1:15 PM ET // WED 13 JAN 2021
Joint Session 8.2 // 9th Symposium on Building a Weather-Ready Nation
101st Annual Meeting of the American Meteorological Society // Virtually Everywhere

From the Horse's Mouth

Mining tropical cyclone forecast discussions to recover patterns and biases in forecast decisions.

Presentation Github: github.com/cpwx/ams21

C. E. Powell // Office of Projects, Planning, and Analysis // NOAA/NESDIS
Z. Jelenak, P. S. Chang // Center for Satellite Applications and Research // NOAA/NESDIS

Agenda

How we got here and where we're going.

1. **Why bother?**
2. **What is this?**
3. **How does this work and lead to actionable data?**
4. **What conclusions can you draw?**

Agenda

How we got here and where we're going.

1. **Why bother?**
   We don't study forecaster decisions as much as the "big-ticket" data sources.

2. **What is this?**
   An exploratory dataset built from mined NHC forecast discussions + applications.

3. **How does this work and lead to actionable data?**
   We apply a classification model to text data, and analyze the relationships between words and other metadata.

4. **What conclusions can you draw?**
   Human forecasters make decisions, both deliberate and unconscious, about data quality, representativeness, and utility. Many of these decisions can be teased out by treating forecast discussions as research artifacts.

Why Bother?

Quantify the human element.

1. Humans are <u>critical</u> to the forecast

   - C.A. Doswell (1986)
     The human element in weather forecasting. NWS Digest.
   - Stuart et al. (2006)
     10.1175/BAMS-87-11-1497
   - Novak et al. (2008)
     10.1175/2008WAF2222142.1
   - Hoffman et al. (2017)
     10.2307/j.ctt1t88w2v

Why Bother?

Quantify the human element.

1. Humans are critical to the forecast
2. Forecast skill improvement is often priced in arbitrary or oblique metrics
   - Produces quantifiable metrics
   - A semblance of reproducibility
   - A sense of rigor

Observation Skill	What is it?	What does it really measure?
Observing System Simulation Experiment (OSSE)	Synthetic data to project value-add of observation	The value of a synthesized data source to a synthetic atmosphere
Observing System Experiment (OSE)	Data denial trial	What happens to a model forecast if you remove one dataset and nothing else
Forecast Sensitivity to Observations Impact (FSOI)	Linearized sensitivity to adjoint	How much a model forecast changes if you modify each observation input by a small
Model Skill	**What is it?**	**What does it really measure?**
500 mb geopotential height anomaly correlation	Correlation coefficient for deviations from model climatology geopotential height at 500 mb	How well a model places deviations of geopotential height at 500 mb
250 mb wind RMSE	Domain root mean square error for wind fields at 250 mb	How well a model generates forecasts of wind fields at 250 mb given a set domain in space/time
[…]	[…]	[…]

Why Bother?

Quantify the human element.

1. Humans are <u>critical</u> to the forecast
2. Forecast skill improvement is often priced in arbitrary or oblique metrics
   - Produces quantifiable metrics
   - A semblance of reproducibility
   - A sense of rigor

   <u>But</u>: *model skill is not forecast skill!*

Observation Skill	What is it?	What does it really measure?
Observing System Simulation Experiment (OSSE)	Synthetic data to project value-add of observation	The value of a synthesized data source to a synthetic atmosphere
Observing System Experiment (OSE)	Data denial trial	What happens to a model forecast if you remove one dataset and nothing else
Forecast Sensitivity to Observations Impact (FSOI)	Linearized sensitivity to adjoint	How much a model forecast changes if you modify each observation input by a small
Model Skill	**What is it?**	**What does it really measure?**
500 mb geopotential height anomaly correlation	Correlation coefficient for deviations from model climatology geopotential height at 500 mb	How well a model places deviations of geopotential height at 500 mb
250 mb wind RMSE	Domain root mean square error for wind fields at 250 mb	How well a model generates forecasts of wind fields at 250 mb given a set domain in space/time
[…]	[…]	[…]

Why Bother?

Quantify the human element.

1. Humans are <u>critical</u> to the forecast
2. Forecast skill improvement is often priced in arbitrary or oblique metrics
   - Produces quantifiable metrics
   - A semblance of reproducibility
   - A sense of rigor

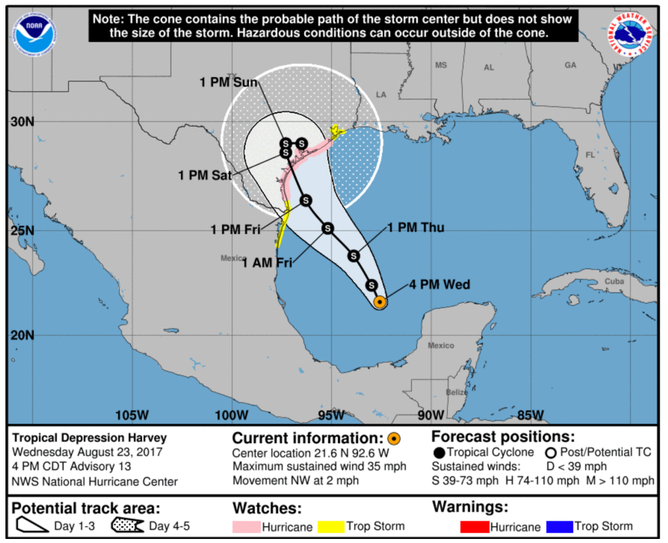
   <u>But</u>: *model skill is not forecast skill!*

<u>ANSWER THE TOUGH QUESTIONS</u>

1. *"How useful are these data?"*
2. *"What would make these data more valuable?"*
3. *"Why did you do _____ ?"*

What is this?

Forecast discussions as a dataset.




```
Tropical Depression Harvey Discussion Number  13
NWS National Hurricane Center Miami FL       AL092017
400 PM CDT Wed Aug 23 2017

High-resolution visible satellite images show that the cloud pattern
of Harvey is a little better organized than it was this morning, but
the system lacks distinct banding features.  Surface synoptic
observations, ASCAT data, and Dvorak classifications from TAFB and
SAB indicated that the cyclone has not strengthened, so the current
intensity is held at 30 kt.  The global models predict that an
upper-level low over the northwest Gulf of Mexico will essentially
dissipate in a day or so.  Therefore, Harvey is expected to remain
in a relatively low-shear environment up to the Texas coast.
Moreover, ocean analyses show that a warm eddy that broke off from
the Loop Current has drifted westward across the Gulf to a location
near the projected path of Harvey.  This would also be conducive to
strengthening, so it is likely that the system will become a
hurricane prior to landfall, although this is not explicitly shown
in the NHC forecast for which landfall is indicated between 48 and
72 hours.
```

```
Based on the scatterometer data and geostationary satellite fixes
the center hasn't moved much this afternoon, although recent imagery
suggests a northwestward drift at about 320/2 kt.  A weak mid-level
ridge to the northeast of Harvey should cause the cyclone to move
on a northwestward or north-northwestward track through 48 hours.
Later, steering currents weaken as a ridge builds over the
southwestern United States and a trough drops down from the Plains.
As a result, Harvey should decelerate while making landfall and
move very slowly just inland of the coast.  Some of the track
guidance models, such as the HWRF, have shifted southwestward in
comparison to their previous run.  The official track forecast is
very close to the previous one through 48 hours and is a little
slower and to the west after that time.  This is very close to the
latest dynamical model consensus, TVCN.  It should be noted that
synoptic surveillance data are currently being collected by the NOAA
G-IV jet aircraft and these data will be assimilated into, and
hopefully improve the forecasts by, the global models.
```

What is this?

