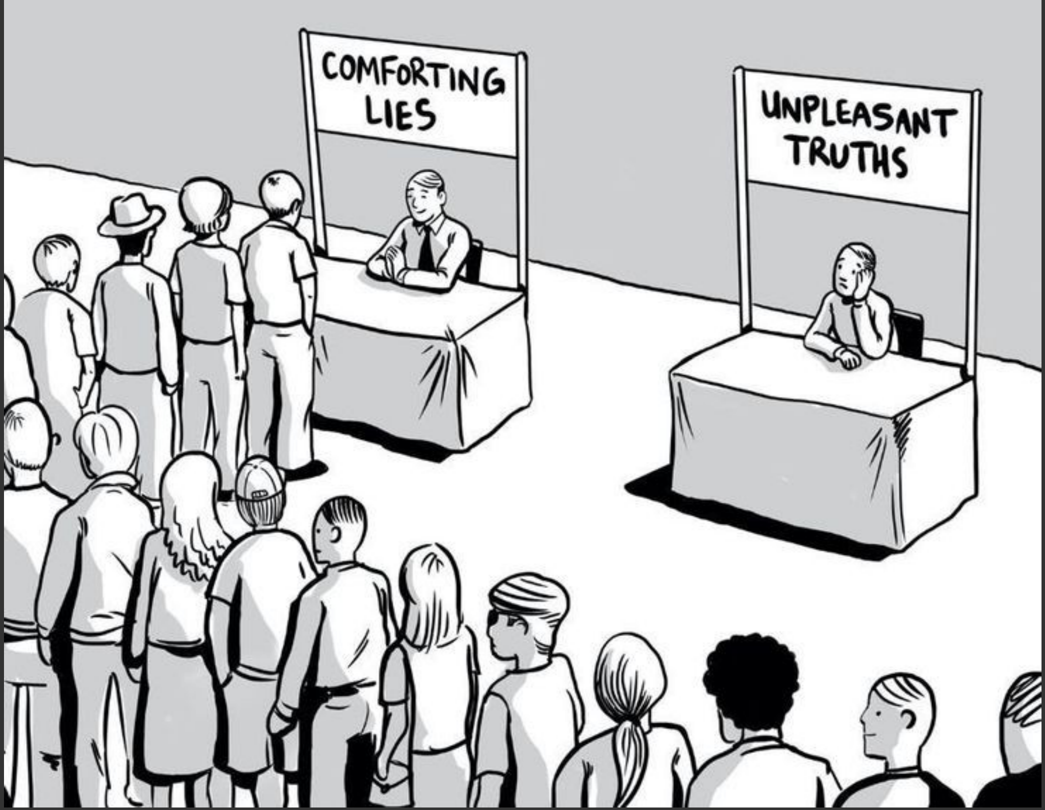
This project is supported by NASA under grant 80NSSC20K1580

University of Colorado Boulder

CIRES

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

The unpleasant truth...



The unpleasant truth...

Machine Learning is revolutionizing the world...

The unpleasant truth…

Machine Learning is revolutionizing the world…

...and it is **reinventing Space Weather**

What can ML do for Space Weather?
(a non-comprehensive list)

- ML works better than physics-based simulations to forecast global/average indexes such as Dst
  - Why? Because in a physics-based approach of a complex system you need to get ‘every single piece right’

The Dst (Disturbance storm time) index is an index of magnetic activity derived from a network of near-equatorial geomagnetic observatories

Space Weather

RESEARCH ARTICLE

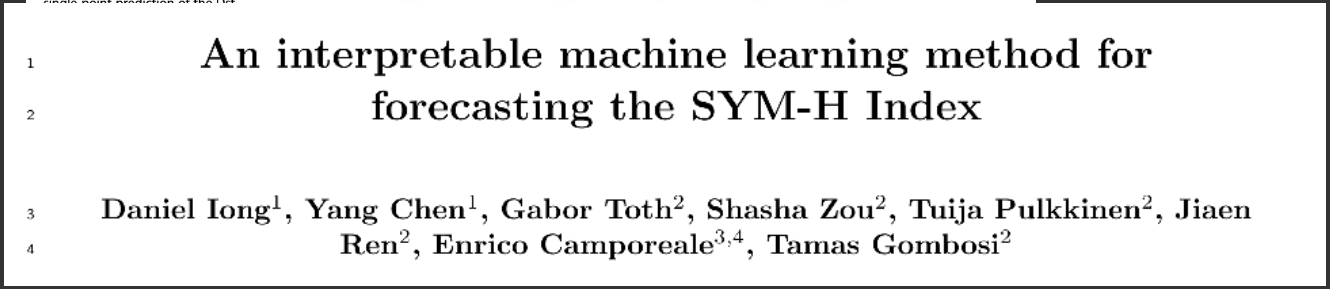
10.1029/2018SW001898

Key Points:

- First use of a Long Short-Term Memory network to provide single point prediction of the Dst

