U.S. NAVAL RESEARCH LABORATORY

# Graph Theoretic Observation Thinning for Satellite Radiances

[a]William F. Campbell (presenter), [a]Hui Christophersen, [b]Christopher M. Hartman

*[a]U.S. Naval Research Laboratory (NRL), Marine Meteorology Division, Monterey, CA*

*[b]SAIC, Monterey, CA*



25th International TOVS Study Conference, Goa, India, May 8-14, 2025

U.S. NAVAL RESEARCH LABORATORY

# Satellite Data Usage

## Why don't we use all of the available satellite data?

- Computational cost of pre-processing and QC
- Ensemble and variational solvers may take too long or even fail to converge
- Spatially correlated observation error
  - A function of how close two similar obs are in space and time
  - Properly accounting for spatially correlated error is an open research question for most data types
  - If ignored, data are treated as if their errors are independent, degrading the analysis
  - We need to select data that are far enough apart to be truly independent, free from correlated observation error

U.S. NAVAL RESEARCH LABORATORY

# Flaws in Current Practice

## How do operational/research centers select data?

- Choose a fixed, approximately equispaced grid on Earth's surface
- Then, for each instrument at each gridpoint,
- Choose the single observation closest to that gridpoint
  - Absent further constraints, the MEAN distance between observations is equal to the grid spacing; however,
  - The MINIMUM distance is unconstrained, allowing correlated error
  - Additional constraints enforcing a larger minimum distance between obs locations leave data gaps
- Averaging (a.k.a. superobbing) instead is not the answer; averaging treats data as independent, baking in any spatially correlated error
- The FUNDAMENTAL problem is that observation selection needs to happen in observation space, not on an arbitrary grid

U.S. NAVAL RESEARCH LABORATORY

# A Problem from Graph Theory

- Given a set of geometric shapes (e.g. circles on a 2D surface), find the maximum independent set (MIS), which consists of the largest number of those shapes that do *not* overlap any other shape
- The MIS problem can be posed as a graph with a node for each shape, and edges between shapes that overlap
- There are known but very slow methods to find the optimal solution, and approximate methods to find good solutions quickly
- The MIS problem maps *directly* onto data selection:
  - Draw a circle at each observation location, with D reflecting the desired local observation density (smaller D => denser obs) – some circles will overlap
  - ***Note that these circles are an abstraction, NOT the satellite footprint***
  - Choose the largest possible set of non-overlapping circles (i.e. solve the MIS)
- The result will be the (nearly) largest set of observations that obey local and global density constraints, and are never closer than the smallest chosen D

