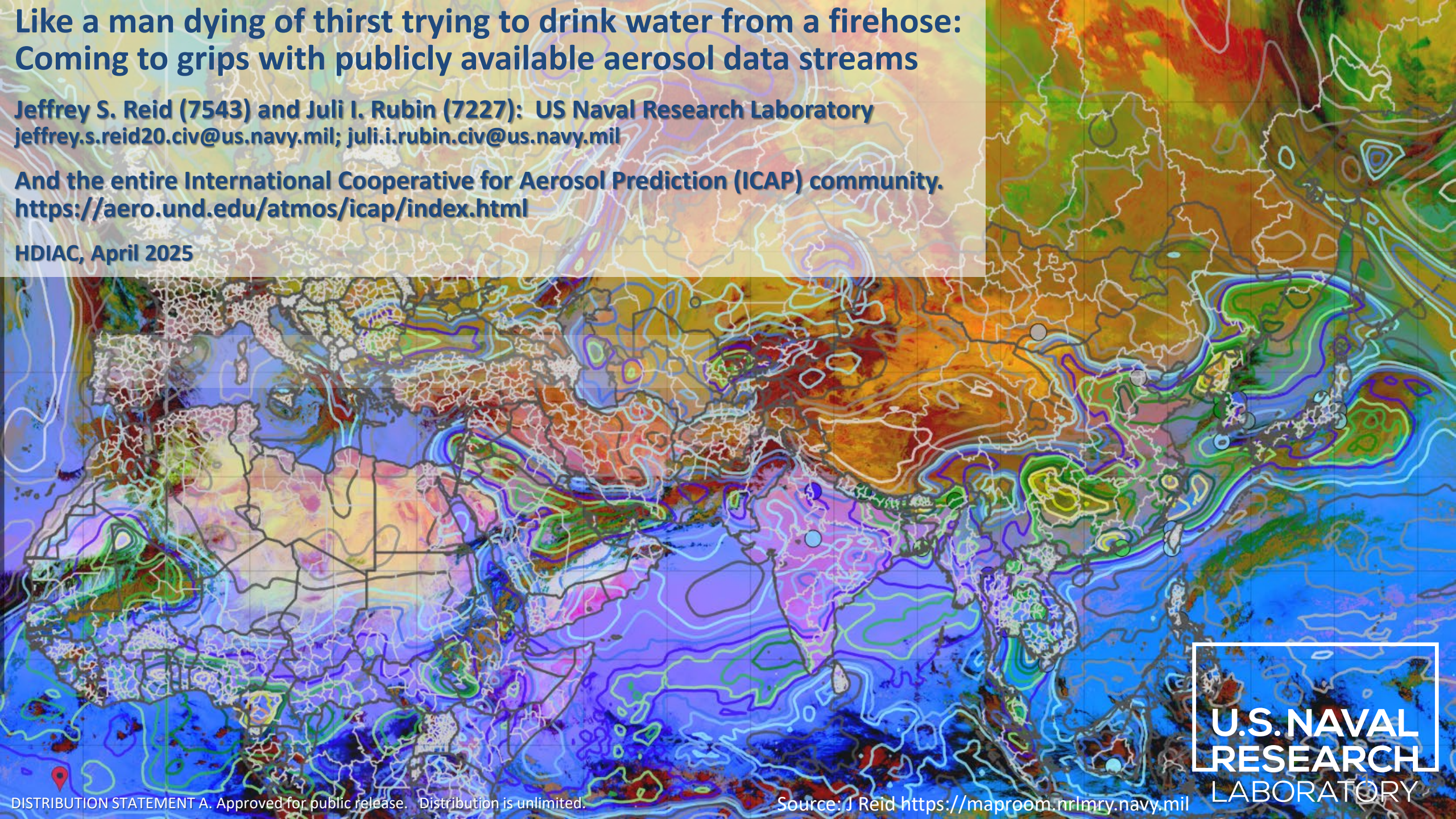
Like a man dying of thirst trying to drink water from a firehose: Coming to grips with publicly available aerosol data streams

Jeffrey S. Reid (7543) and Juli I. Rubin (7227): US Naval Research Laboratory
jeffrey.s.reid20.civ@us.navy.mil; juli.i.rubin.civ@us.navy.mil

And the entire International Cooperative for Aerosol Prediction (ICAP) community.
https://aero.und.edu/atmos/icap/index.html

HDIAC, April 2025



Source: J Reid https://maproom.nrlmry.navy.mil

U.S. NAVAL RESEARCH LABORATORY

Bottom Lines Up Front

Global environmental monitoring and prediction is like "space." To quote Douglas Adams:

"Space is big. Really big. You just won't believe how vastly hugely mind-bogglingly big it is. I mean, you may think it's a long way down the road to the chemist, but that's just peanuts to space"

Monitoring and prediction is fundamentally diverse, as it is funded by a near infinite number of sources for different purposes. Ultimately, the field is an amalgamation of lots of independent parts.

Data are broadly classified as in situ measurements, satellite measurements, and models, all pulled together in syntheses and data assimilation.

And time are changing, with AI/ML and commercial development in environmental monitoring, prediction, and decision aids.

The successful ones in the business keep their head together by being efficient on data, do only what is necessary and paying attention to error and uncertainty.

Source: ImgFlip

Goals for today's seminar:

Provide a very brief overview of the nature of Numerical Weather and Aerosol prediction systems. Whole degree programs are based on this. But it is a start.

- DoD context: Why we care about weather and aerosol forecasting?
- The nature of battlespace prediction.
- A huge stream of data: Observations, satellite, and models, oh my.
- Community practices for data distributions.
- Some thoughts on AI and Industry.
- Caveat emptor.

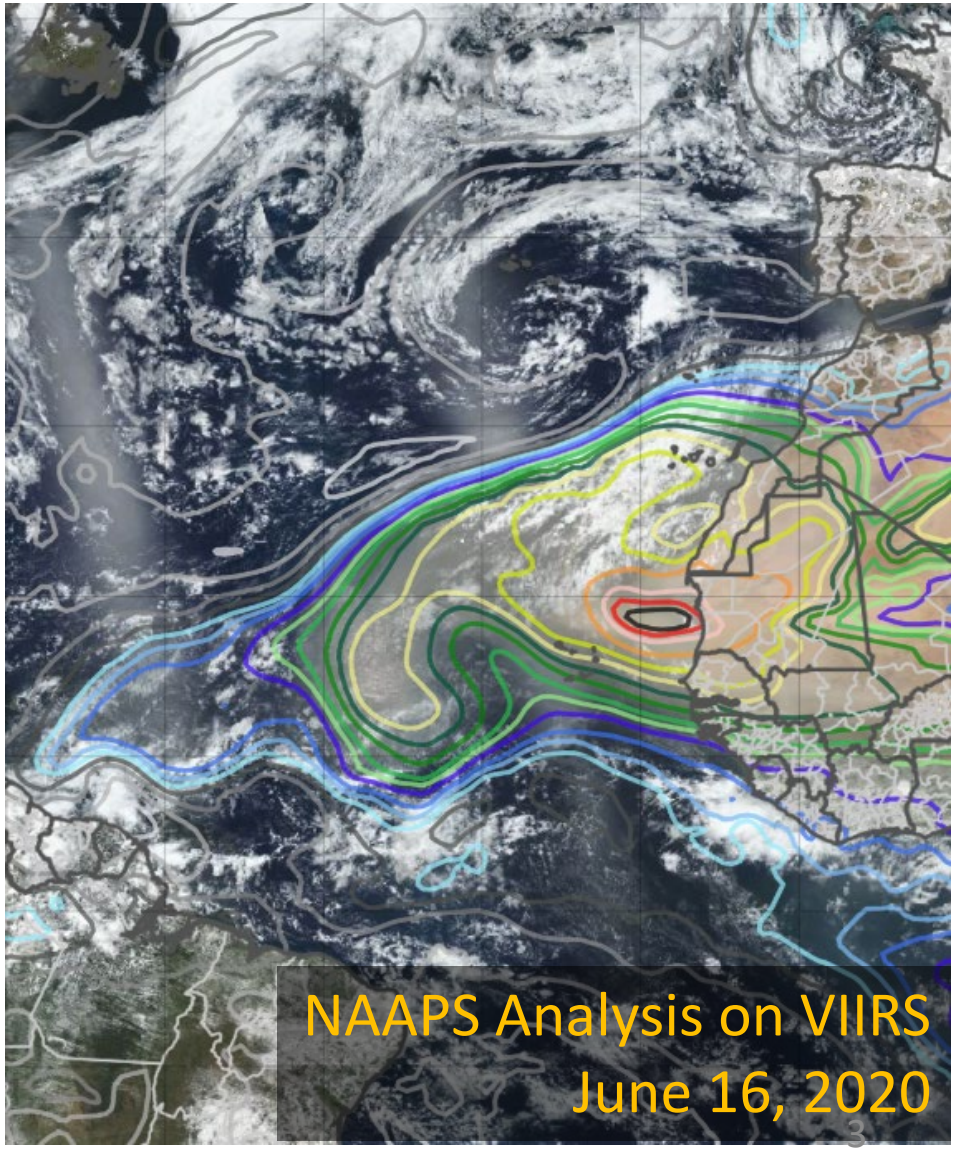
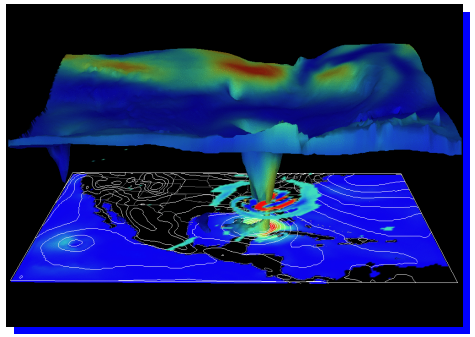