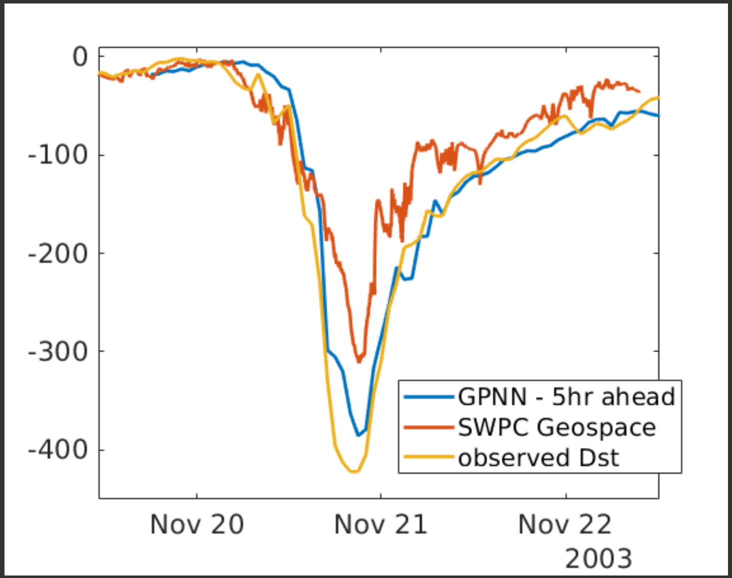
Multiple-Hour-Ahead Forecast of the Dst Index Using a Combination of Long Short-Term Memory Neural Network and Gaussian Process

M. A. Gruet[1], M. Chandorkar[2], A. Sicard[1], and E. Camporeale[2]



An interpretable machine learning method for forecasting the SYM-H Index

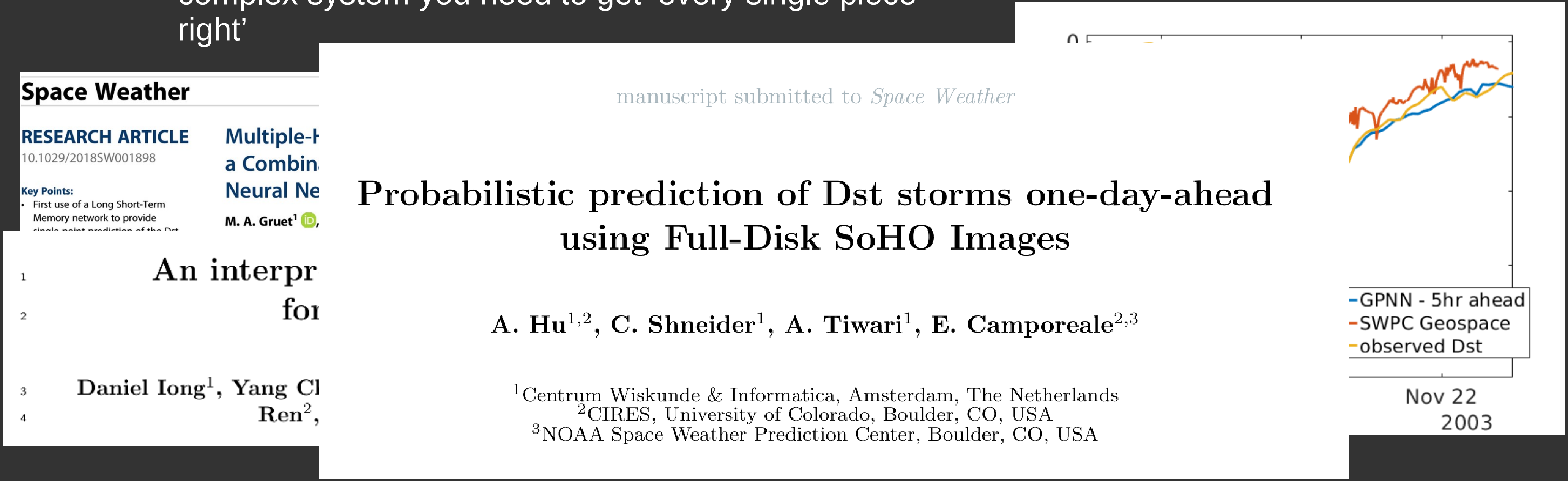
Daniel Iong[1], Yang Chen[1], Gabor Toth[2], Shasha Zou[2], Tuija Pulkkinen[2], Jiaen Ren[2], Enrico Camporeale[3,4], Tamas Gombosi[2]



What can ML do for Space Weather?
(a non-comprehensive list)

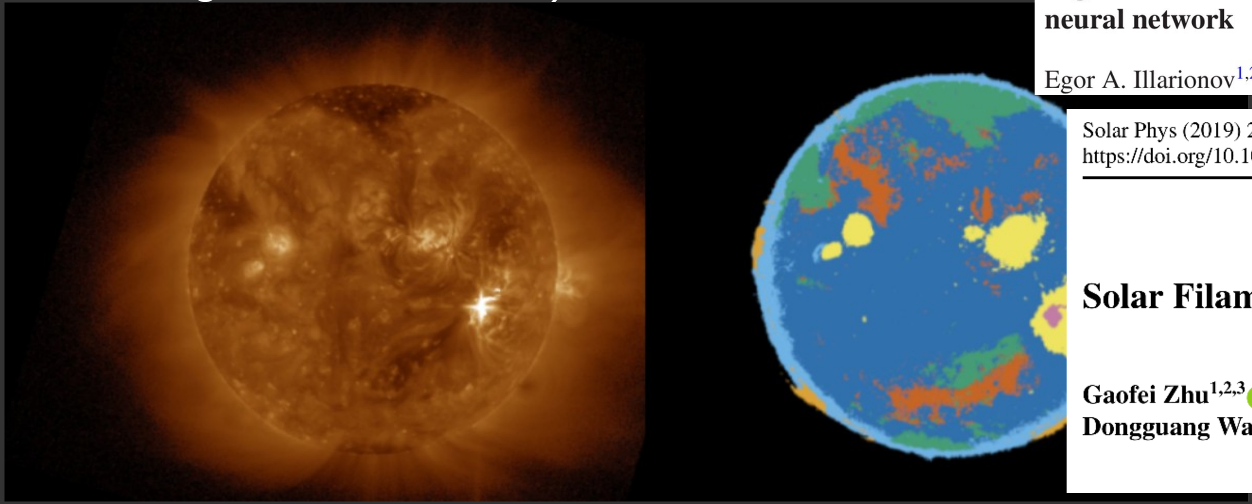
- ML works better than physics-based simulations to forecast global/average indexes such as Dst
  - Why? Because in a physics-based approach of a complex system you need to get ‘every single piece right’