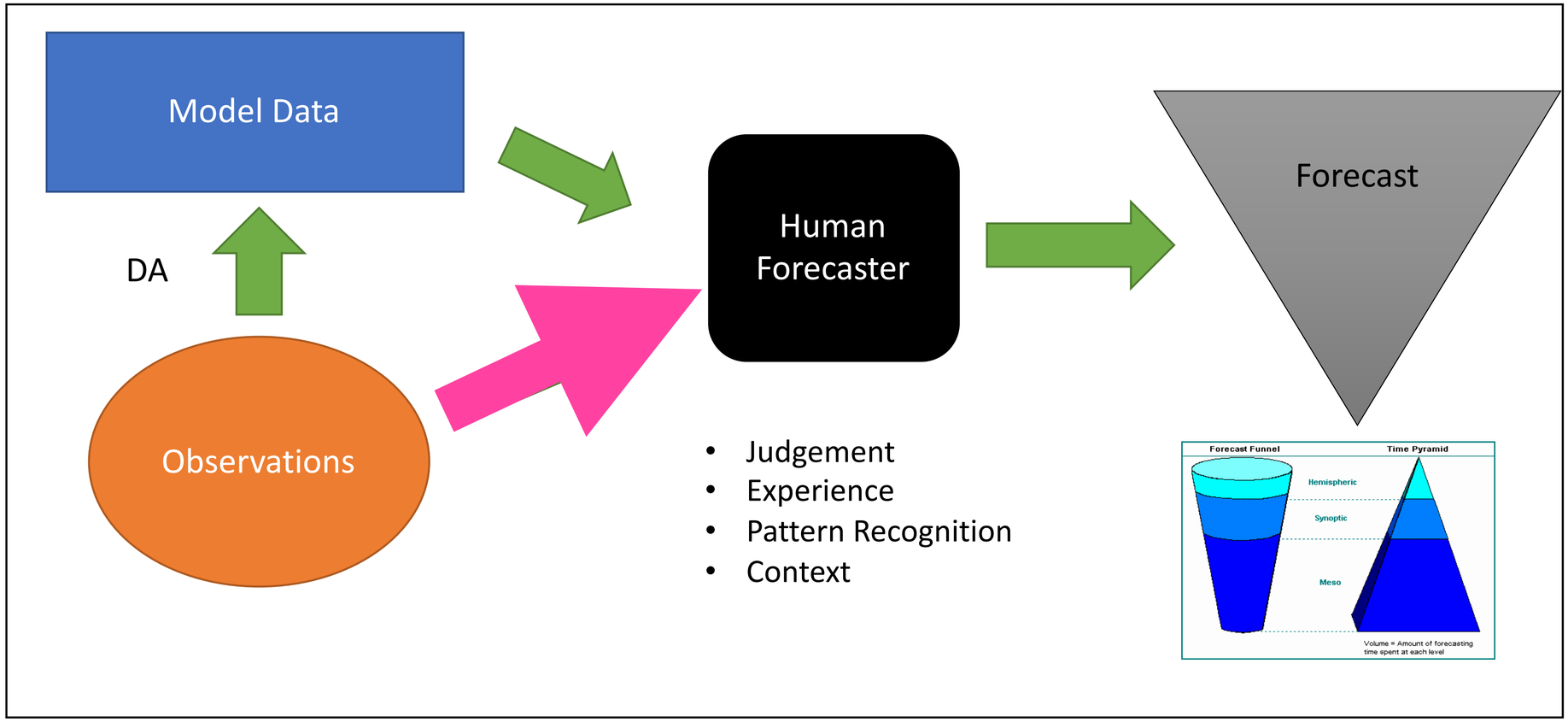
Forecast discussions as a dataset.

- **Justificatory**
  - Discusses motivations and considerations for a forecast decision
- **Paired with forecasts**
  - Means there's a forecast-justification matched pair
- **Precise, expert language**
  - Primary audience is other meteorologists, and assumes expert parsing
- **Metadata rich**
  - Contains storm + forecaster information
- **Information rich**
  - Decades of experience + terabytes of data summarized in a few words
- **Unstructured, but patterned**
  - No concrete style guide, but forecast discussions have very similar content

What is this?

Forecast discussions as a dataset.

The forecast value chain (simplified)



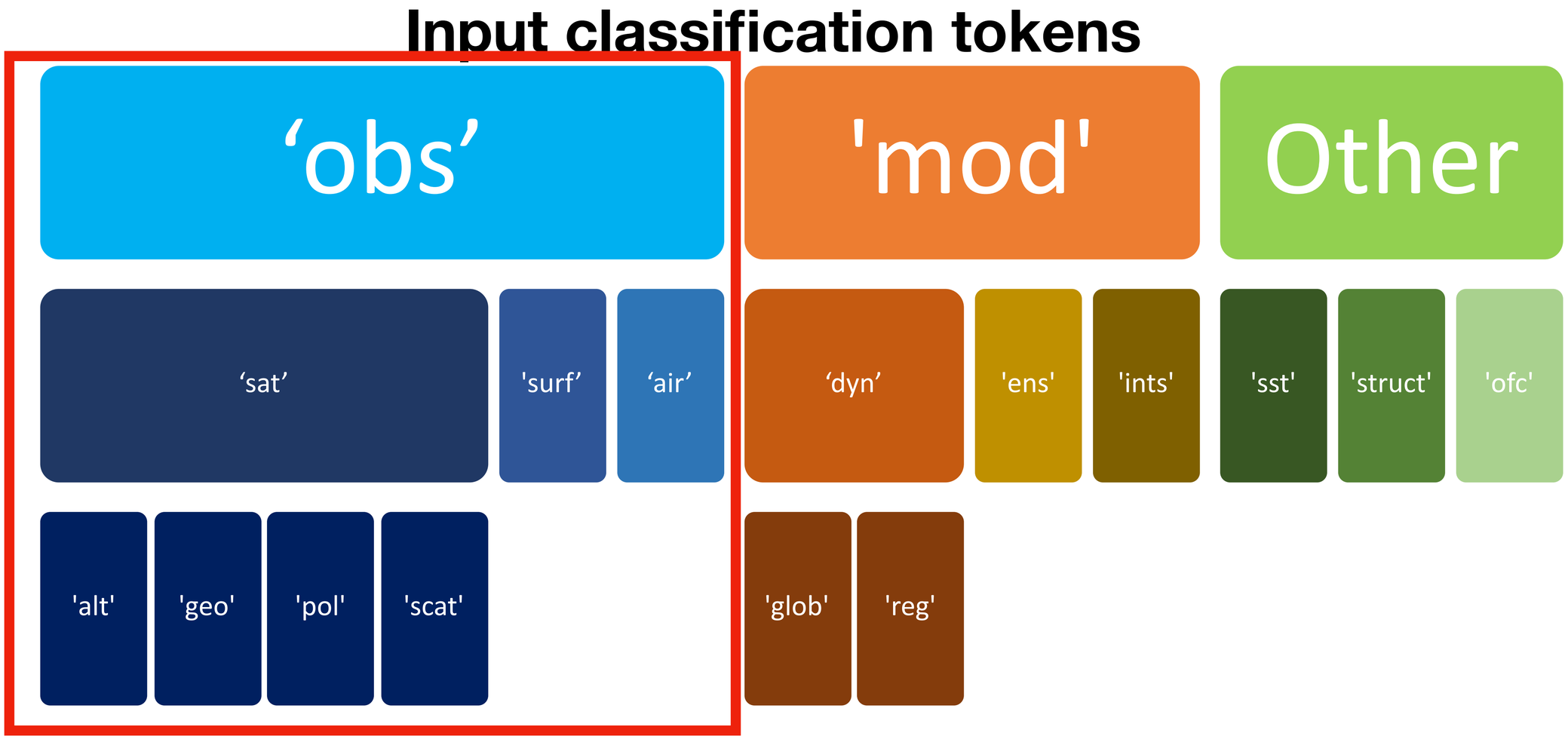
How does this work?

A text classification model for forecast discussions.

Another simple model of the TC Forecast Process

How does this work?

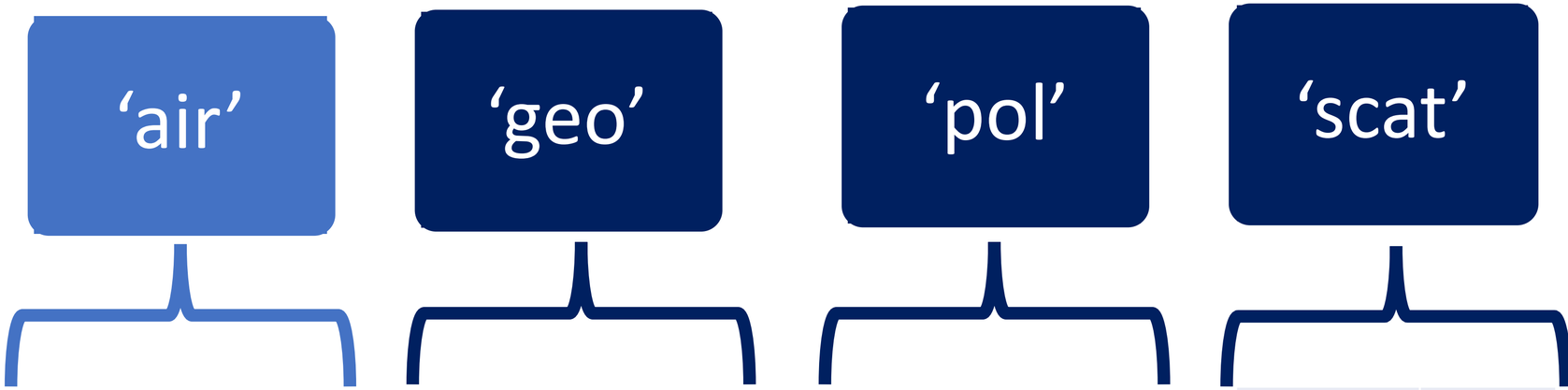
A text classification model for forecast discussions.