Image Source: Reid J. /NRL https://maproom.nrlmry.navy.mil

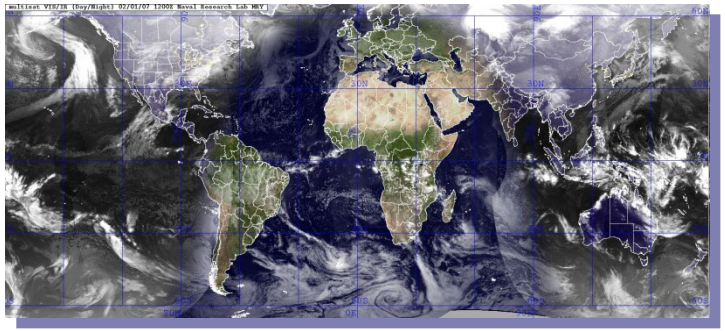
NRL Monterey MARINE METEOROLOGY DIVISION

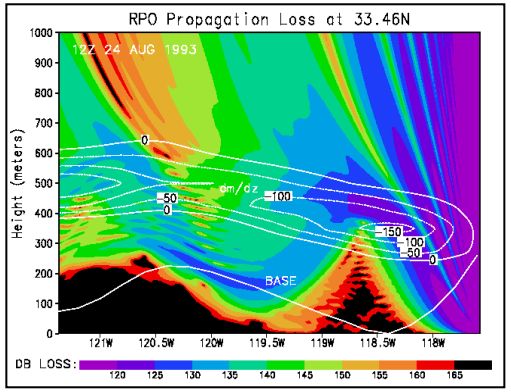
Our Core Business of ~ 100 Scientists and Engineers



To sample, understand, and simulate the behavior of the atmosphere and its constituents on local, regional, and global scales, including its interaction with the ocean, land, cryosphere, and middle atmosphere.

To invent, develop, and implement new capabilities and systems for objective environmental analysis and prediction.





To integrate new capabilities into automated systems for assessing risk and analyzing impact of atmospheric conditions on Navy/DoD operations and weapons systems.

Bridging the Valley of Death

NRL's nationally unique vertical integration:

Basic research feeds applied science, that drives prototype development for operations. Operational products then in turn feedback basic research.

Basic Research ↔ ***Applied Science*** ↔ ***Prototype*** ↔ ***Operations***

Key Sponsors: Navy, Air Force, NASA, NOAA, + Industry

Images Source: Reid J. /NRL

Example: Consequences of poor weather and visibility planning

Iranian Hostage
Op Eagle Claw
Dust Storm Disrupted Mission

Osama Bin Laden: Op Neptune Spear
SEAL team waited a day in part for weather

A “closer to home” example for me.

USA TODAY | MOSQUITO SEASON Hotspots ranked | NEWS TO YOUR INBOX Start the day smarte... | GRADUATION TIME Funniest cap messag...

U.S. Politics Sports Entertainment Life Money Travel Opinion Crossword

NATION Monterey Add Topic +

'Terrifying': Monterey County declares emergency after massive lithium battery plant fire

The massive blaze that began a week ago forced the evacuation of 1,200 residents before it burned itself out. Now residents are worried about safe cleanup and what the future holds.

James Powel and Roseann Cattani USA TODAY NETWORK

Jan. 23, 2025 | Updated Jan. 24, 2025, 10:40 a.m. ET

https://www.usatoday.com/story/news/nation/2025/01/23/moss-landing-lithium-battery-plant-fire-vistra/77912642007/

NOWCAST Watch on Demand

MENU KSBW 8 ACTION NEWS SUBSCRIBE TO EMAIL 63° WEATHER

Monterey, CA 93940 63° Sunny 1% 1 / 1

EPA says testing shows air is safe after Moss Landing fire

https://www.ksbw.com/article/epa-testing-chemical-residue-air-moss-landing-fire/63468705a

NBC BAY AREA Watch 24/7

TRENDING SF Pride guide | Pride is Universal | Birthright citizenship | Illegal fireworks sei...

INVESTIGATIVE UNIT

Heavy metals found in dust miles from Moss Landing battery fire

By Jaxon Van Derbeken • Published March 25, 2025 • Updated on March 26, 2025 at 8:59 am

Log in or create a free profile to save articles

New data shows that heavy metals from the smoke at the Moss Landing battery storage plant fire have turned up miles away at levels that a toxicologist calls “concerning.” Jaxon Van Derbeken reports.

https://www.nbcbayarea.com/investigations/air-quality-data-moss-landing-battery-facility-fire/3828127/

How Aerosol Particles Impact Naval and DoD Activities:

Each one of these has its own development community focused on specific things

Operations, Safety of Navigation & Carrier Landings

Image Source: US Navy

Sensors and Target Acquisition

Image Source: Reid J./NRL

Plumes

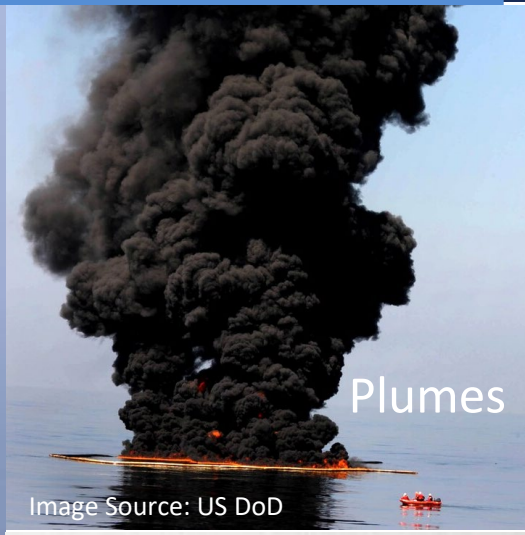


Image Source: US DoD

Obscurants

Image Source: US Navy

Intelligence, Surveillance & Reconnaissance

Image Source: NASA

Directed Energy

Image Source: US Navy

Chem/Bio

Image Source: US Army

A framework on how meteorology and data can help you make a decision

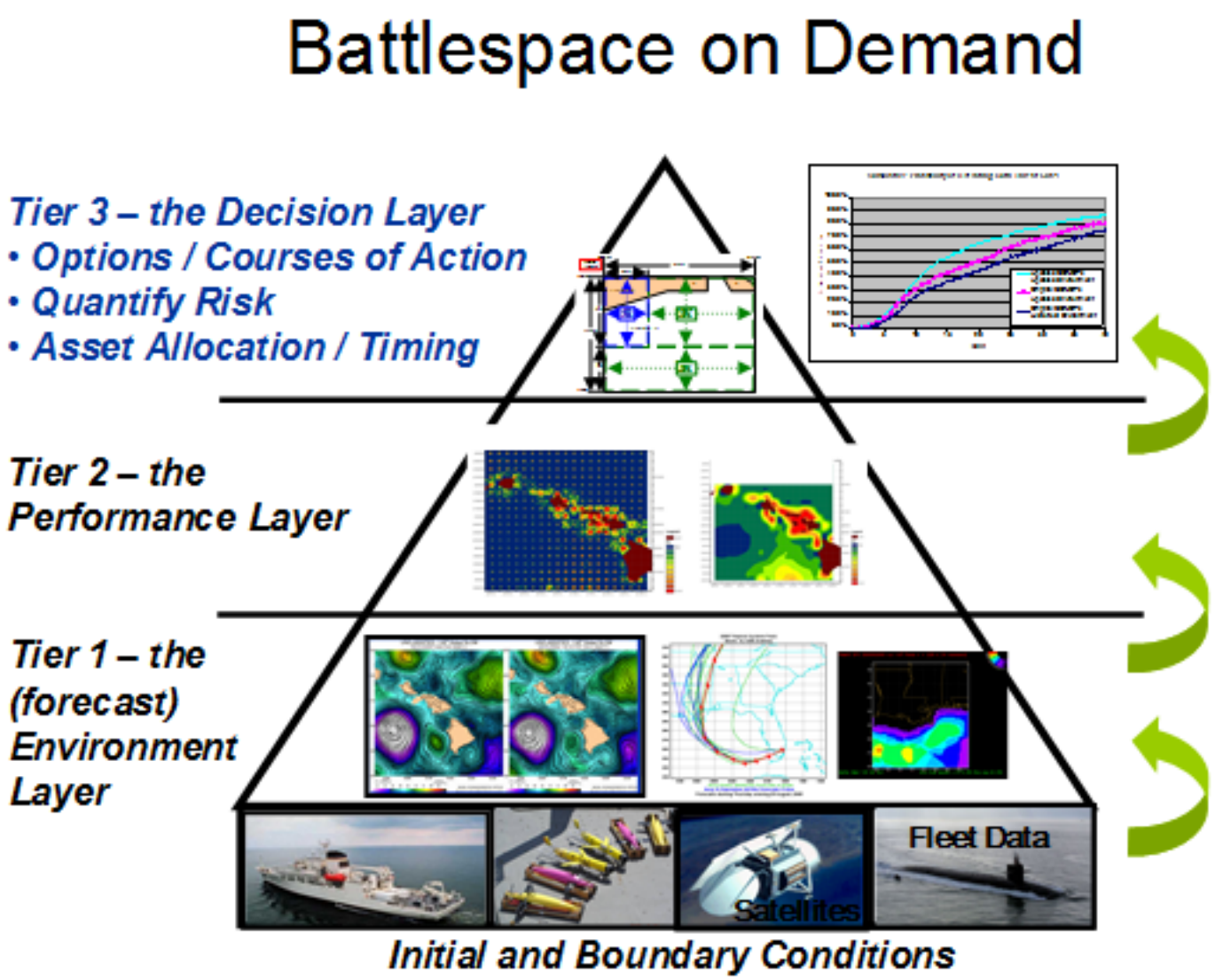
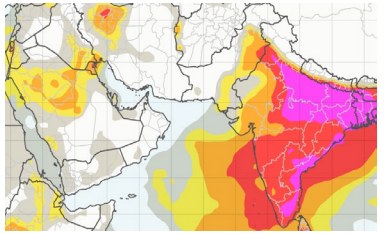