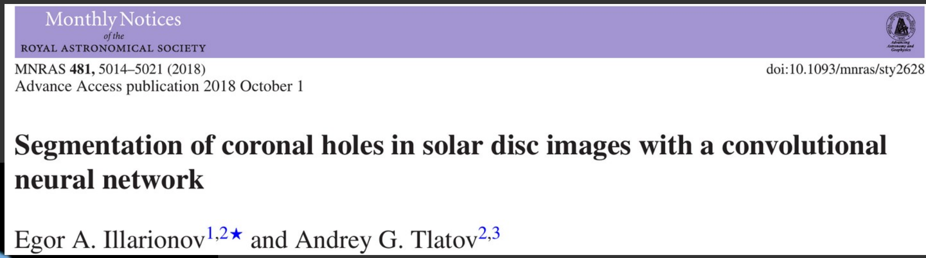
The Dst (Disturbance storm time) index is an index of magnetic activity derived from a network of near-equatorial geomagnetic observatories



What can ML do for Space Weather?
(a non-comprehensive list)

- Segmentation of solar disk images (supervised or unsupervised):
  - Automatically extract different solar regions (that are associated with different solar wind/geoeffectiveness)



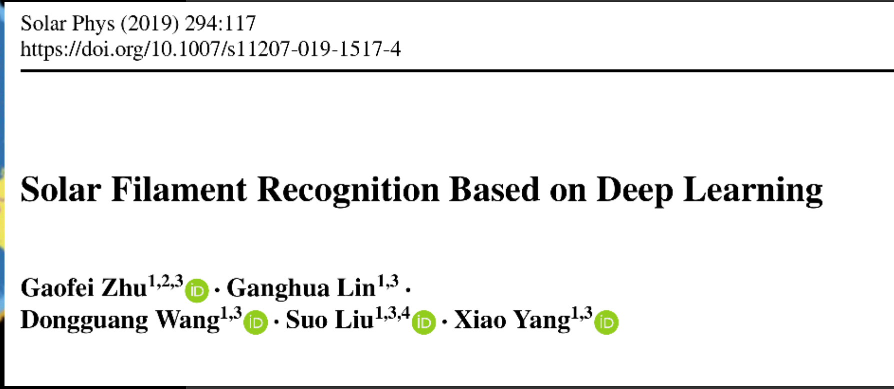


Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **481,** 5014–5021 (2018) doi:10.1093/mnras/sty2628
Advance Access publication 2018 October 1

Segmentation of coronal holes in solar disc images with a convolutional neural network

Egor A. Illarionov[1,2★] and Andrey G. Tlatov[2,3]



Solar Phys (2019) 294:117
https://doi.org/10.1007/s11207-019-1517-4

Solar Filament Recognition Based on Deep Learning

Gaofei Zhu[1,2,3] · Ganghua Lin[1,3] · Dongguang Wang[1,3] · Suo Liu[1,3,4] · Xiao Yang[1,3]

Courtesy of Dan Seaton and J. Marcus Hughes, NCEI, CIRES, and University of Colorado Boulder

What can ML do for Space Weather?
(a non-comprehensive list)

- Solar flare prediction

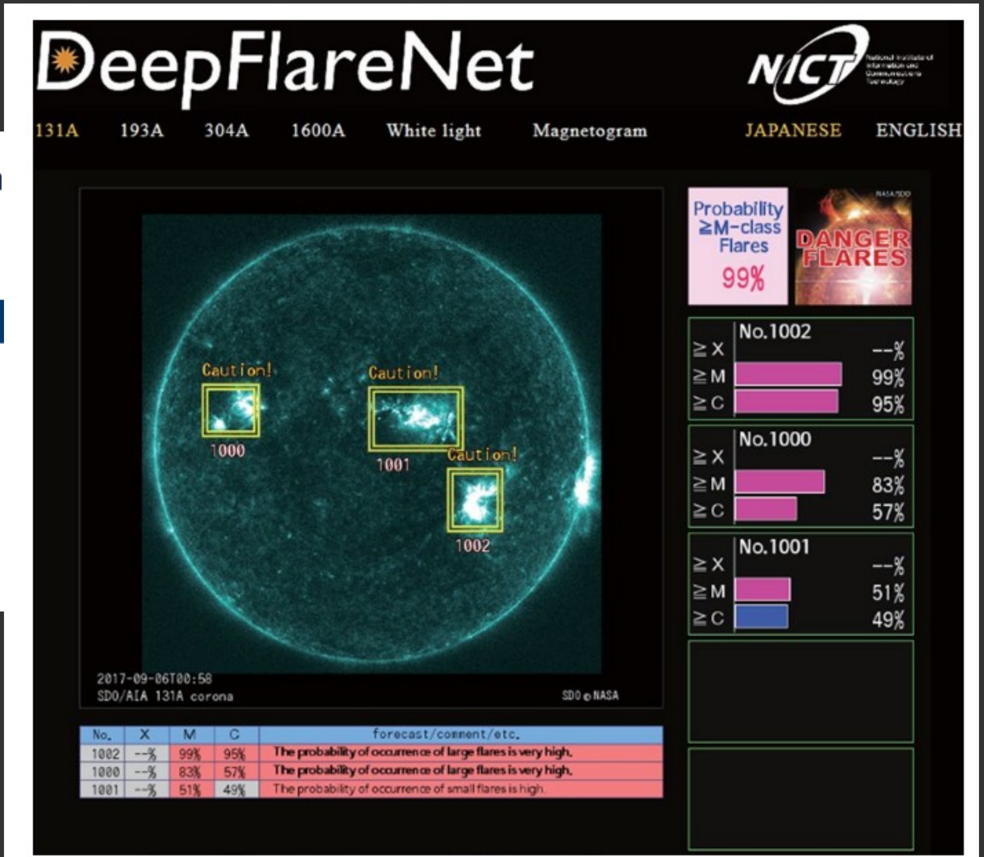
Nishizuka *et al. Earth, Planets and Space* (2021) 73:64
https://doi.org/10.1186/s40623-021-01381-9