U.S. NAVAL RESEARCH LABORATORY

# Graph Thinning Visualization

## Visualizing graph theoretic abstraction for data selection

After legacy thinning on fine grid

**New algorithm**

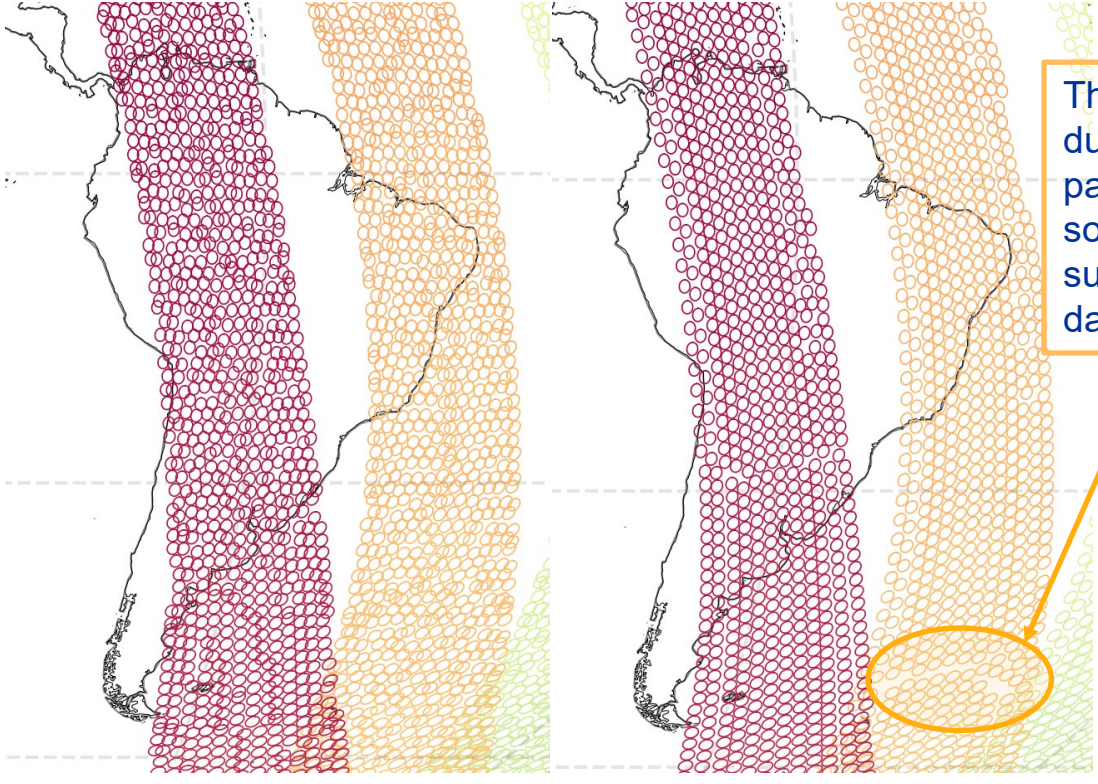
Each ob is a node (coin), two obs that might be spatially correlated share an edge (overlap), algorithm removes overlapping nodes to find the maximum disjoint set of obs

Final observations retained

U.S. NAVAL RESEARCH LABORATORY

# Grid Thinning vs Graph Thinning

SSMIS F18 20210715T1800Z



The gap is due to the parallel solve of subsets of data

U.S. NAVAL RESEARCH LABORATORY

# Experiments with SSMIS

- **Control run with operational grid thinning, experiment with graph thinning (SSMIS only), flat thinning at grid thinning size**
- **NAVGEM cycling DA runs using all operational conventional and satellite data from June 29 – Oct 2, 2022**
  - ***N.B.*** *grid pre-thinning enhances the speed of the graph thinning with little effect on the final result*

| Control | Graph_flat |
|---|---|
| Grid thinning for all radiances, **~41K** SSMIS profiles | Graph thinning for **SSMIS only**, **~32K** SSMIS profiles, 22% fewer |
| 1 ob per 140km box per 30 minutes | **Prethin:** 1 ob per 33 km box per 30 minutes |
| No diameter assigned; no graph thinning | D=140km for all SSMIS |

U.S. NAVAL RESEARCH LABORATORY

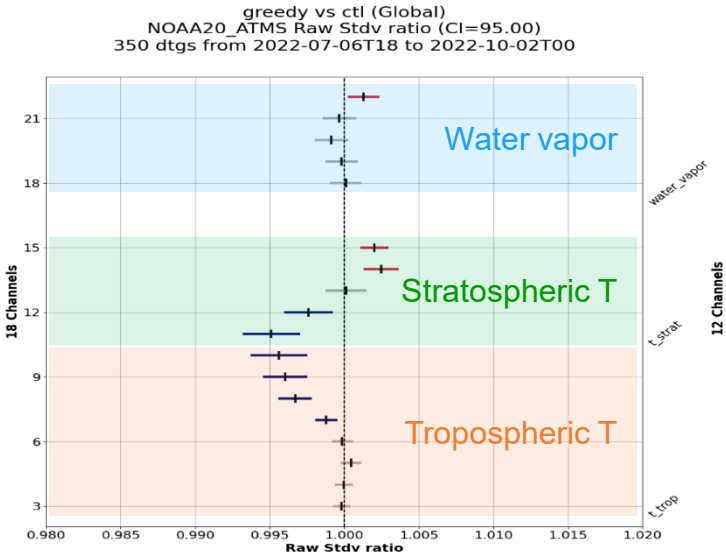
# Verification Metrics

- **Satellite vs simulated obs from short forecasts (aka fit2obs)**
  - Time series of ratio (between experiments) of spatial standard deviation of obs minus background for each channel
  - If the confidence interval (CI) contains the ratio 1.0, then the test cannot distinguish between the two experiments for that channel
- **Fit2obs categorical scorecard (subjective)**
  - Attempt to put similar channels in categories, and evaluate the category as a function of fit2obs wins, losses, and ties (ratio CI contains 1.0)
  - Still a work in progress; wlt_score is very conservative
- **Forecast verification against independent ECMWF analyses)**
  - Each panel shows pressure vs forecast lead time for a given variable (row) and a given region (column) for an experiment vs the control
  - Colors indicate percent improvement or degradation w.r.t. ECMWF analyses; hashing indicates 95% statistical significance