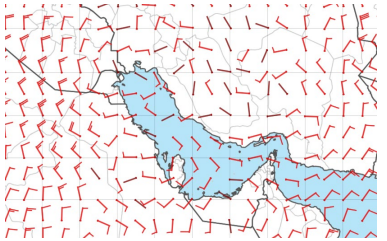
Image Source: US Navy/ RADM Titley (Ret)

Question: Is visibility going to be good enough for me to go into port in two days and resupply?

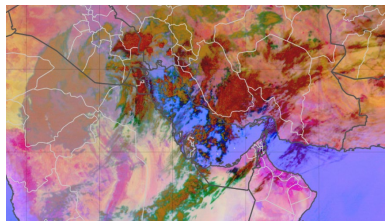
Consequence of a bad decision: Supply chains honked up. Can't have people waiting on the dock.



Best estimate of surface visibility and probabilities of reaching a threshold



NAAPS and COAMPS forecast of dust around the Persian Gulf driven by weather model output



Satellite observations of where storms are occurring now

Image Source: Reid J. /NRL

Ultimately, “Battlespace on Demand” is how the larger community operates for a diverse customer base for environmental data custom to order….

Boundary Conditions and Tier 1: Obs and Forecasts

Finish Meteorological Institute

Tier 2: Performance

Tier 2+3: Performance + Custom Decision Aids:

Plus to name a few:

General sensible weather:
AccuWeather, IBM/weather underground/Weather Channel/, Windy

Power industry:
3Tier, ESIG, Meteomativcds, SRECTrade

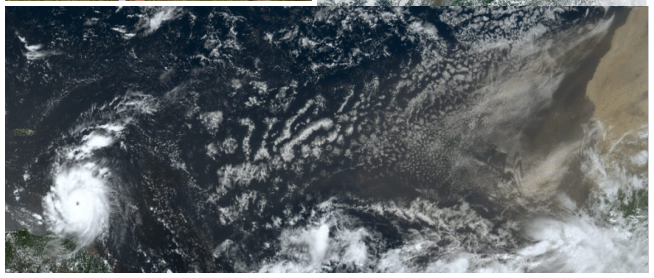
Routing
JEPPESEN, Weather Routing Inc, SOFAR

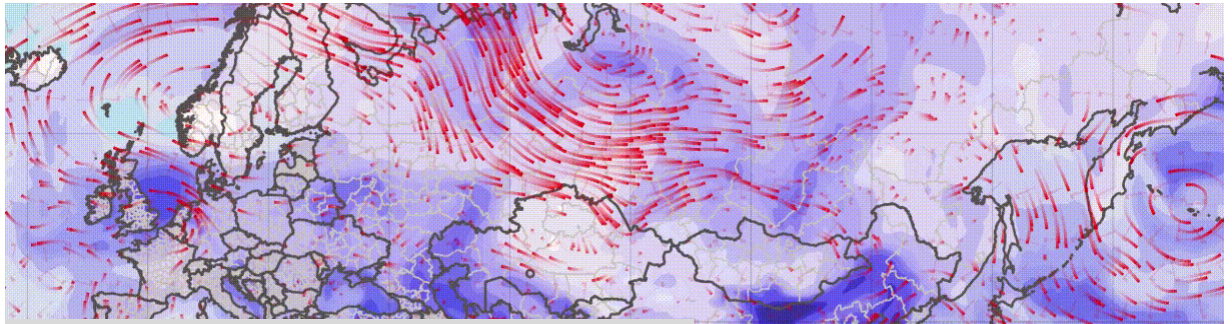
Getting a handle on scale: Examples of the Navy's framework

Obs and Data Assimilation

*What you can see by just looking?
The models need observations to keep them on track.
"Data Assimilation" is the process of combining the previous forecast with current observations to create a new starting point for the model.
Impactful observations include:

- *Weather balloons/dropsondes.
- *Satellite obs in the IR & microwave for temperature/moisture.
- *Tracking clouds for winds.
- *Radar, lidar, and microwave for precipitation and clouds.
- *Aircraft, ship, and ground obs.
- *Hand diagnosed storm centers.





Global Scale: NAVGEM, NEPTUNE, ESPC, NAAPS.

*Weather, ice, aerosols, connected to ocean
*Global Coverage from 1° to ~10-20 km
*5d Operational medium range guidance
10d Ensemble, Seasonal and Reanalysis
Literally a fluids model, but with sub grid parametrizations

Meso Scale: COAMPS, soon regional NEPTUNE

*Weather, ice, aerosol, ocean coupling.

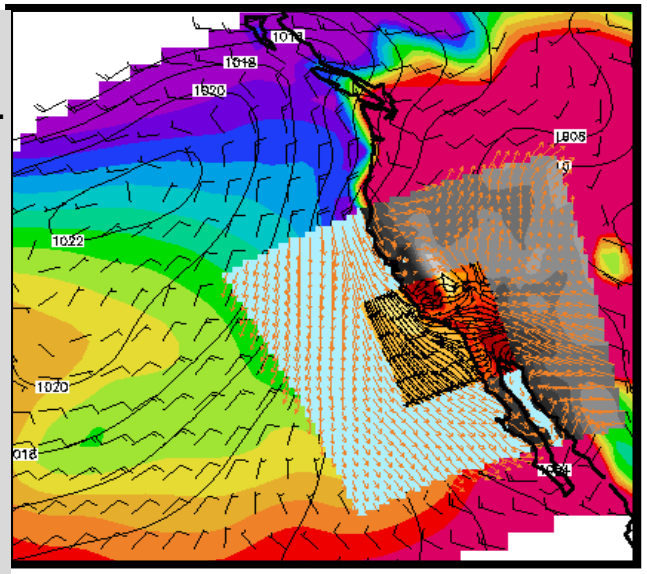
*Regional coverage, 1-18 km.

*Also fluids model but with more complete physics and a better ability to account for terrain and individual storms.

*Used operationally to forecast but also to support science and exercises.

*Run at a center but also locally on ships and even jeeps.

Image Sources: Reid J. /NRL



Data Fusion, Analysis, and Decision Aids.

Get back to making a decision.

NWP, satellite, and obs feed a host of decision aids that do calculations and visualization.

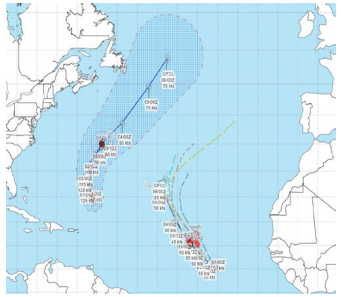
Examples:
Automated Tropical Cyclone Forecasting system-Will I get hit by a hurricane?

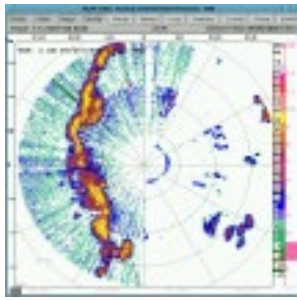
Joint METOC Viewer/Ladder: Ship rout.

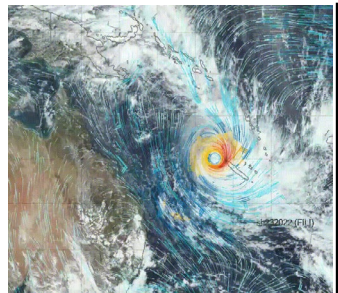
Builder: Radar propagation.

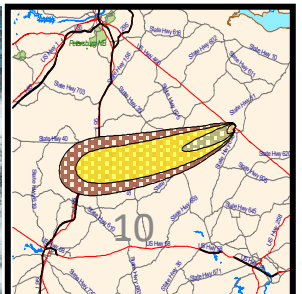
Host of EO TDAS: See and be seen, power in bucket, etc.