Earth, Plan

FULL PAPER

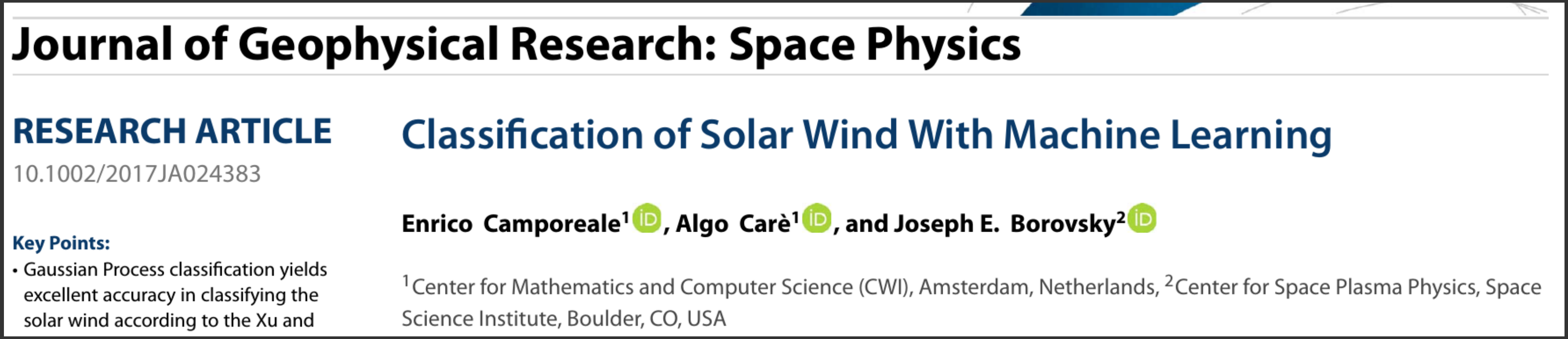
Operational solar flare prediction model using Deep Flare Net

Naoto Nishizuka[1*], Yûki Kubo[1], Komei Sugiura[2], Mitsue Den[1] and Mamoru Ishii[1]



What can ML do for Space Weather?
(a non-comprehensive list)

- Solar wind classification (supervised):
  - Extending human labeled database from 8000 hrs (<1 year) to 40+ years



Journal of Geophysical Research: Space Physics

RESEARCH ARTICLE

10.1002/2017JA024383

Key Points:

- Gaussian Process classification yields excellent accuracy in classifying the solar wind according to the Xu and

Classification of Solar Wind With Machine Learning

Enrico Camporeale[1], Algo Carè[1], and Joseph E. Borovsky[2]

[1]Center for Mathematics and Computer Science (CWI), Amsterdam, Netherlands, [2]Center for Space Plasma Physics, Space Science Institute, Boulder, CO, USA

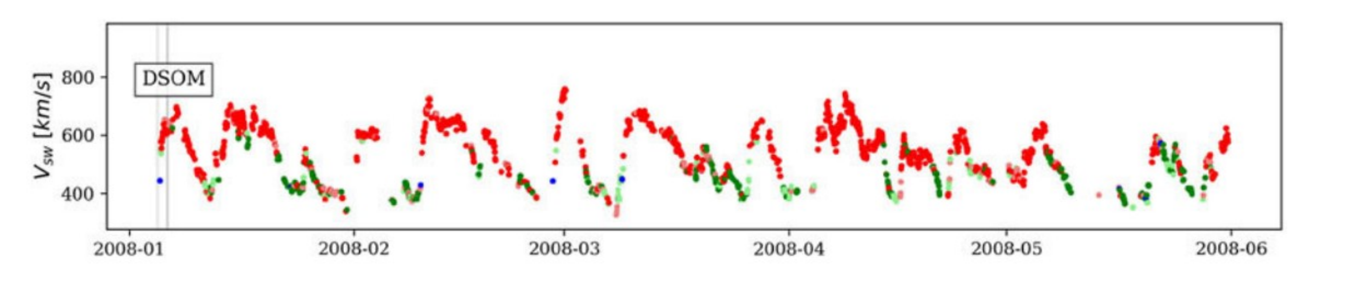
What can ML do for Space Weather?
(a non-comprehensive list)

- Solar wind classification (unsupervised):