How does this work?

A text classification model for forecast discussions.

Zooming in to ‘obs’



air force	obs-air
aircraft	obs-air
dropsonde	obs-air
flight	obs-air
flight-level	obs-air
g-iv	obs-air
hurricane hunter	obs-air
p-3	obs-air
plane	obs-air
sfmr	obs-air
stepped frequency microwave radiometer	obs-air

cimss	obs-sat-geo
dvorak	obs-sat-geo
first-light	obs-sat-geo
geostationary	obs-sat-geo
goes 16	obs-sat-geo
goes imagery	obs-sat-geo
goes infrared	obs-sat-geo
goes satellite	obs-sat-geo
goes shortwave	obs-sat-geo
goes visible	obs-sat-geo
goes-15	obs-sat-geo
goes-16	obs-sat-geo
goes-17	obs-sat-geo
goes-e	obs-sat-geo
goes-east	obs-sat-geo
goes-w	obs-sat-geo
goes-west	obs-sat-geo
infrared	obs-sat-geo
uw-cimss	obs-sat-geo

37 ghz	obs-sat-pol
37-ghz	obs-sat-pol
83 ghz	obs-sat-pol
83-ghz	obs-sat-pol
85 ghz	obs-sat-pol
85-ghz	obs-sat-pol
amsr	obs-sat-pol
amsr-2	obs-sat-pol
amsr2	obs-sat-pol
amsu	obs-sat-pol
avhrr	obs-sat-pol
gcom	obs-sat-pol
gpm	obs-sat-pol
metop	obs-sat-pol
metop-a	obs-sat-pol
metop-b	obs-sat-pol
microwave	obs-sat-pol
noaa-19	obs-sat-pol
noaa-20	obs-sat-pol
polar orbiter	obs-sat-pol
polar orbiting	obs-sat-pol
sounder	obs-sat-pol
ssmi	obs-sat-pol
ssmis	obs-sat-pol
windsat	obs-sat-pol

ascat	obs-sat-scat
ascat-a	obs-sat-scat
ascat-b	obs-sat-scat
ascat-c	obs-sat-scat
oscat	obs-sat-scat
quickscat	obs-sat-scat
quikscat	obs-sat-scat
rapidscat	obs-sat-scat
scatsat	obs-sat-scat
scatterometer	obs-sat-scat

How does this work?

A text classification model for forecast discussions.

'obs'

'track'

'mod'

'ints'

'ofc'

Tropical Depression Harvey Discussion Number 13
NWS National Hurricane Center Miami FL AL092017
400 PM CDT Wed Aug 23 2017

High-resolution visible satellite images show that the cloud pattern of Harvey is a little better organized than it was this morning, but the system lacks distinct banding features. Surface synoptic observations, ASCAT data, and Dvorak classifications from TAFB and SAB indicated that the cyclone has not strengthened, so the current intensity is held at 30 kt. The global models predict that an upper-level low over the northwest Gulf of Mexico will essentially dissipate in a day or so. Therefore, Harvey is expected to remain in a relatively low-shear environment up to the Texas coast. Moreover, ocean analyses show that a warm eddy that broke off from the Loop Current has drifted westward across the Gulf to a location near the projected path of Harvey. This would also be conducive to strengthening, so it is likely that the system will become a hurricane prior to landfall, although this is not explicitly shown in the NHC forecast for which landfall is indicated between 48 and 72 hours.

Based on the scatterometer data and geostationary satellite fixes the center hasn't moved much this afternoon, although recent imagery suggests a northwestward drift at about 320/2 kt. A weak mid-level ridge to the northeast of Harvey should cause the cyclone to move on a northwestward or north-northwestward track through 48 hours. Later, steering currents weaken as a ridge builds over the southwestern United States and a trough drops down from the Plains. As a result, Harvey should decelerate while making landfall and move very slowly just inland of the coast. Some of the track guidance models, such as the HWRF, have shifted southwestward in comparison to their previous run. The official track forecast is very close to the previous one through 48 hours and is a little slower and to the west after that time. This is very close to the latest dynamical model consensus, TVCN. It should be noted that synoptic surveillance data are currently being collected by the NOAA G-IV jet aircraft and these data will be assimilated into, and hopefully improve the forecasts by, the global models.

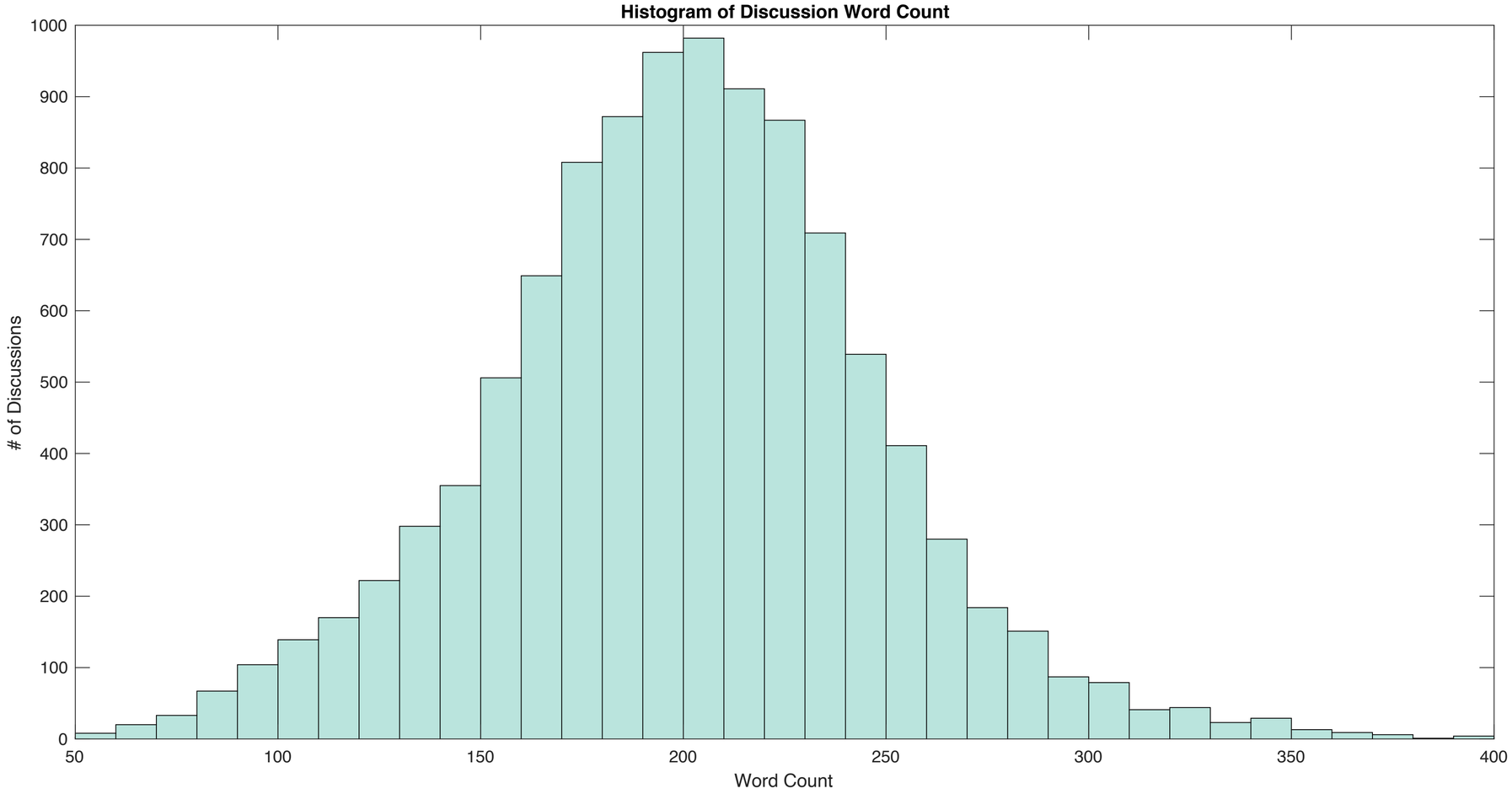
How does this work?

A text classification model for forecast discussions.