Dispersion









And that was just NRL. Here is the tip of the iceberg of meteorology and aerosol data that we often draw from

Network Observations: a) Airfields, ships and aircraft report visibility; b) Many counties publish near-real-time (NRT) air quality data; c) Personal federated networks such as Purple Air; d) cal/val and long-term monitoring sites.

Remote Sensing: US: NASA, NOAA, USDA; Europe: ESA & EUMETSAT; Asia: JMA/JAXA; KMA; ISRO; and yes CMA.

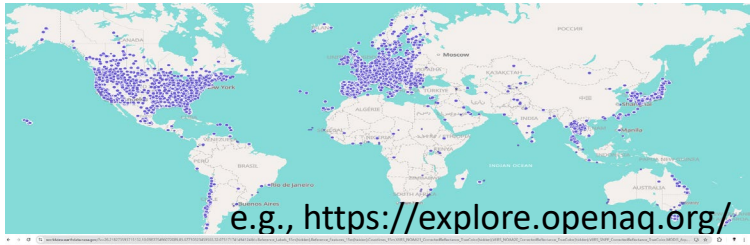
Global Modeling: Nine international centers perform operational or consistent NRT global aerosol forecasting and make their products public. Lots more keep their products internal.

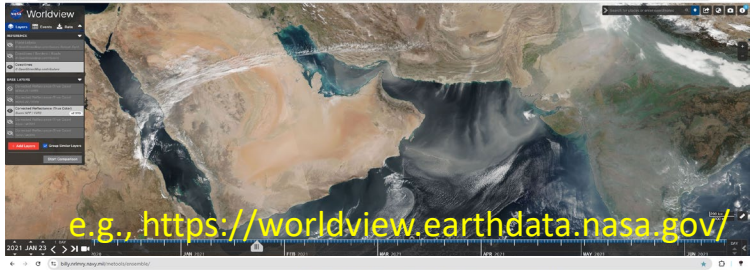
1)Barcelona Super Computing Center: Dust storms; **2)Copernicus/ECMWF/Europe:** Air quality and climate **3)Finish Met Service:** Air quality and biomass burning; **4)Japan Meteorological Agency:** Air quality and dust storms **5)MeteoFrance:** Air Quality; **6)NASA GMAO: Science and campaign support**; **7)NOAA**: Air Quality; **8)UK Met office**: Dust forecasting for MoD; **9) US Navy**: Operations.

Finer mesoscale (domains that cover more than 300 km): I lost count.

Plume Dispersion: I don't even try and keep count.

Field campaigns: Dozens per year with focus on specific products.





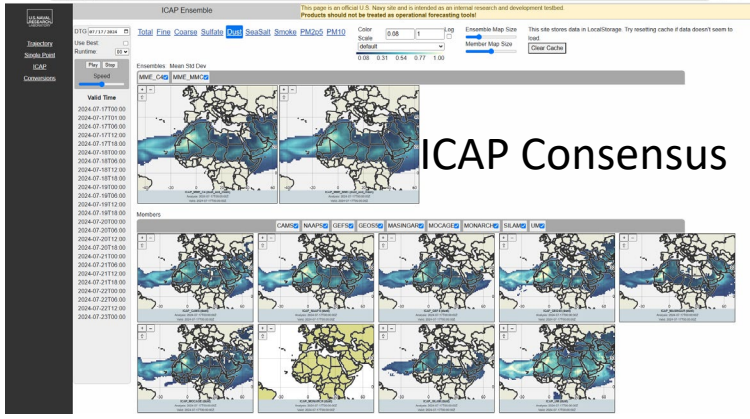
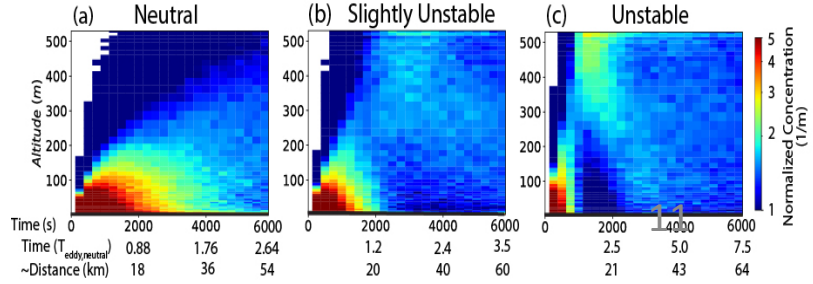


Image Source: Reid J. /NRL



Numerical Weather Prediction Observations

- Most often we think of observations as being made by in situ
  - Weather stations
  - Weather balloons
  - Airline/aircraft AMDAR
  - Voluntary Marine Observations (VOS)
  - Observer Reports/PIREPS
- In actuality, the vast majority of weather observations are from remote sensing. Each has its own strengths and weakness. E.g.,
  - Cloud tracking for winds.
  - Infrared for cloud top heights and temperature.
  - Infrared sounding for temperature & water vapor.
  - Microwave floor clouds and water vapor.
  - Radar, precip and scatterometer/SAR surface winds.
  - Radio/GPS occultation for humidity.
  - Constituents, including ozone and aerosol.


Image Source: NOAA


Image Source: US Navy

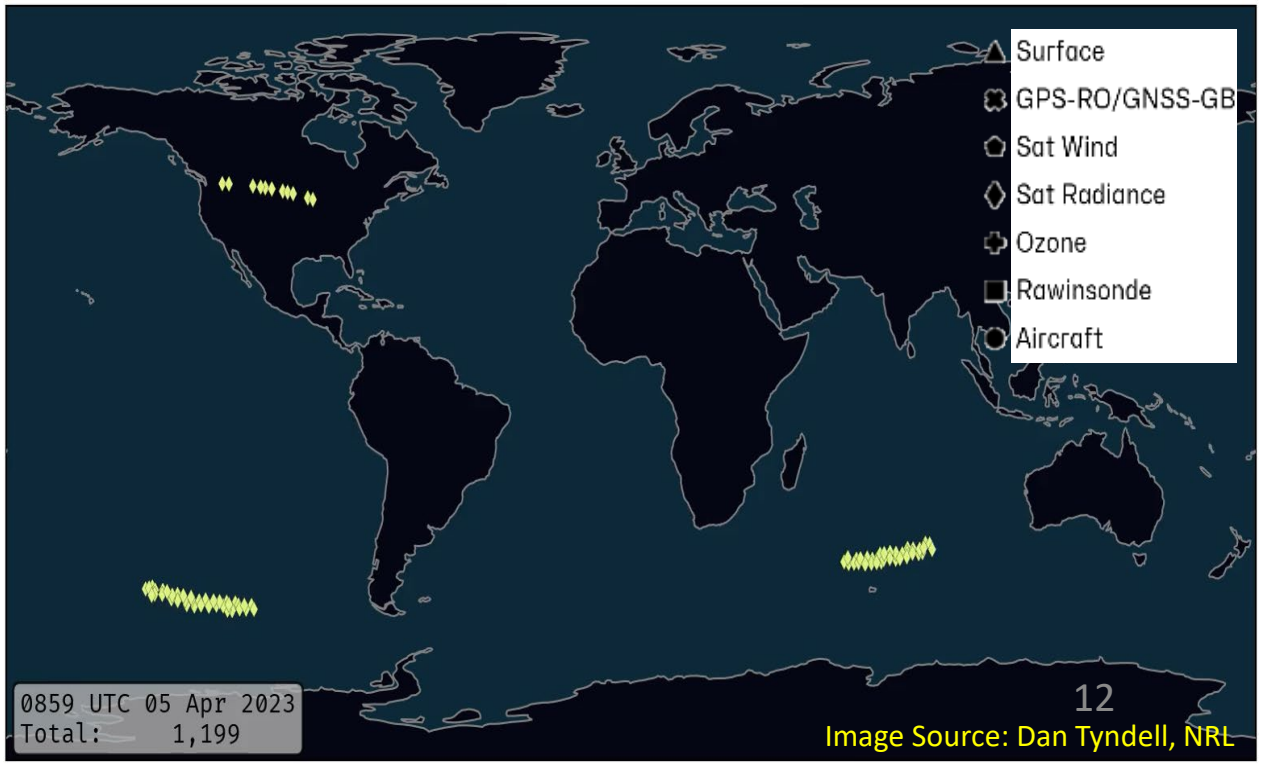
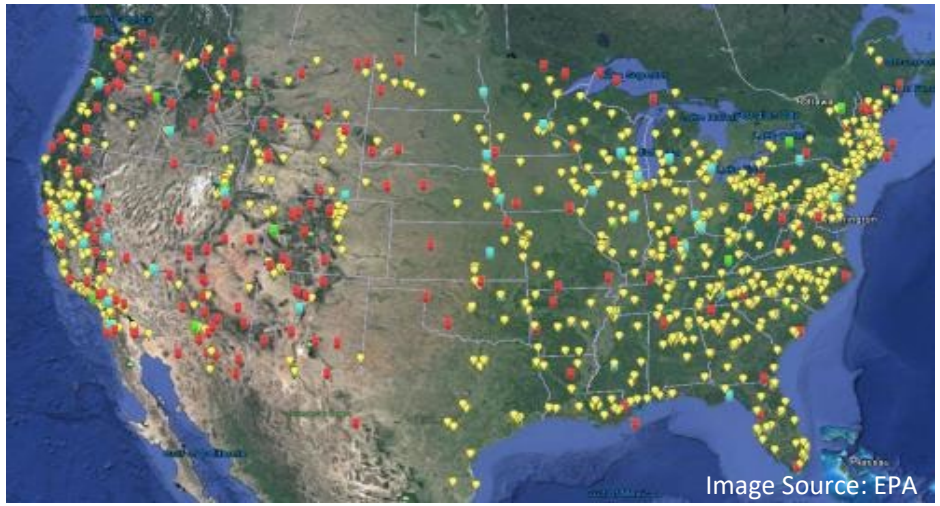
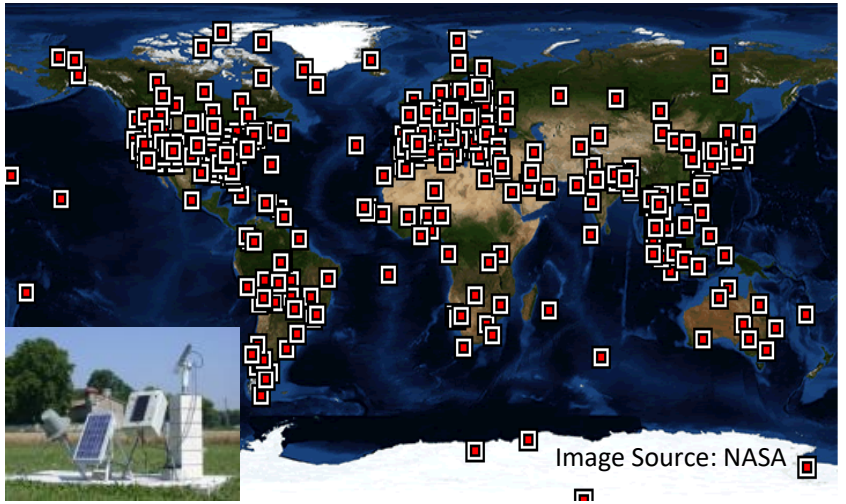