Visualizing and Interpreting Unsupervised Solar Wind Classifications

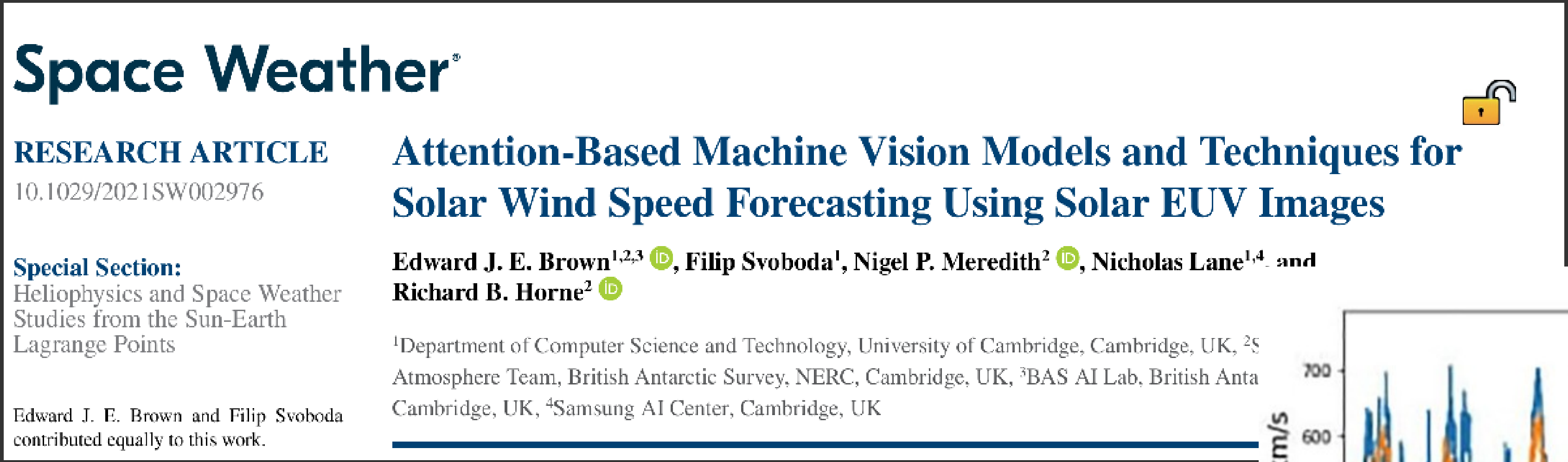
Jorge Amaya, Romain Dupuis, Maria Elena Innocenti and Giovanni Lapenta*

Mathematics Department, Centre for Mathematical Plasma-Astrophysics, KU Leuven, Leuven, Belgium



What can ML do for Space Weather?
(a non-comprehensive list)

- Solar wind speed forecast



Space Weather

RESEARCH ARTICLE
10.1029/2021SW002976

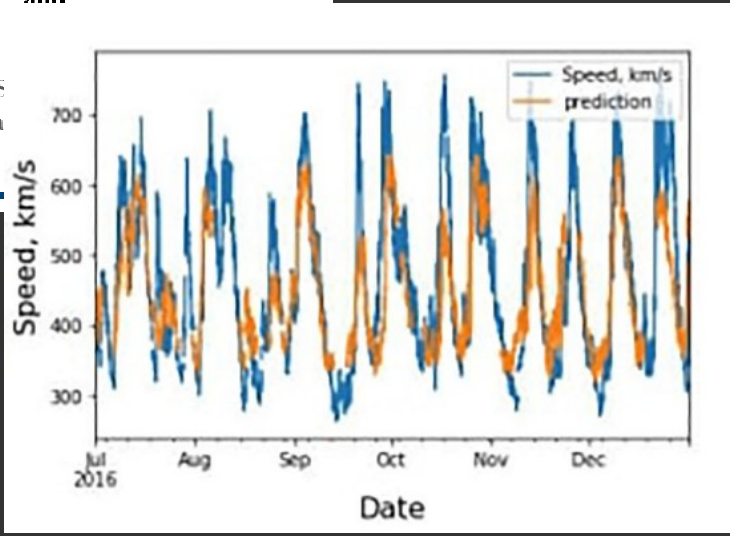
Attention-Based Machine Vision Models and Techniques for Solar Wind Speed Forecasting Using Solar EUV Images

Edward J. E. Brown[1,2,3], Filip Svoboda[1], Nigel P. Meredith[2], Nicholas Lane[1,4], and Richard B. Horne[2]

[1]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK, [2]S
Atmosphere Team, British Antarctic Survey, NERC, Cambridge, UK, [3]BAS AI Lab, British Anta
Cambridge, UK, [4]Samsung AI Center, Cambridge, UK

Special Section:
Heliophysics and Space Weather Studies from the Sun-Earth Lagrange Points

Edward J. E. Brown and Filip Svoboda contributed equally to this work.



What can ML do for Space Weather?
(a non-comprehensive list)

- Radiation belt physics

Space Weather

RESEARCH ARTICLE

10.1029/2021SW002808

Relativistic Electron Model in the Outer Radiation Belt Using a Neural Network Approach

Xiangning Chu[1], Donglai Ma[2], Jacob Bortnik[2], W. Kent Tobiska[3], Alfredo Cruz[3], S. Dave Bouwer[3], Hong Zhao[4], Qianli Ma[2,5], Kun Zhang[6], Daniel N. Baker[1], Xinlin Li[1], Harlan Spence[7], and Geoff Reeves[8]

Key Points:

- A neural network model was developed to forecast relativistic electron fluxes with energies

Research

Data-driven discovery of Fokker-Planck equation for the Earth's radiation belts electrons using Physics-Informed Neural Networks

E. Camporeale[1,2], George J. Wilkie[3], Alexander Drozdov[4], Jacob Bortnik[4]

What can ML do for Space Weather?
(a non-comprehensive list)

- Regression problems, i.e. predict:
  - The value of a geomagnetic index (Dst, Kp, etc.);
  - The arrival time of a Coronal Mass Ejection;
  - Global Total Electron Content (TEC) maps;
  - Solar wind speed;
  - Relativistic electrons at GEO;
  - Ground magnetic field (dB/dt)
  - Electron precipitation

What can ML do for Space Weather?
(a non-comprehensive list)

- Classification problems, i.e. what is the probability that:
  - An active region will flare in the next 24 hours?
  - dB/dt will exceed a given value?
  - The solar wind is originated by coronal holes/ejecta, etc.
  - A region of the Sun belongs to a coronal hole

Why does it work (so well) ?