- Built a web-scraping engine to collect the last 15 years of NHC tropical cyclone forecasts
- ~280 Storms
- ~10.5k forecasts
- ~2.6MM words
- ***What do the forecasters tell us?***

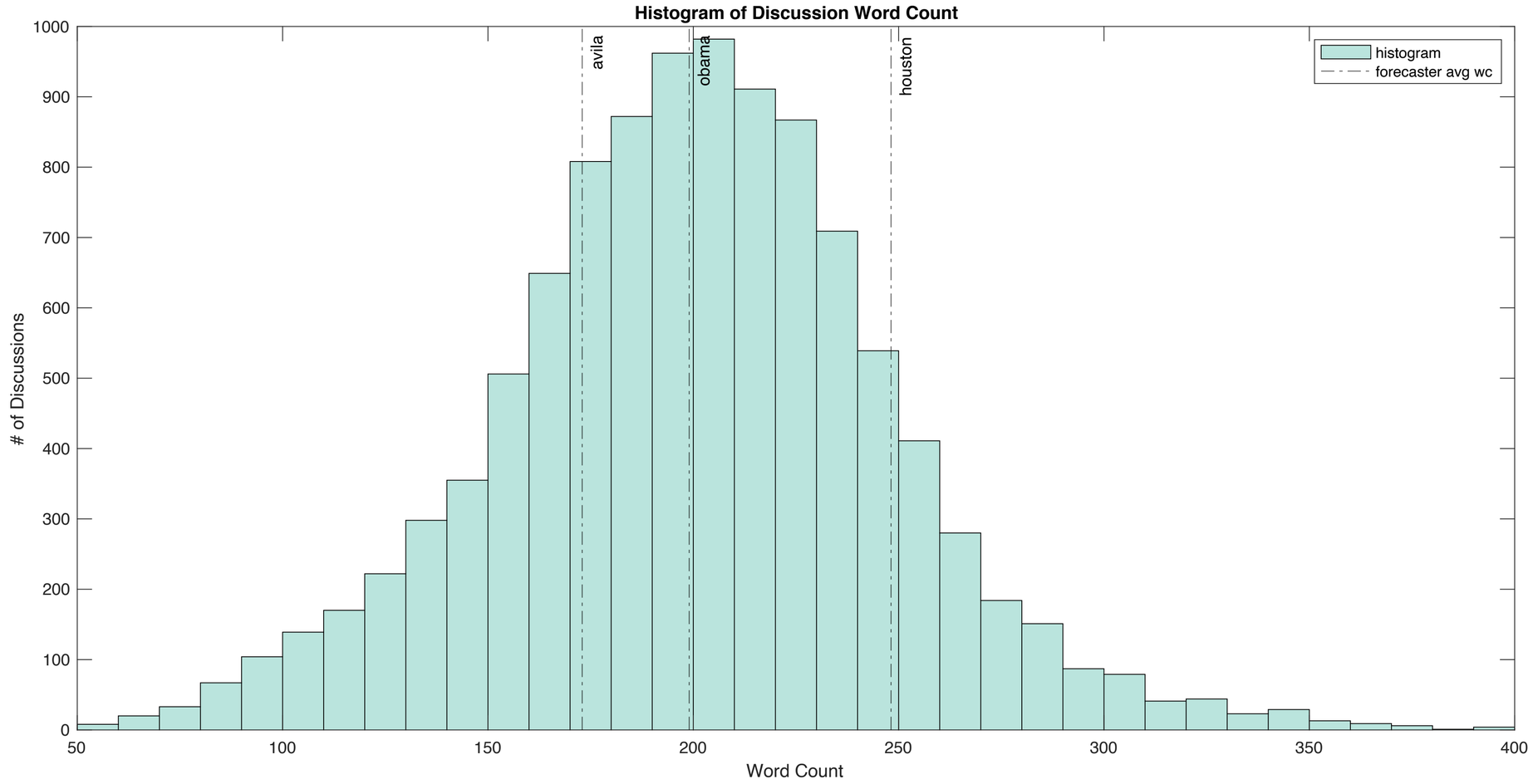
Summary statistics

Word count.



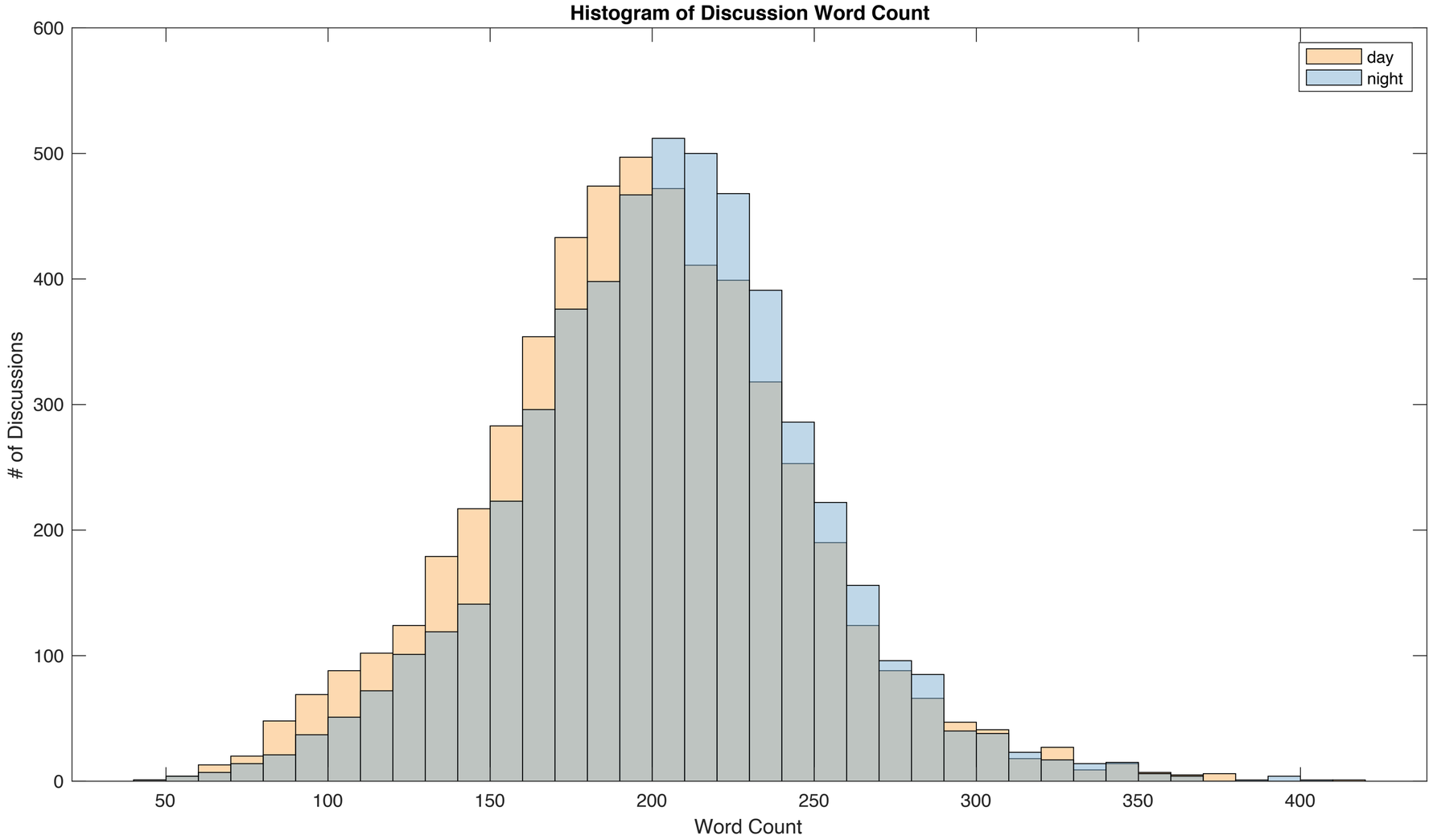
Summary statistics

Word count.