Image Source: Dan Tyndell, NRL

Now how about atmospheric chemistry networks: Each has their strengths and weaknesses: Quality, coverage, period, aggregation, format, etc.



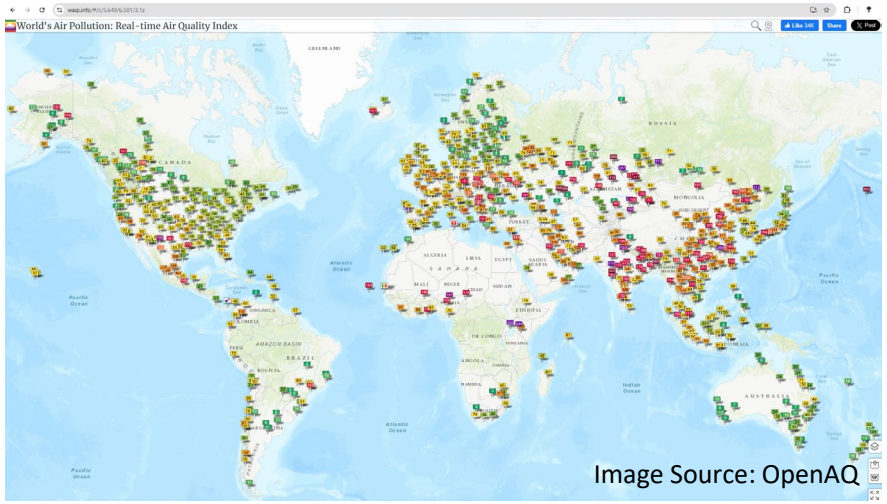
Regional air quality networks



NASA AERONET sun photometers

Image Source: WMO

GAW sites



Third party aggregators

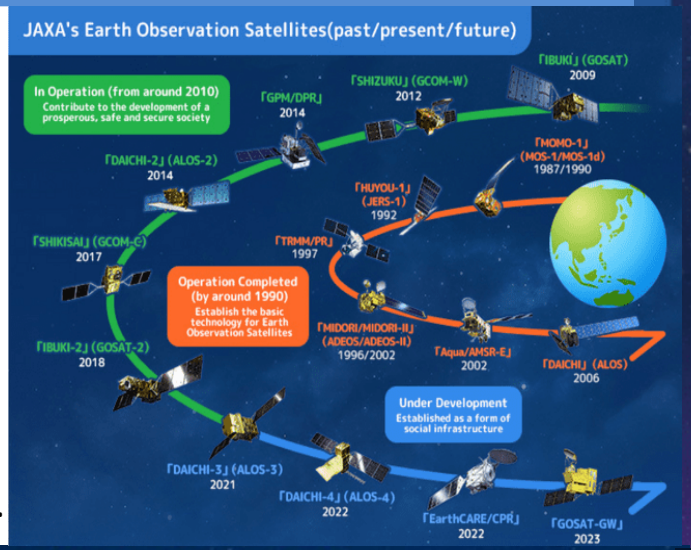
GAW Lidar Networks: GALION

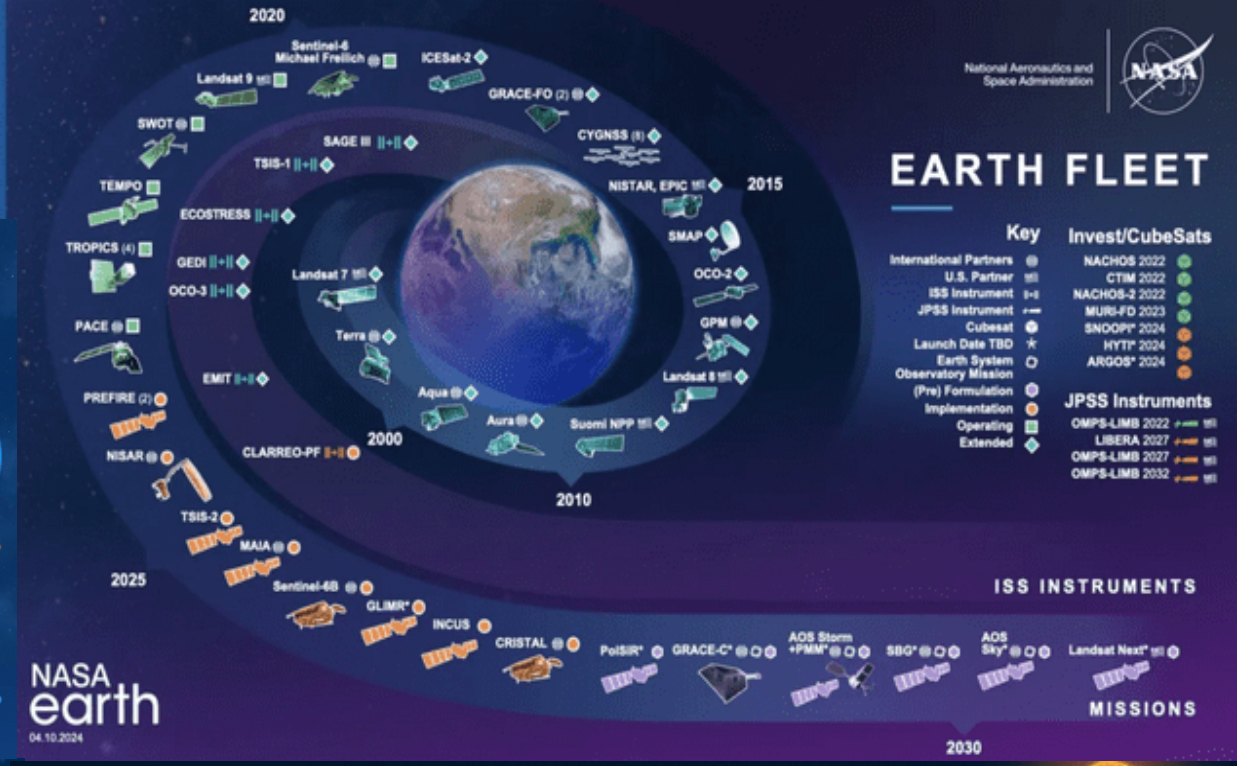
Ad hoc networks & field campaigns

The remote sensing busy board

Wow, there are a lot of satellite data streams to look after. Questions…

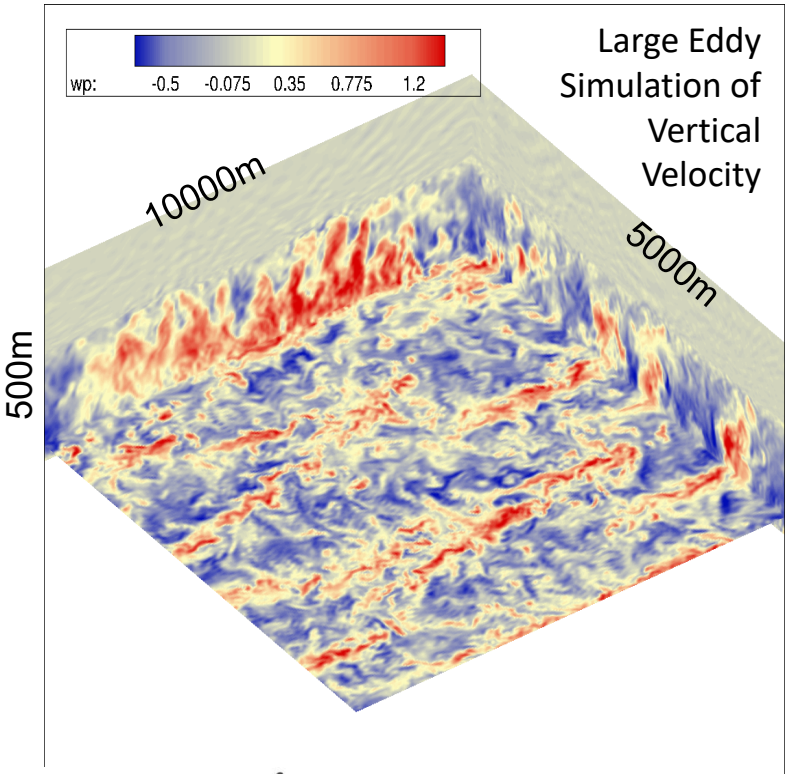
- So many. Sampling, skill, technique?
- And application? Visualization? Product? DA?
- What is your baseline?
- And how do we manage all of the data feeds?
- But who has the most impact on the power curve?
- Despite diversity, aerosol operations focuses on a few well characterized and timely datasets.