A short digression

The Unreasonable Effectiveness of Mathematics in the Natural Sciences

Richard Courant Lecture in Mathematical Sciences delivered at New York University, May 11, 1959

EUGENE P. WIGNER

Princeton University

“The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve.”

Why does it work (so well) ?

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

The unreasonable effectiveness of deep learning in artificial intelligence

Terrence J. Sejnowski[a,b,1]

[a]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037; and [b]Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093

We are not in the same boat with image and text recognition, self-driving, or recommendation systems!

Why does it work (so well) ?

Physics to the rescue!

- Physical properties such as invariance, symmetry, conservation laws, etc. reduce drastically the ‘search space’ of parameters
- Any system that follows ‘laws of physics’ should be learnable by Machine Learning
- Any simulation can be emulated by ML
- The major hurdle is **Data Quality & Quantity!**

J Stat Phys (2017) 168:1223–1247
DOI 10.1007/s10955-017-1836-5

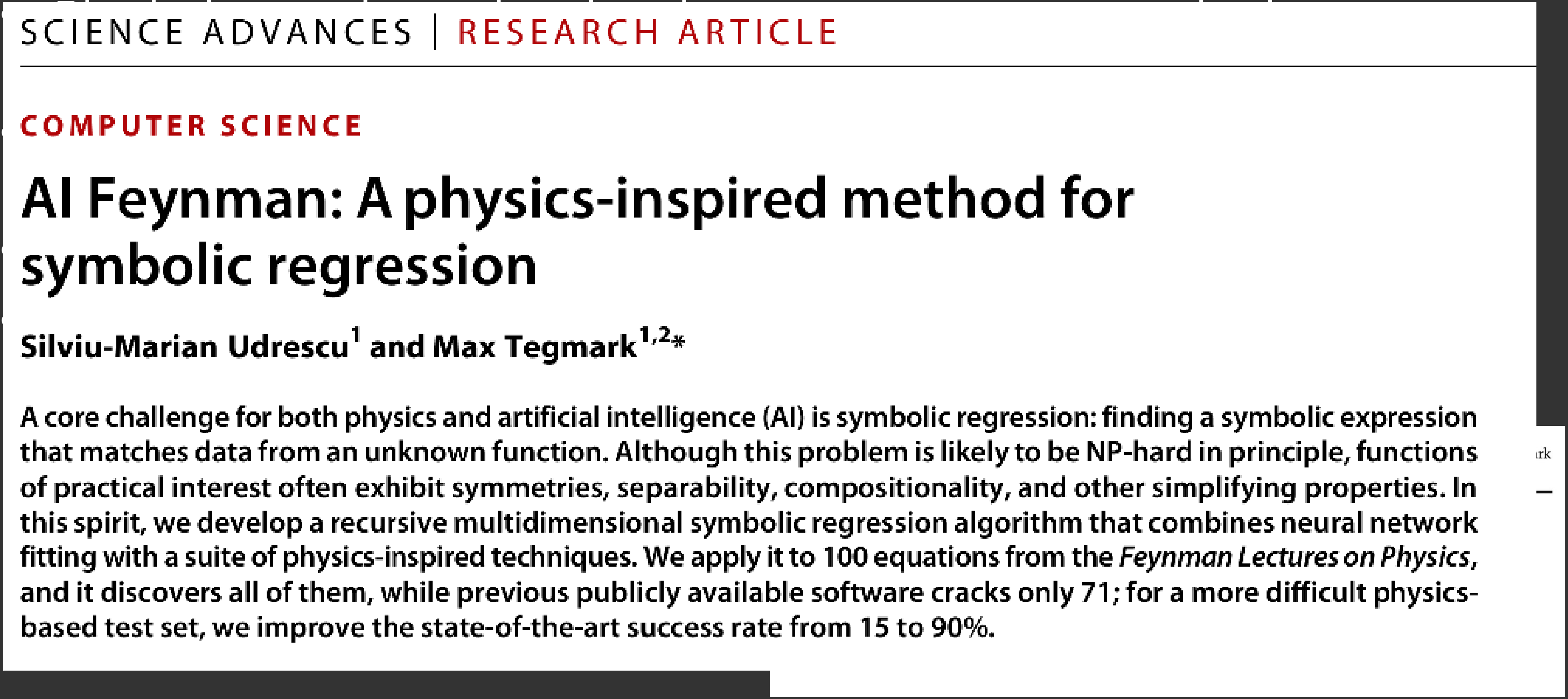
CrossMark

Why Does Deep and Cheap Learning Work So Well?

Henry W. Lin[1] · Max Tegmark[2] · David Rolnick[3]

Why does it work (so well) ?

Physics to the rescue!



SCIENCE ADVANCES | RESEARCH ARTICLE

COMPUTER SCIENCE

AI Feynman: A physics-inspired method for symbolic regression

Silviu-Marian Udrescu[1] and Max Tegmark[1,2]*

A core challenge for both physics and artificial intelligence (AI) is symbolic regression: finding a symbolic expression that matches data from an unknown function. Although this problem is likely to be NP-hard in principle, functions of practical interest often exhibit symmetries, separability, compositionality, and other simplifying properties. In this spirit, we develop a recursive multidimensional symbolic regression algorithm that combines neural network fitting with a suite of physics-inspired techniques. We apply it to 100 equations from the *Feynman Lectures on Physics*, and it discovers all of them, while previous publicly available software cracks only 71; for a more difficult physics-based test set, we improve the state-of-the-art success rate from 15 to 90%.