Summary statistics

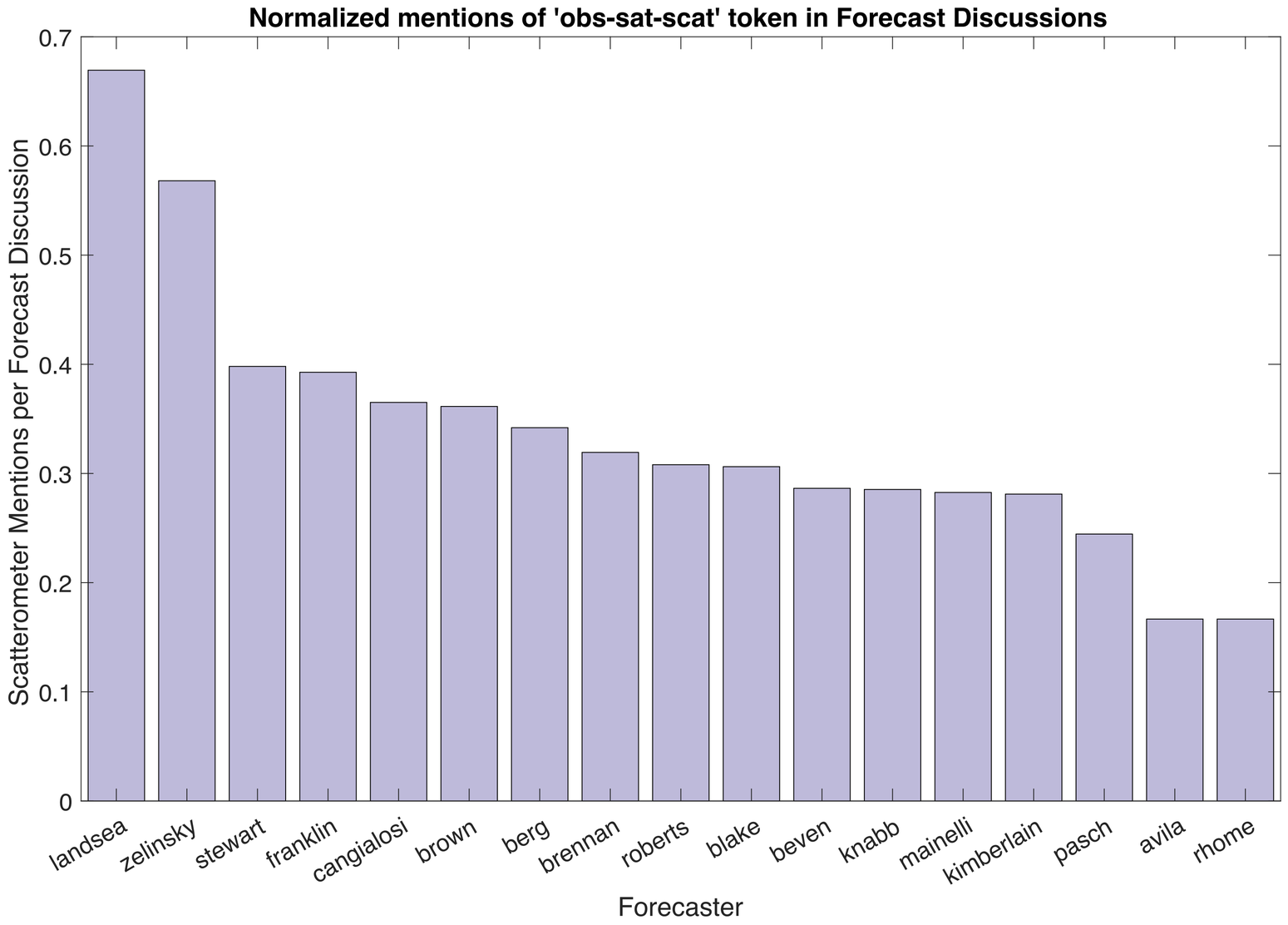
Word count.



Summary statistics

Token mentions.

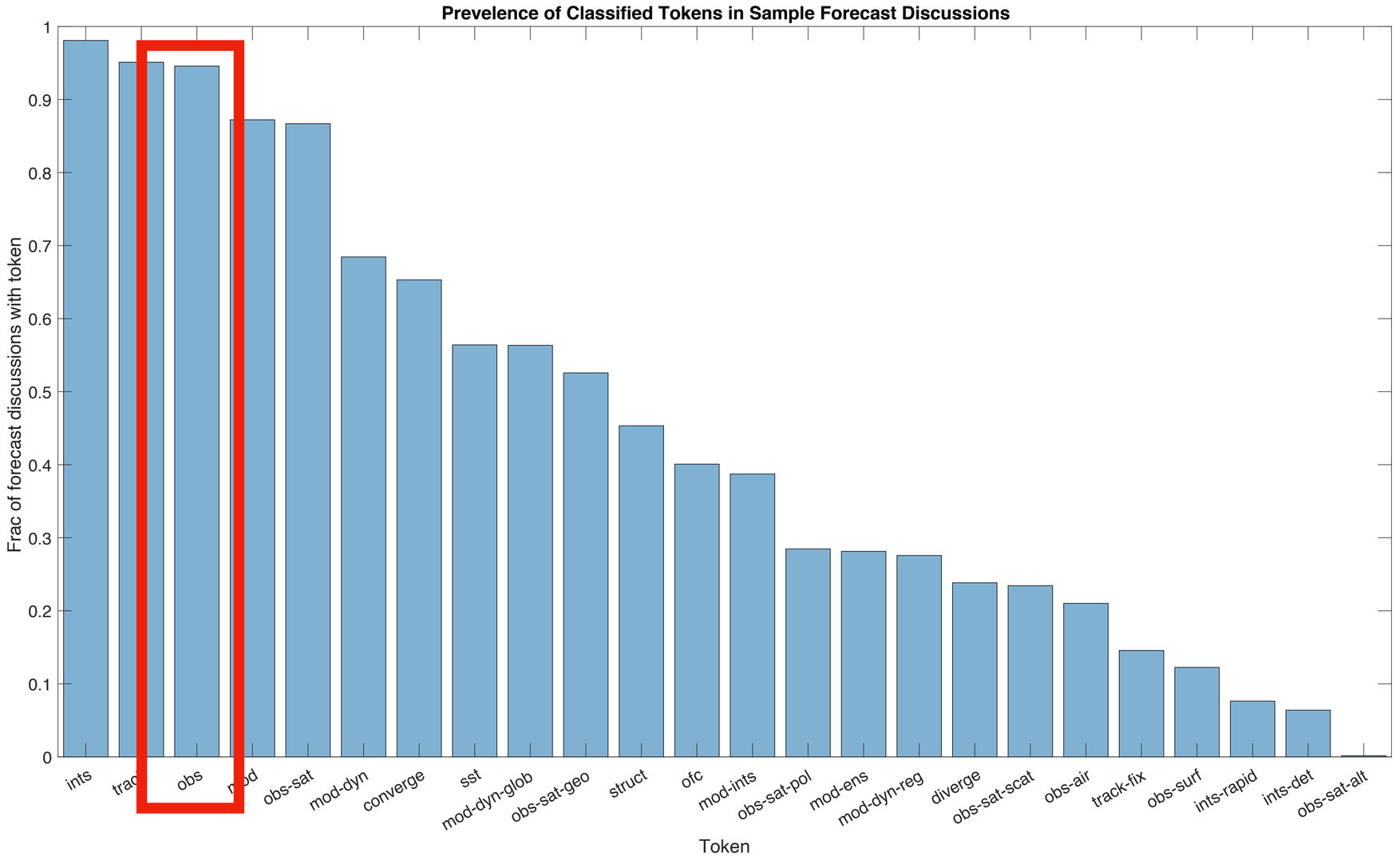
- *Why do different forecasters cite different data in their forecast decisions?*
- *Why do some cite certain observatories more than others?*
- *Are these data representative of the forecaster thought processes?*



Summary statistics

Token mentions.

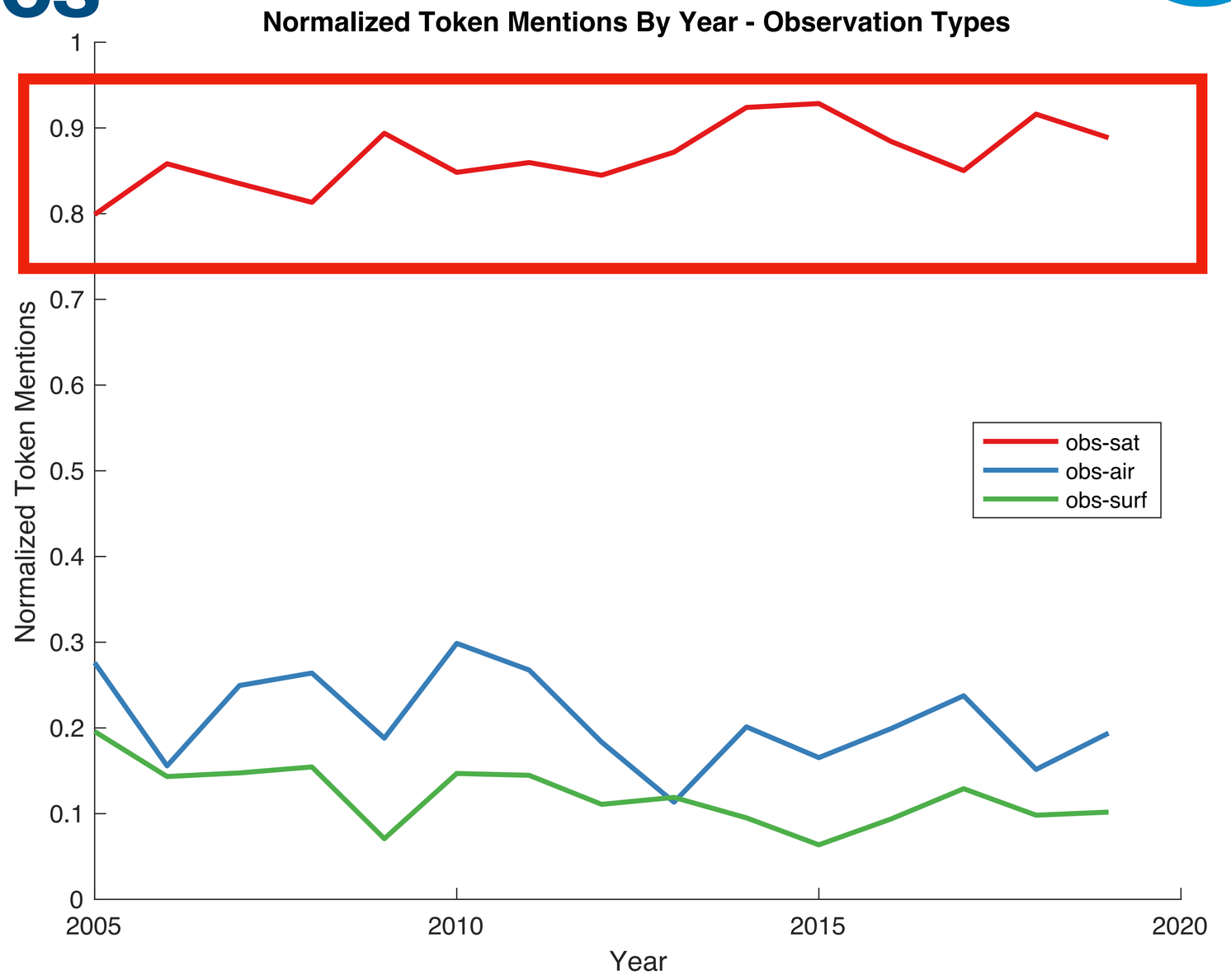
- *Nearly every forecast discussion has an ‘output’ token (an intensity or track forecast)*
- *Nearly every forecast has an ‘input’ token (explicit reference to a model or observatory)*
- ***Most have all of the above***