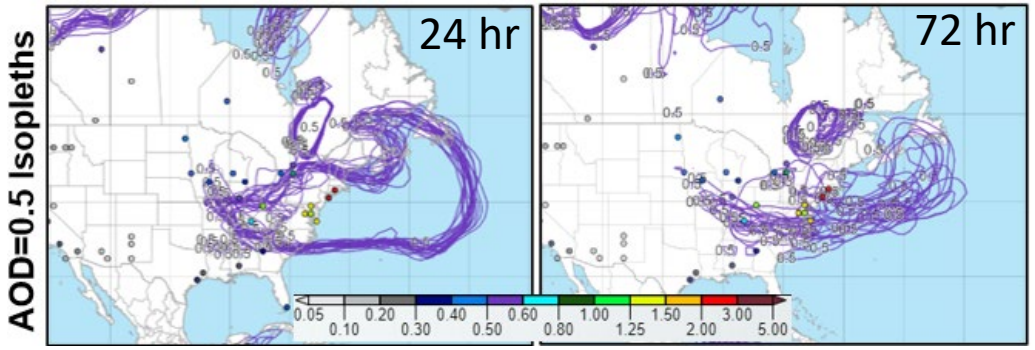
Flavors of Aerosol Modeling

- Where do you get your meteorology?
  - Inline versus offline versus type it in by hand?
  - Scales: Global, mesoscale (to 1 km), large eddy simulation (10-100 m), direct numerical simulation (m), analytic.
- How do you model your aerosol particles? Parametric versus Eulerian versus Lagrangian/trajectory.
- Do you need to know what happened or a forecast?
- How do you deal with uncertainty?
  - Deterministic versus ensemble, or faith?



Large Eddy Simulation of Vertical Velocity

Image Source: Rubin J/NRL



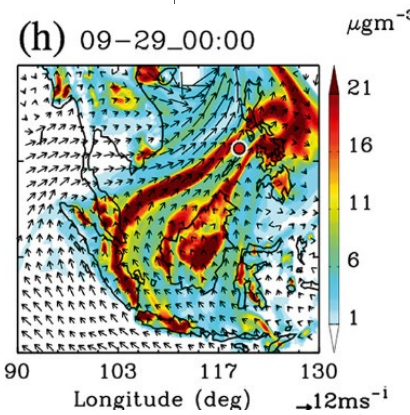


Image Source: Park & Reid /NRL

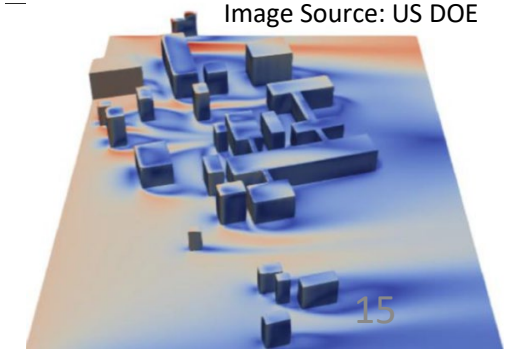
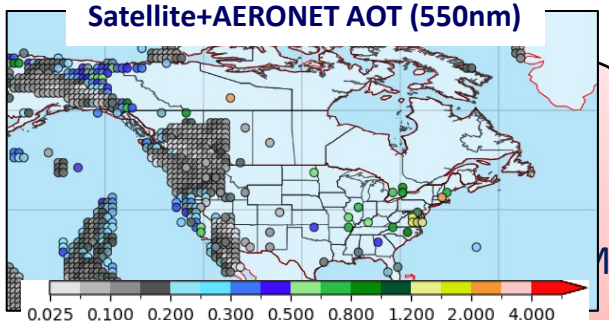


Image Source: US DOE

What it takes to use an observation in NWP:
Data Assimilation for aerosol particles, and anything else

Satellite+AERONET AOT (550nm)

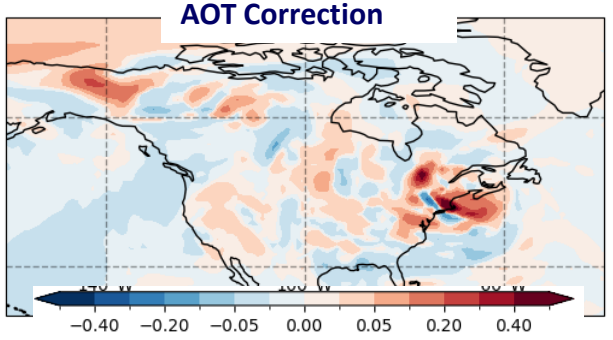


Obs Error Covariance: R

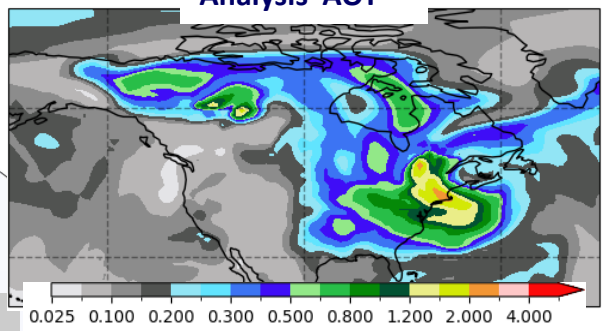
Prognostic Error Model for Level 3 Data (Table C2, Hyer et al. 2011)

	Global	NA Boreal	E. CONUS	W. CONUS	Cent. Am.	N. SA	S. SA	Australia
Terra CLIM	$0.06\|0.02+0.19\tau$	$0.05\|0.06+0.10\tau$	$0.04\|0.00++0.22\tau$	$0.07\|0.07+0.20\tau$	$0.06\|0.07+0.11\tau$	$0.05\|0.02+0.16\tau$	$0.066\|0.11+0.01\tau$	$0.04\|0.05+0.11\tau$
Aqua CLIM	$0.06\|0.02+0.19\tau$	$0.05\|0.06+0.11\tau$	$0.05\|0.02+0.17\tau$	$0.06\|0.04+0.44\tau$	$0.06\|0.03+0.17\tau$	$0.05\|0.02+0.14\tau$	$0.06\|0.06+0.22\tau$	$0.04\|-0.03+0.42\tau$
	S. Africa	Eq. Africa	N. Africa	S. Europe	Eurasian Boreal	East Asia Mid-Lat.	Penin. SE Asia Subcontinent	Indian
Terra CLIM	$0.05\|0.02+0.15\tau$	N/A	$0.09\|0.6+0.14\tau$	$0.05\|-0.01+0.26\tau$	$0.04\|0.01+0.15\tau$	$0.10\|0.02+0.22\tau$	$0.08\|0.02+0.19\tau$	$0.22\|0.05+0.13\tau$
Aqua CLIM	$0.05\|0.02+0.15\tau$	$0.05\|-0.01+0.20\tau$	$0.09\|-0.01+0.29\tau$	$0.06\|-0.01+0.29\tau$	$0.04\|0.03+0.20\tau$	$0.10\|0.03+0.20\tau$	$0.09\|0.06+0.11\tau$	$0.22\|0.04+0.17\tau$

AOT Correction

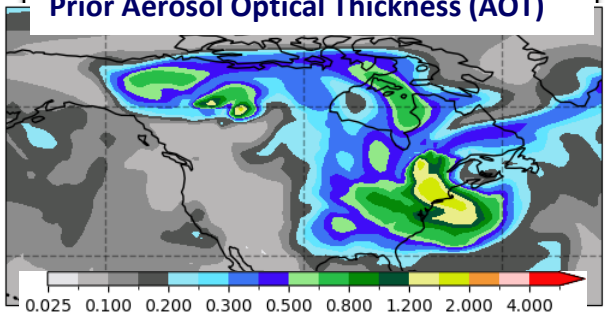


Analysis AOT



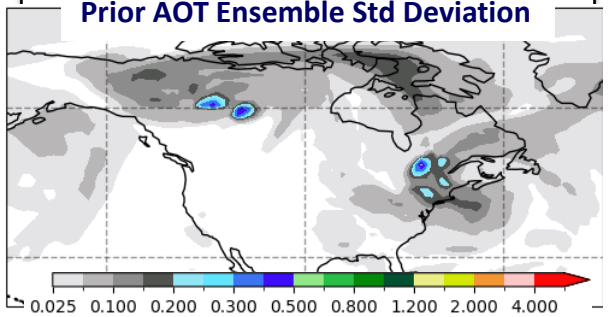
Model Forecast: x^b

Prior Aerosol Optical Thickness (AOT)



Model Error Covariance: B

Prior AOT Ensemble Std Deviation



Data Assimilation Correction:

$$x_a = x_b + K[y_o - Hx_b)],$$
$$K = BH^T(R + HBH^T)^{-1}$$

K=Kalman gain matrix

H=obs operator, maps model to obs space

$y_o - Hx_b$=Innovation (QC, Bias Correction)

BH^T=model error in obs space

$(R + HBH^T)^{-1}$=total error

Solver: 3DVar, 4DVar, Ensemble, Hybrid

Analysis: x^a

Used to initialize the next forecast.