Path forward for ML in SWx

Freely adapted from:

Space Weather

FEATURE ARTICLE

10.1029/2018SW002061

GRAND CHALLENGES
CENTENNIAL COLLECTION

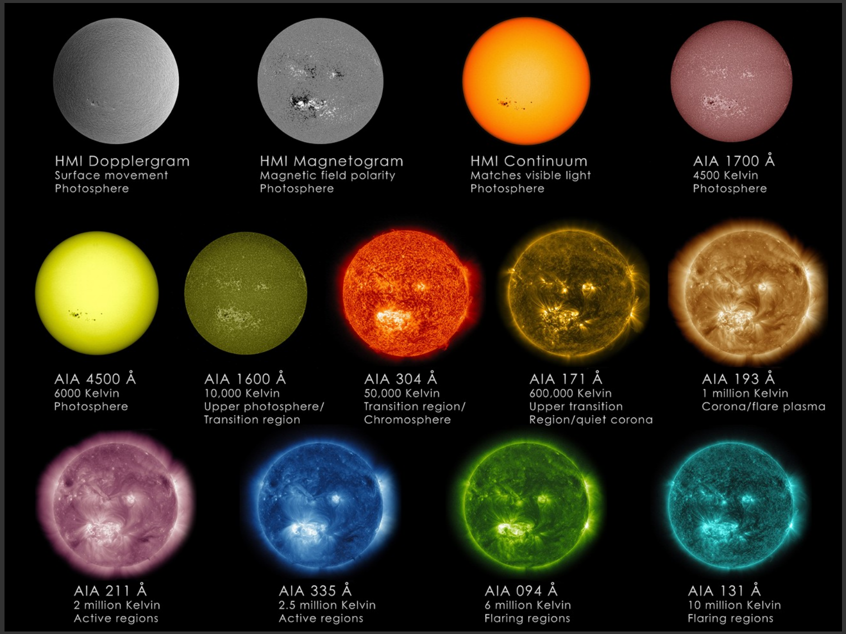
The Challenge of Machine Learning in Space Weather: Nowcasting and Forecasting

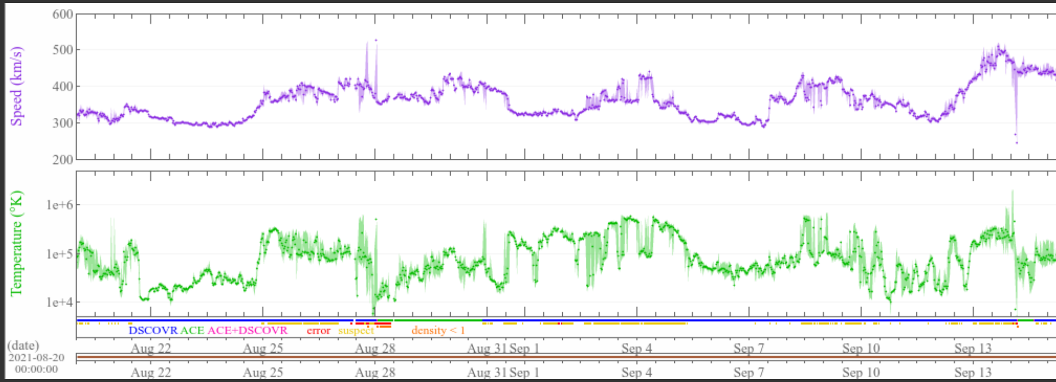
E. Camporeale[1,2]

[1]CIRES, University of Colorado Boulder, Boulder, CO, USA, [2]Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

Path forward for ML in SWx

- *The information problem*: What is the minimal physical information required to make a forecast?

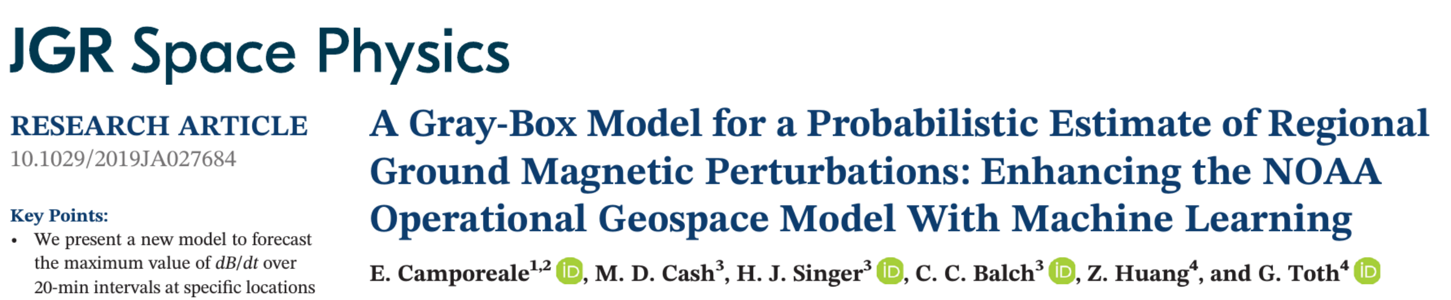




200M pixels → 1 scalar value

Path forward for ML in SWx

- *The gray-box problem:* What is the best way to make an optimal use of both our physical understanding and our large amount of data in the Sun-Earth system?