U.S. NAVAL RESEARCH LABORATORY

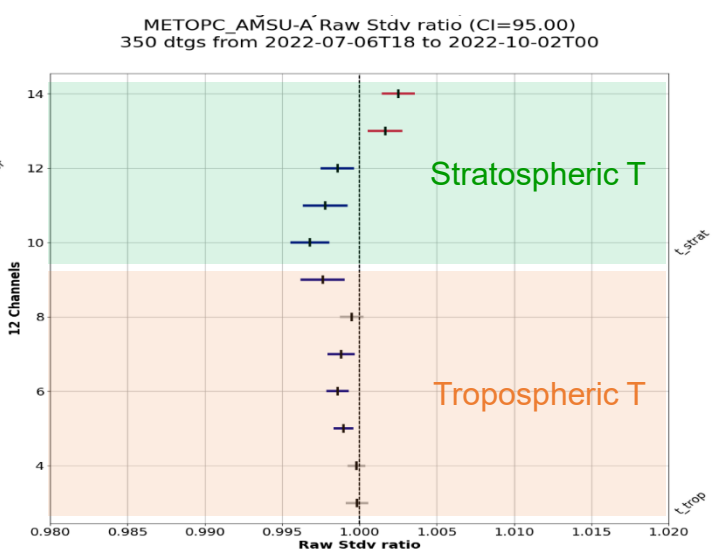
# Graph_flat vs Control

## NOAA20 ATMS




6 wins; 9 ties; 3 losses

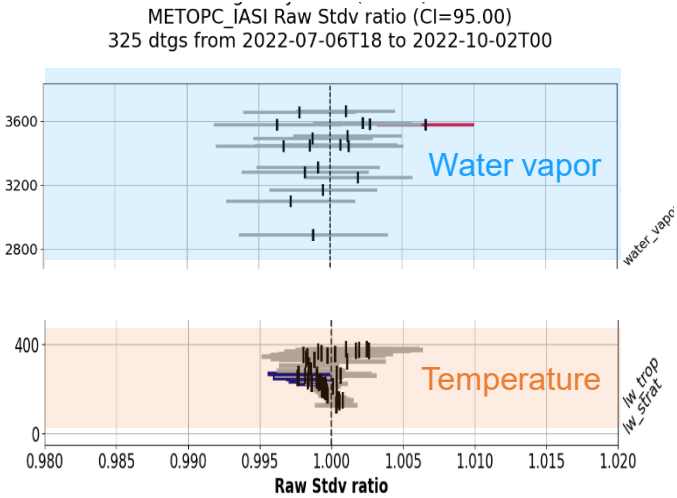
## METOPC AMSU-A




7 wins; 2 ties; 2 losses

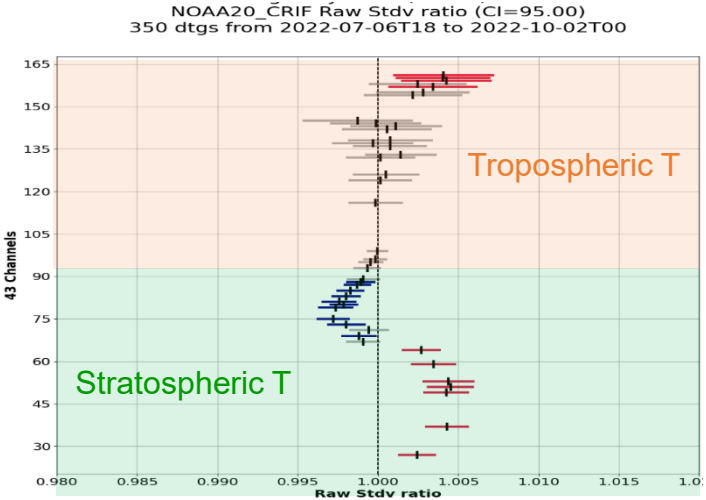
# Graph_flat vs Control

## METOPC IASI




6 wins; 61 ties; 1 loss

## NOAA20 CrIS FSR

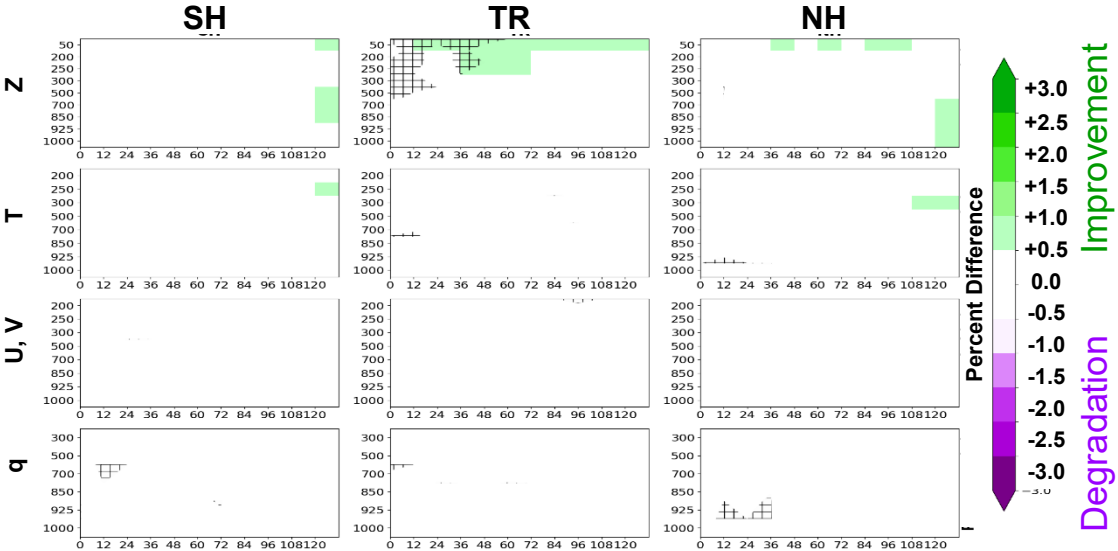



10 wins; 22 ties; 11 losses

U.S. NAVAL RESEARCH LABORATORY

# Graph_flat vs Control Stats

## Verification against ECMWF Analyses



U.S. NAVAL RESEARCH LABORATORY

# Graph Thinning Advantages

- The new method sets an INVIOLATE MINIMUM distance between two similar observations, while simultaneously retaining the maximum number of observations obeying that constraint
- It works in observation space, independent of any grid, and dynamic in space and time
- The freedom to assign each observation its own radius given by a density function, where the users expertise is implemented to influence which obs will be more or less likely to be retained, with a myriad of potential applications
- The density function is arbitrary: it can depend on obs error, other metadata such as orbital parameters, surface type, flow of the day, etc.
- The maximum density is constrained by the minimum safe distance to mitigate spatially correlated error; minimum density is chosen by the user
- Even if the density function is constant, there is a significant benefit, because we are restricting the MINIMUM obs distance rather than the MEAN obs distance

U.S. NAVAL RESEARCH LABORATORY

# Future Work

- **Apply the methodology to existing sensors of interest such as ATMS and CrIS FSR, and future sensors such as TROPICS, TEMPEST, and COWVR, starting with flat thinning in a NAVGEM context**
- **Expand the science by investigating application of novel density functions, for example targeting increased observation density near developing storms or hurricanes**
- **Investigate spatially correlated error with complete control over minimum distance between observations**
- **Anisotropic data thinning across frontal boundaries, geographic features**
- **Subset thinning for different channels at the same spatial location (e.g. greater density for moisture channels than temperature channels while retaining full channel profiles where possible)**
- **Weighted superobbing accounting for spatially correlated error may be more straightforward to address on a graph in observation space**

U.S. NAVAL RESEARCH LABORATORY

# Questions?

Email me @ william.f.campbell54.civ@us.navy.mil

Dhanyavaad.