Image Sources: Rubin J/NRL

Take a deep breath: Where do I start? Define your problem set clearly.

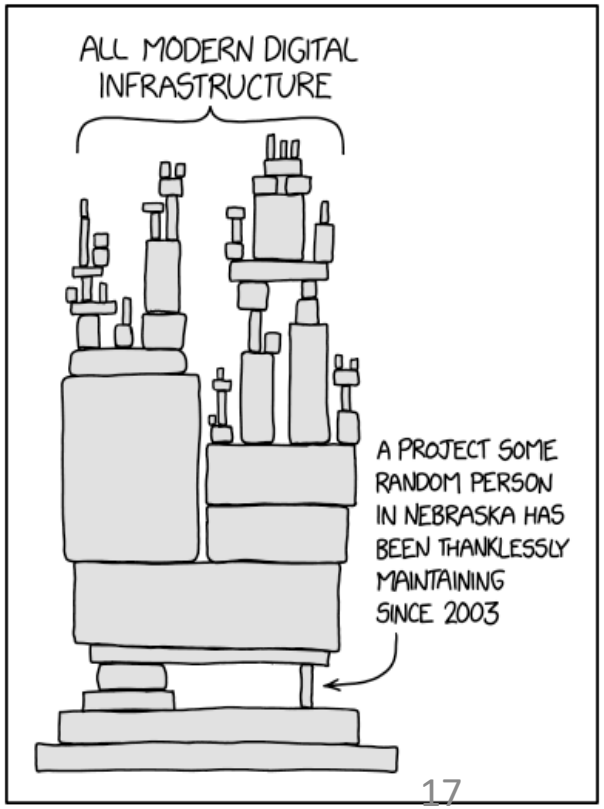
Image Sources: Reid J/NRL

- First, what is it you *really* need to know?
  - Plume position?
  - Climatology versus where it will be in 2 days?
  - Deposition/exposure?
- Second, how well do you *really* need to know it? What if I'm wrong? When is a bad answer better or worse than no answer?
  - Qualitative: What just hit me? Or will I get hammered tomorrow?
  - Quantitative: Will aerosol particles impact clouds in my forecast?
  - What is the difference?

    Even "flagship" products have errors.

    Just because someone posts a product doesn't mean its right.

    Precision is not the same as accuracy.
- Third, are there any resource requirements?
  - Do I know what it means to use a product? Can I even get there from here?
  - Do I have the time/capability, hard disk space, and need to dig into the data?
  - Do I know of any upstream dependencies?
  - What happens if you need a forecast and the product doesn't arrive?



https://xkcd.com/2347/

Concepts of data levels and archiving: Following NASA's lead for satellite data

- What level of data do you need? I need…
  - Unprocessed Root data: Level 0
  - Calibration corrections, annotation, georeferencing: Level -1a, b, c
  - Secondary calculations of geophysical data: Level 2
  - Gridded/averaged and systematically timed data: Level 3
  - Consistent and contiguous model-based products that are based on level 1 through 3: Level 4
  - I'm ok with an equation.
- Level 1 and 2 Data: go to the Distributed Active Archive Centers (DAACS)
  - E.g., https://ladsweb.modaps.eosdis.nasa.gov/; or near real time: https://lance.modaps.eosdis.nasa.gov/
- Level 3: Can go to the DAACs but also web portals: https://giovanni.gsfc.nasa.gov/giovanni/
- Level 4 and Beyond: https://gmao.gsfc.nasa.gov/GMAO_products/
- Just need a pretty picture? https://worldview.earthdata.nasa.gov/
- Equation or synopsis? Library….

Good starting places for public data repositories

There are lots of data archive, and they sort of play by the same rules. But each is nuanced. Here are a few :

- NASA
  - GMAO: https://gmao.gsfc.nasa.gov/GMAO_products/
  - LAADS: https://ladsweb.modaps.eosdis.nasa.gov/
  - Worldview: https://worldview.earthdata.nasa.gov/
- NOAA:
  - Data pages: https://www.star.nesdis.noaa.gov/jpss/aerosols.php
  - Imagery Aerosol Watch: https://www.star.nesdis.noaa.gov/smcd/spb/aq/AerosolWatch/
  - HYSPLIT trajectory: https://www.ready.noaa.gov/HYSPLIT.php

- BSC SDS Was dust forecasting: https://dust.aemet.es/
- ECMWF/Copernicus: https://cds.climate.copernicus.eu/
- EUMETSAT: https://view.eumetsat.int/
- JMA P-Tree: https://www.eorc.jaxa.jp/ptree/

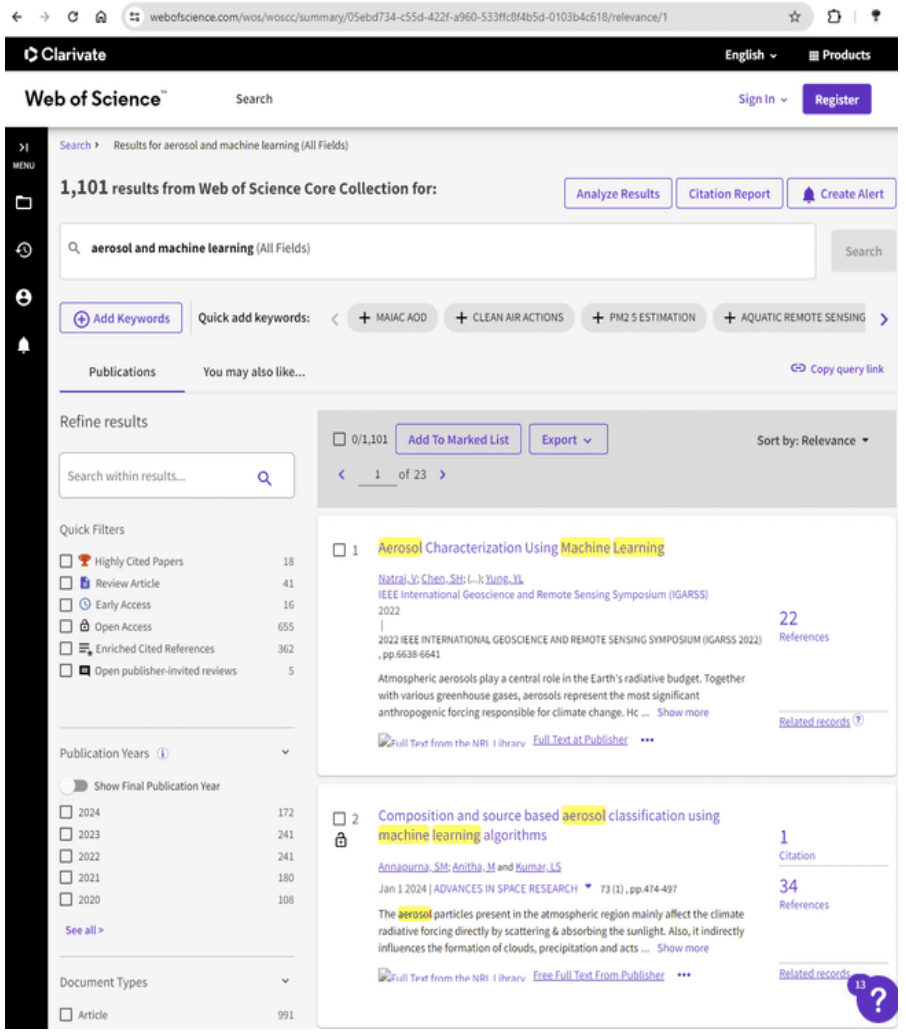
Navy archives and visualization resources.

- US Navy DISTRO-A global model and oceanography data: https://usgodae.org/
- Operational Navy weather and oceanography products (CaC): https://portal.fnmoc.navy.mil/
- Experimental visualization and lowcode (looking for alpha and beta testers):
  - Distro A site- https://maproom.nrlmry.navy.mil:8443/
  - Distro D site- https://maproom.nrlmry.navy.mil

And feel free to drop me a line: jeffrey.s.reid20.civ@us.navy.mil

AI and cloud computing: Can they help with the data vastness problem? Yes, sort of.