JGR Space Physics

RESEARCH ARTICLE

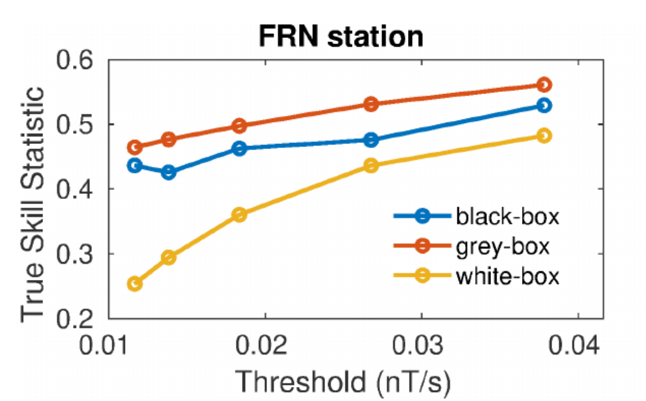
10.1029/2019JA027684

Key Points:

- We present a new model to forecast the maximum value of *dB*/*dt* over 20-min intervals at specific locations

A Gray-Box Model for a Probabilistic Estimate of Regional Ground Magnetic Perturbations: Enhancing the NOAA Operational Geospace Model With Machine Learning

E. Camporeale[1,2], M. D. Cash[3], H. J. Singer[3], C. C. Balch[3], Z. Huang[4], and G. Toth[4]



Path forward for ML in SWx

- *The surrogate problem:* What components in the Space Weather chain can be replaced by an approximated black-box surrogate model? What is an acceptable trade-off between lost of accuracy and speed-up?

Path forward for ML in SWx

- *The uncertainty problem:* Most Space Weather services provide forecast in terms of single-point predictions. There is a clear need of understanding and assessing the uncertainty associated to these predictions. Propagating uncertainties through the Space Weather chain from solar images to magnetospheric and ground-based observations is a complex task that is computationally demanding.

Space Weather

RESEARCH ARTICLE

10.1029/2018SW002026

On the Generation of Probabilistic Forecasts From Deterministic Models

E. Camporeale[1,2], X. Chu[3], O. V. Agapitov[4], and J. Bortnik[5]

Key Points:
- We introduce a new method to estimate the uncertainties associated

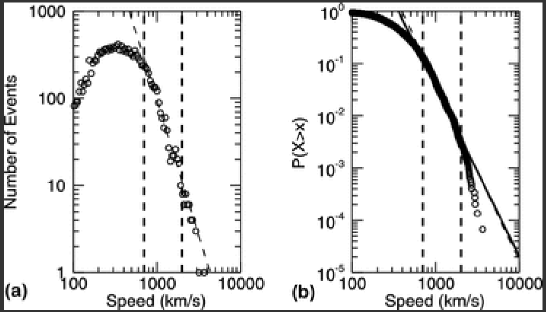
International Journal for Uncertainty Quantification, 11(4):81–94 (2021)

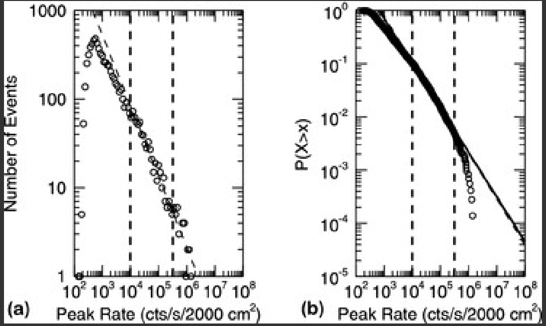
ACCRUE: ACCURATE AND RELIABLE UNCERTAINTY ESTIMATE IN DETERMINISTIC MODELS

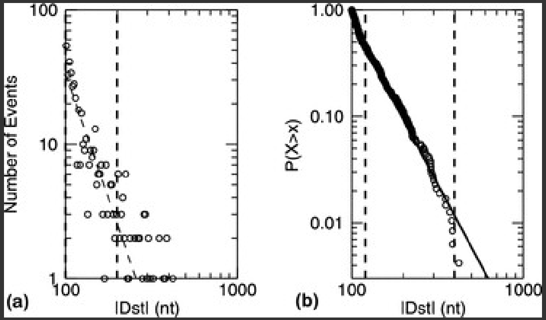
Enrico Camporeale[1,] & Algo Carè[2]*

Path forward for ML in SWx

- *The too often too quiet problem:* Space weather data sets are typically imbalanced: many days of quiet conditions and a few hours of storms. This poses a serious problem for any machine learning algorithm. It is also problematic for defining meaningful metrics that actually assess the ability of a model to predict interesting but rare events.







SPACE WEATHER, VOL. 10, S02012, doi:10.1029/2011SW000734, 2012

On the probability of occurrence of extreme space weather events

Pete Riley[1]

Path forward for ML in SWx

- *The knowledge discovery and explainability problem:* How do we distill some knowledge from a machine learning model and improve our understanding of a given system? How do we open the black-box and reverse-engineer a machine learning algorithm?

arXiv.org > physics > arXiv:2107.14322

Physics > Space Physics

[Submitted on 29 Jul 2021]

Machine-learning based discovery of missing physical processes in radiation belt modeling

Enrico Camporeale, George J. Wilkie, Alexander Drozdov, Jacob Bortnik

Summary

ML 4 SWx is the quintessential interdisciplinary discipline.

These 6 problems not only hinder progress in Space Weather, but pose fundamental challenges in the fields of AI and UQ.

- *The information problem*
- *The gray-box problem*
- *The surrogate problem*
- *The uncertainty problem*
- *The too often too quiet (rare events) problem*
- *The knowledge discovery and explainability problem*

Contact: enrico.camporeale@noaa.gov