Summary statistics

Token mentions.

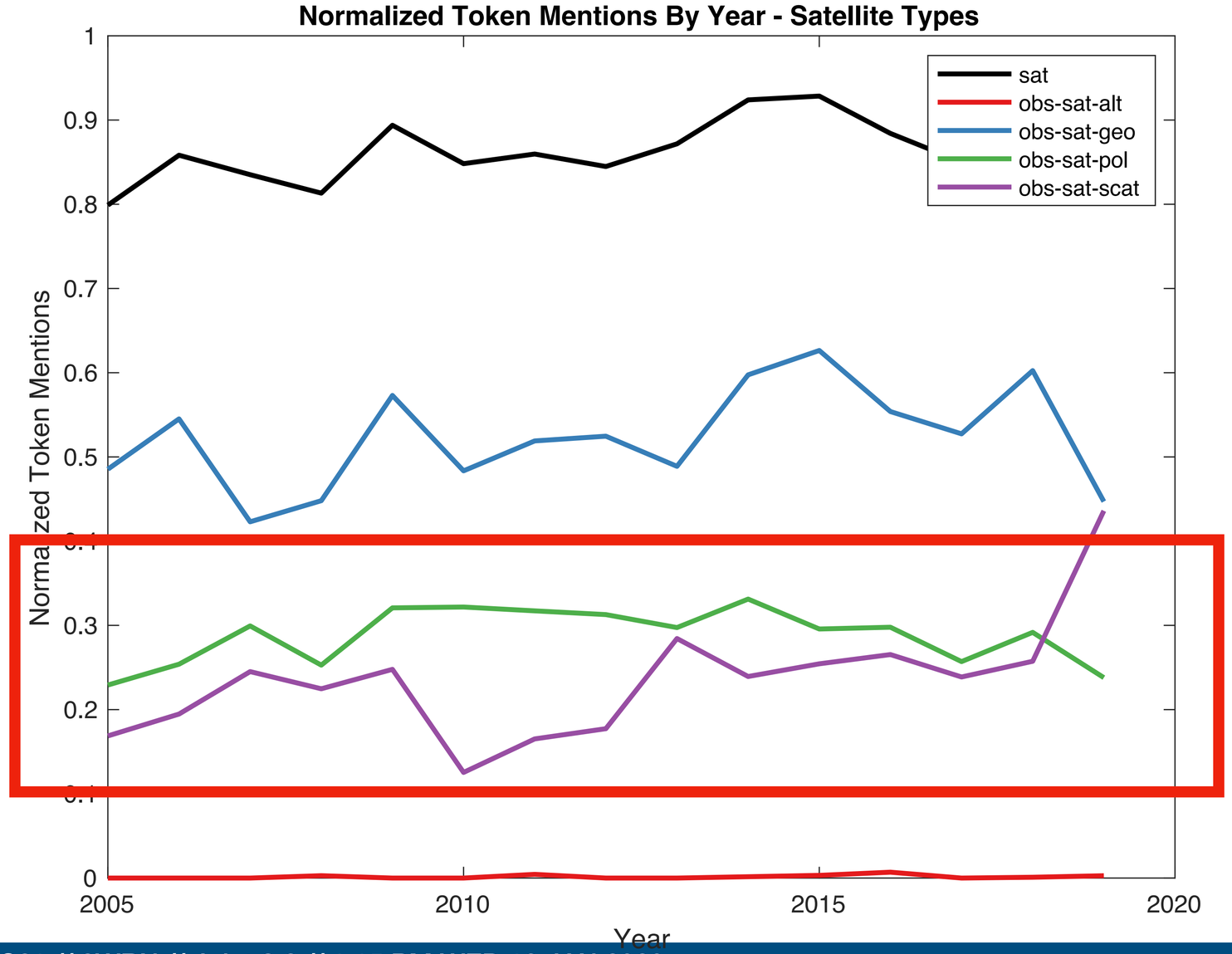
- *There are far more frequent references to satellite observations than other types of observations.*
- *Given the ubiquitous and global nature of satellite data, this makes intuitive sense.*
- *Does that mean satellite data are more valuable, or simply more available?*



Summary statistics

Token mentions.

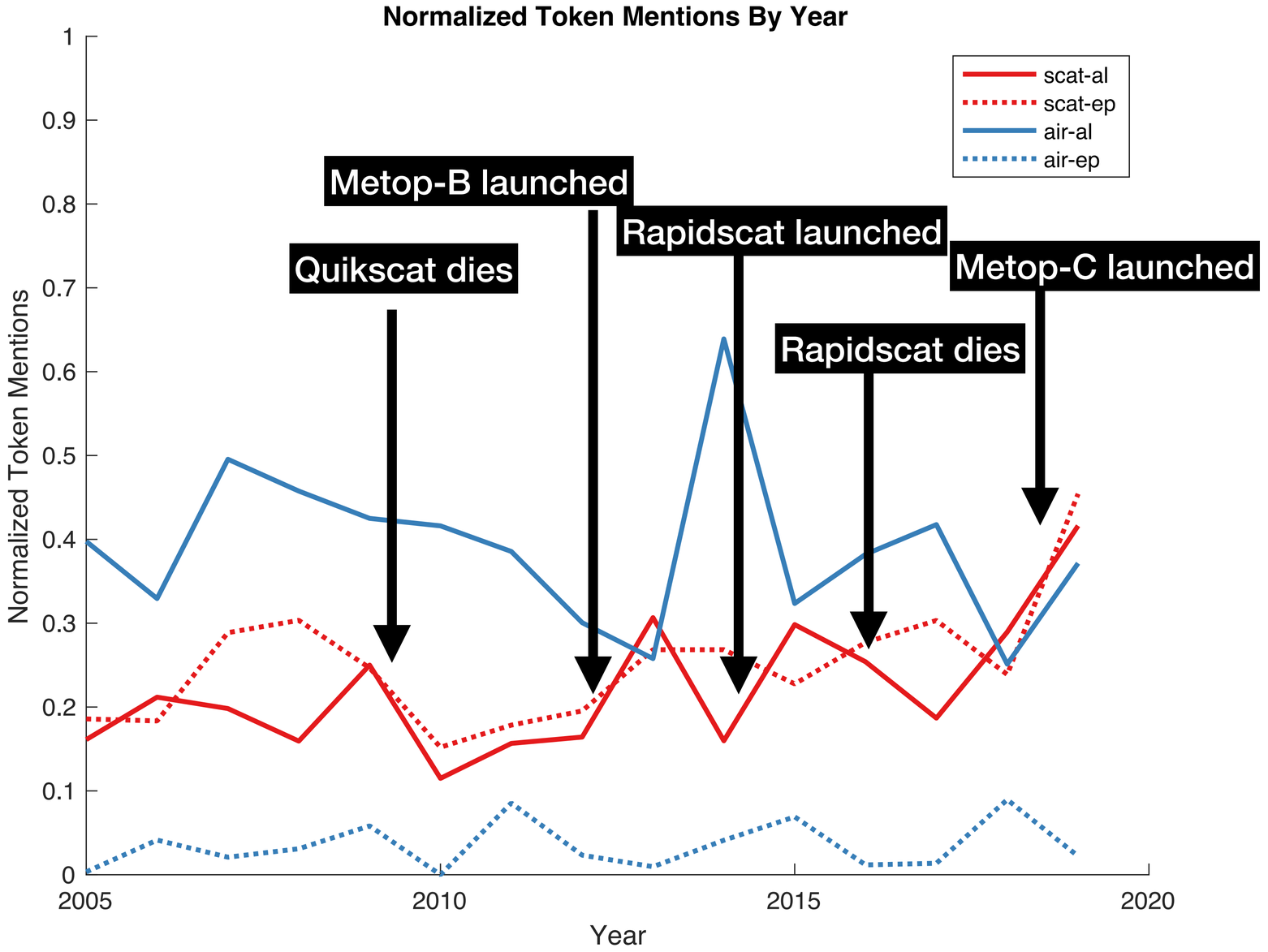
- *Given the the "staring" nature of geostationary satellites, there is a unique role in TC forecasting*
- *The "polar" satellites are restricted to references of microwave sounders*
- *Explicit mentions of scatterometers are broken out on a separate line*