- AI processing and cloud computing are often in the same presentation/paragraph. And AI platforms and large language models are popping up like mushrooms.
  - Good at harvesting and communicating qualitative information.
  - Getting there for reformatting data and analytics.
  - Still poor at equations, physics, and abstractions.
- Scientists have been begging data centers for access to do their processing local to the data. And AI/ML methods are data heavy.
- Well, this future is here, and hard questions on what, where's, and how's of cloud computing and with it AI efforts.
  - Centers pay for data storage, who really pays for processing?
  - Account management.
  - Before you bought your machine on your grant, how does computer access change or go away with grant cycles?
- AI/ML is more "individual product-based" whereas science is more physics-based. Where is weather prediction? Depends on your point of view, but more towards product-based but AI is starting to get physics constraints.
- My concern, we have a hard enough time keeping up with traditional flagship products, what about all of AI/ML-based products?
- Development is not necessarily good or bad, but different. Nevertheless, we need to push for rigor. It's easy to make your AI product look like your training dataset. But what about independent verification? And who is responsible?



Industry in observations

- Agencies are experimenting with shifting from government collects to industry data purchase programs. These industry products are often AI based.
- There is an argument to be made against-many not to be politely named.
- But lets be honest, there is an argument in favor regarding efficiency and innovation. Efficient competition is good, especially if you can play by some reasonable ground rules.
- There are many commercial development examples that would never happen in a lab or center.
- And BTW, our retrievals and models are not as physical as the community likes to profess.
- It is not a zero-sum game. Market capitalization is growing rapidly for custom processing and products.
- And hey, we have to partner with industry to build anyway.
- My big concerns, and they are case by case:
  - Lack of knowledge or rigor of the science and ultimately user needs.
  - Black box processing and overfitting, especially in AI.
  - Consistency/calibration/verification and misleading scorecards.
  - Unfair business practices and contracting clauses.
  - Community feedback can become proprietary.
  - Longevity.

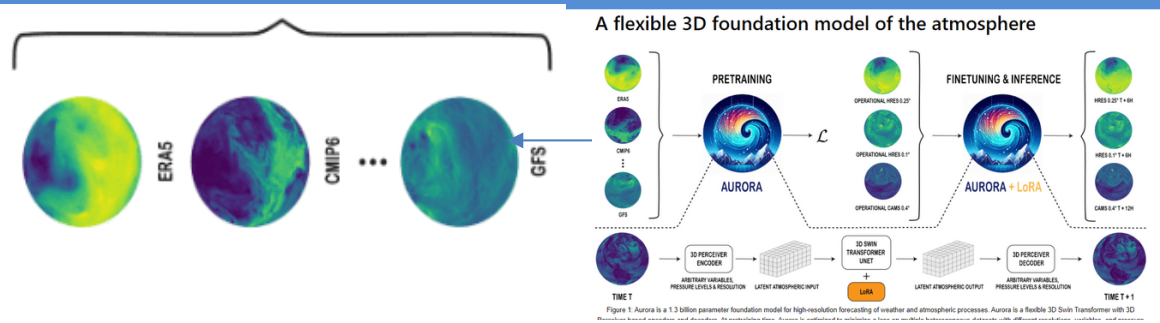
$6k for report

Industry in prediction

Microsoft Aurora: Bruinsma et al.,

https://www.microsoft.com/en-us/research/blog/introducing-aurora-the-first-large-scale-foundation-model-of-the-atmosphere/

- Industry also has a rapidly increasing market capitalization for commoditized NWP and custom weather/climate impact products.
- Energy traders have been successful at the ground level. Why? Lots of money at stake. Even personal subscriptions to weather underground etc.
- How about air quality, to what extent is there value added above center products? Certainly regulatory.
- Oh, and all of these model products are based on government furnished data streams. What if they dry up?
- And while they scorecard well, their outputs are sort of weird. Like observations, need for rigor plus consistency/calibration/verification.
- Predicting for isolated products is much easier than if you need consistency across variables?
- So, I am not suggesting we should be dismissive. Almost certainly there is adaptable technology, especially for debiasing prediction, which largely is what they are doing.
- Ultimately will industry dominate? Well time will tell, as will our ability to generate better baselines. But the need for basis sets is not going away. But they are squeezed.



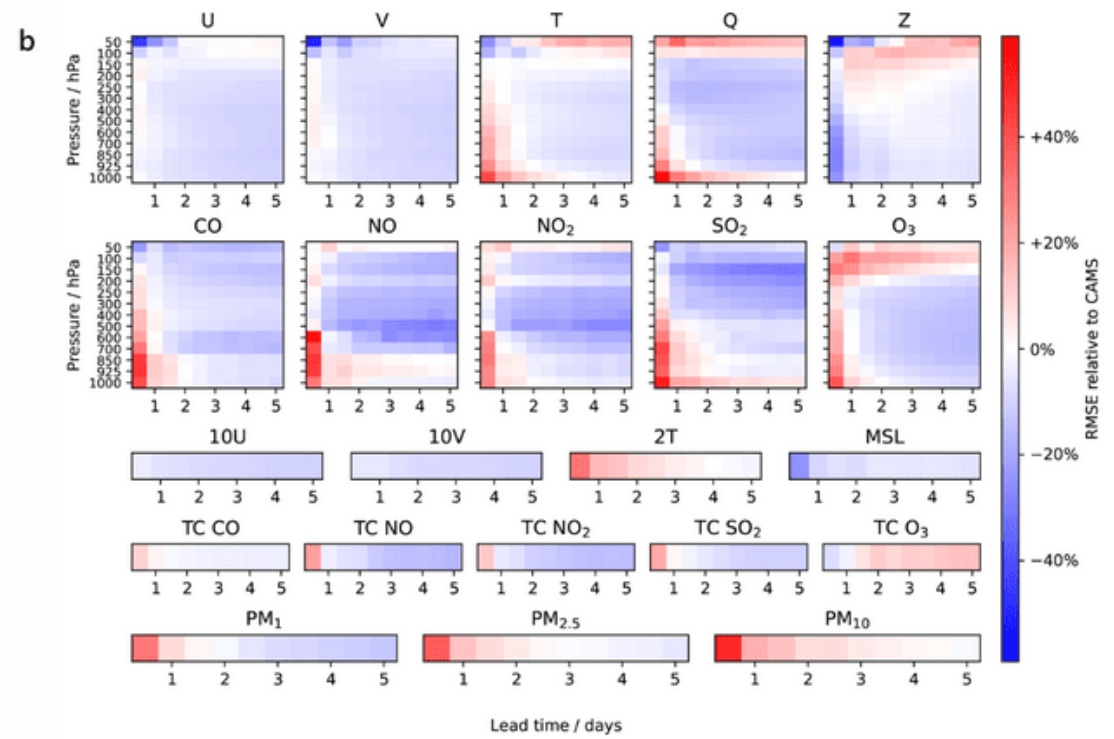


Figure 2: Aurora outperforms operational CAMS across many targets. (a) Sample predictions for total column nitrogen dioxide by Aurora compared to CAMS analysis. Aurora was initialized with CAMS analysis at 1 Sep 2022 00 UTC. Predicting atmospheric gases correctly is extremely challenging due to their spatially heterogeneous nature. In particular, nitrogen dioxide, like most variables in CAMS, is skewed toward high values in areas with large anthropogenic emissions, such as densely populated areas in East Asia. In addition, it exhibits a strong diurnal cycle; e.g., sunlight reduces background levels via a process called photolysis. Aurora accurately captures both the extremes and background levels. (b) Latitude-weighted root mean square error (RMSE) of Aurora relative to CAMS, where negative values (blue) mean that Aurora is better. The RMSEs are computed over the period Jun 2022 to Nov 2022 inclusive. Aurora matches or outperforms CAMS on 74% of the targets.

A final word on *Caveat Emptor and real-world trade spaces.*

Just because people say their data is from sophisticated sources and will do the job, doesn't mean it will.

- Aerosol (and all manner of environmental) prediction is a big big big field.
- There is pressure to sell, publish and declare success, and not much funding for independent evaluation.
- More complex is not necessarily better.
- Ask if the data producer uses it themselves for real world problems?
- Search for verification publications, the more independent the better.
- Do your own sniff test. Does it make sense? Does it look like satellite imagery or other like products? Does it even make sense to you?
- Talk to other users, and find out their own level of sophistication.

So how do you feel? It's ok, we have all been there.

- It's a lot for anyone to wrap their brain around, but the nature of data streams reflects the diverse nature of needs, funding, and competition.
- In time, you will come to grips with it all. But here are a few pointers:
  - a) Define your problem well.
  - b) Guage your need to error tolerances.
  - c) Simple is good, and know the rules before you decide to break them.
  - d) Use the library, DTIC and HDIAC
  - e) To keep up, partner.
  - f) I know this isn't Missouri, but "show me."
- Feel free to connect: jeffrey.s.reid20.civ@us.navy.mil

Image Sources: Getty Images

Bibliography

- Current state of global aerosol forecasting: https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3497
- Ensemble data assimilation: https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016JD026067
- Global data requirements: https://acp.copernicus.org/articles/18/10615/2018/
- Dispersion modeling: https://www.mdpi.com/2073-4433/13/10/1640
- Particle trajectories in the MABL: https://egusphere.copernicus.org/preprints/2025/egusphere-2025-576/
- Satellite skill of significant aerosol events: https://amt.copernicus.org/articles/16/2547/2023/