Summary statistics

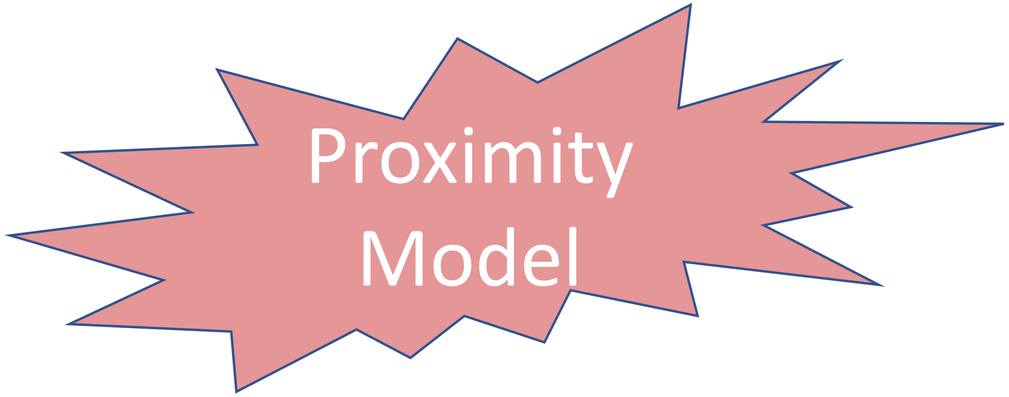
Token mentions.

- *Scatterometer references follow major milestones in the availability of scatterometry data*
- *Aircraft data does not significantly "cannabalize" scatterometry data, it appears forecasters will use data when available*
- *There's significantly more aircraft references in the Atlantic basin than the Eastern Pacific, but scatterometry references are similar across both basins*



Proximity analysis

How are data used?



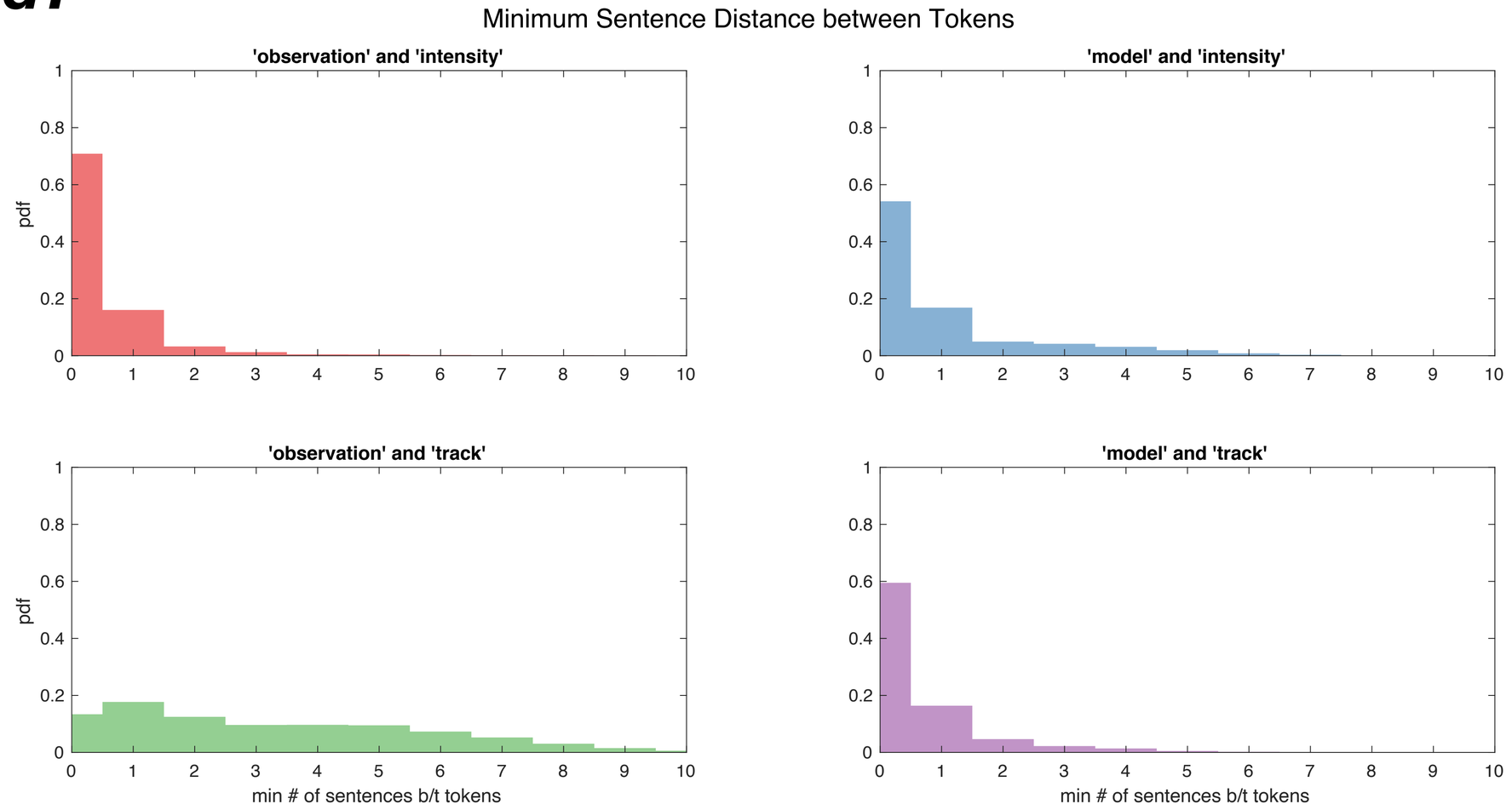
Argument:
Related tokens are closer together

1. Minimum Sentence Distance (MSD)
   - The *minimum* number of sentences between tokens in a FD
     - (Tokens may appear multiple times in a FD, we choose the shortest distance!)
   - Range: [0, ~15-20]
2. Minimum Word Distance (MWD)
   - The *minimum* number of words between tokens in a FD
   - Range: [1, ~150]

Proximity analysis

How are data used?

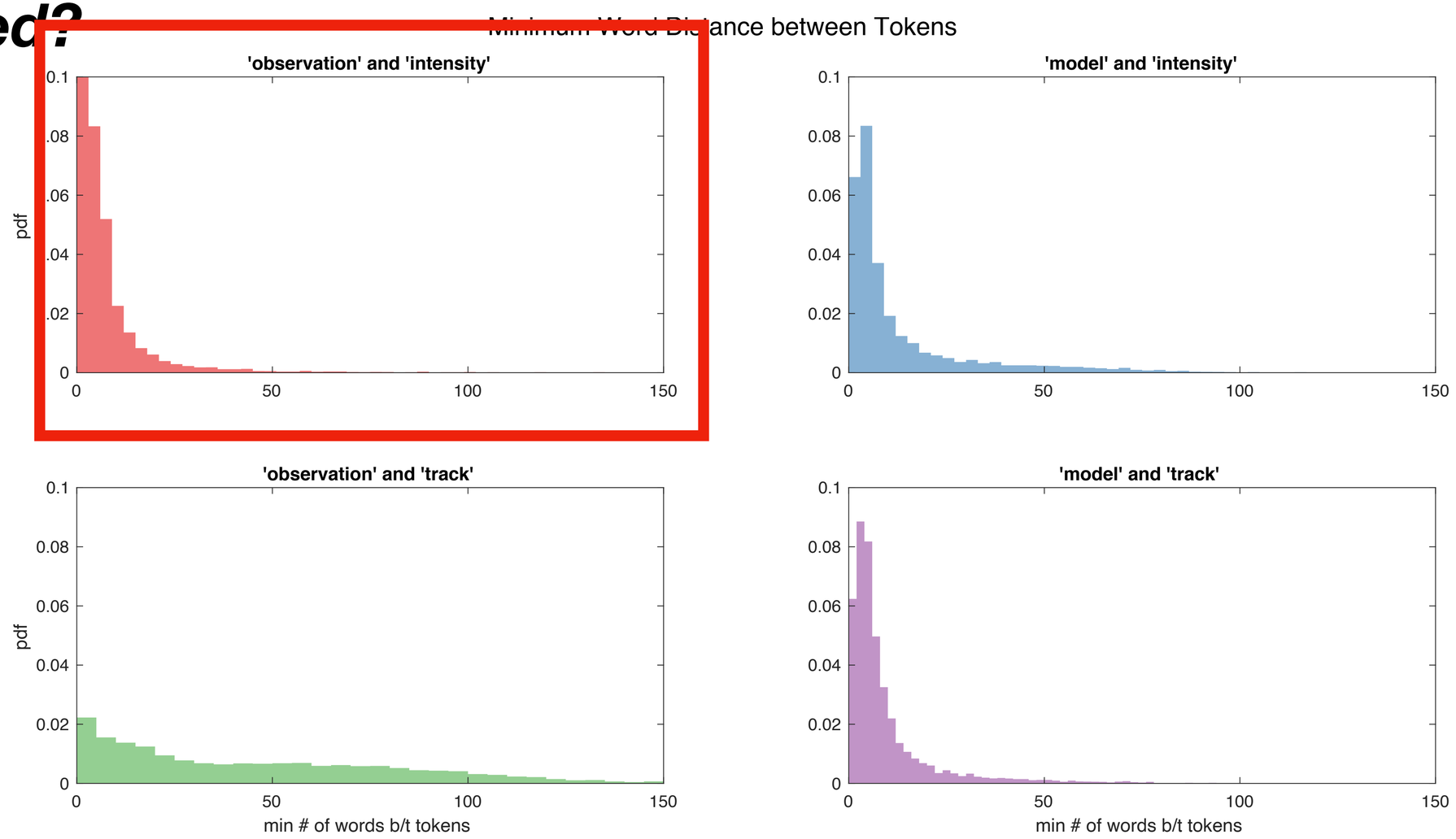
- *Observations are closely related to intensity forecast tokens*
- *Model tokens are also closely related to intensity forecast tokens, but to a lesser degree*
- *Model tokens are closely related to track forecasts*
- *There is no discernible matched token pair relationship between observations and track forecasts*



Proximity analysis

How are data used?

- *Observations are closely related to intensity forecast tokens*
- *Model tokens are also closely related to intensity forecast tokens, but to a lesser degree*
- *Model tokens are closely related to track forecasts*
- *There is no discernible matched token pair relationship between observations and track forecasts*

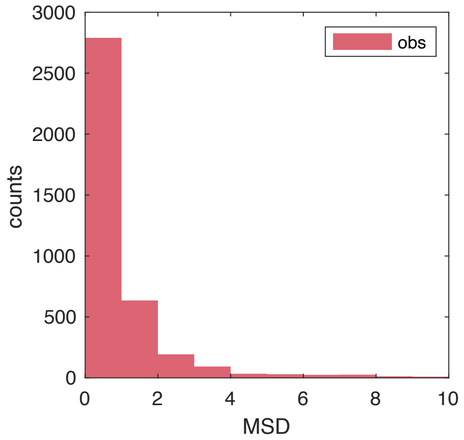


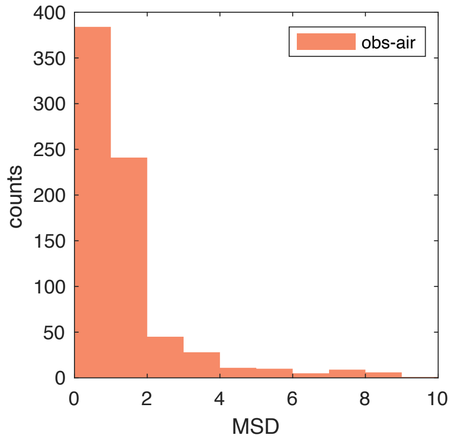
Proximity analysis

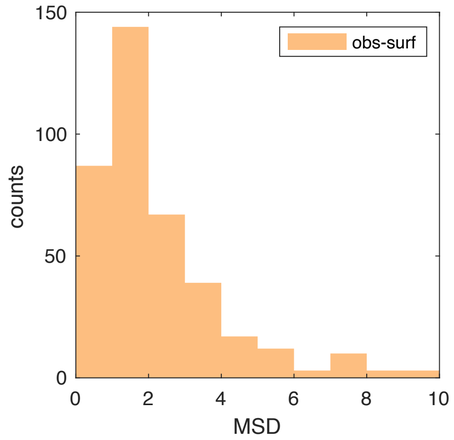
How are data used?

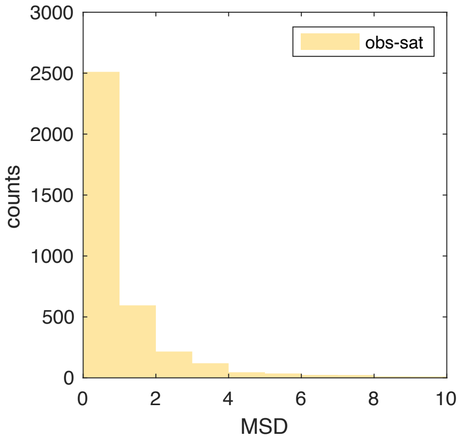
- *Strongest relationships*
  - *Aircraft data*
  - *Geostationary*
  - *Scatterometer*
- *Limited data*
  - *Altimeter*
- *Note the dynamic scales*

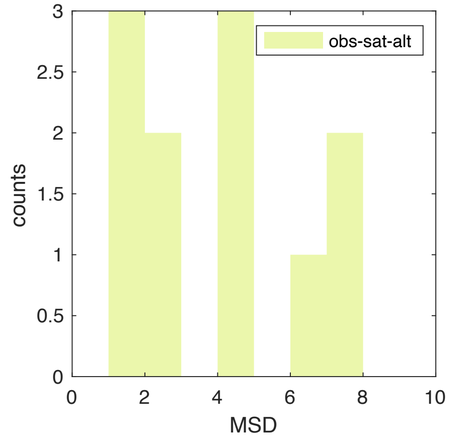
Minimum Sentence Distance Between 'intensity' and 'obs' tokens

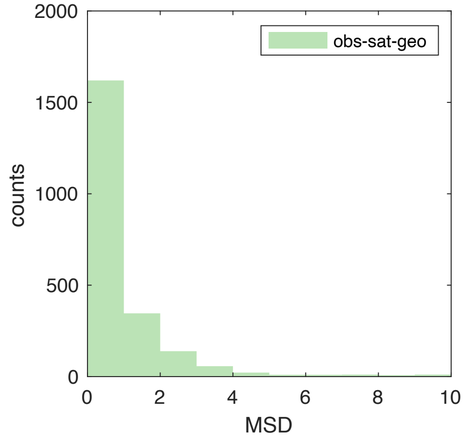


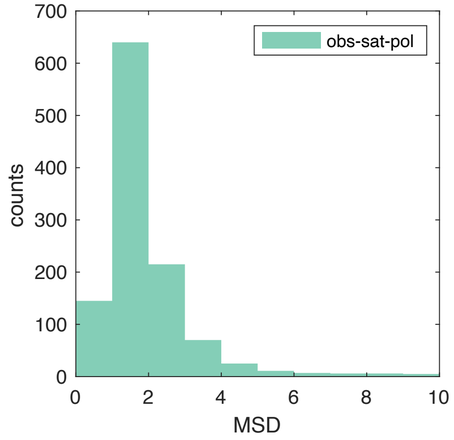


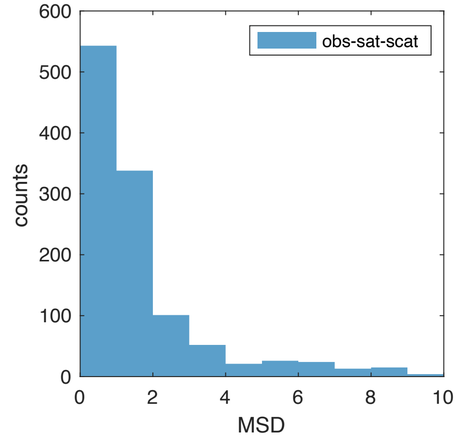








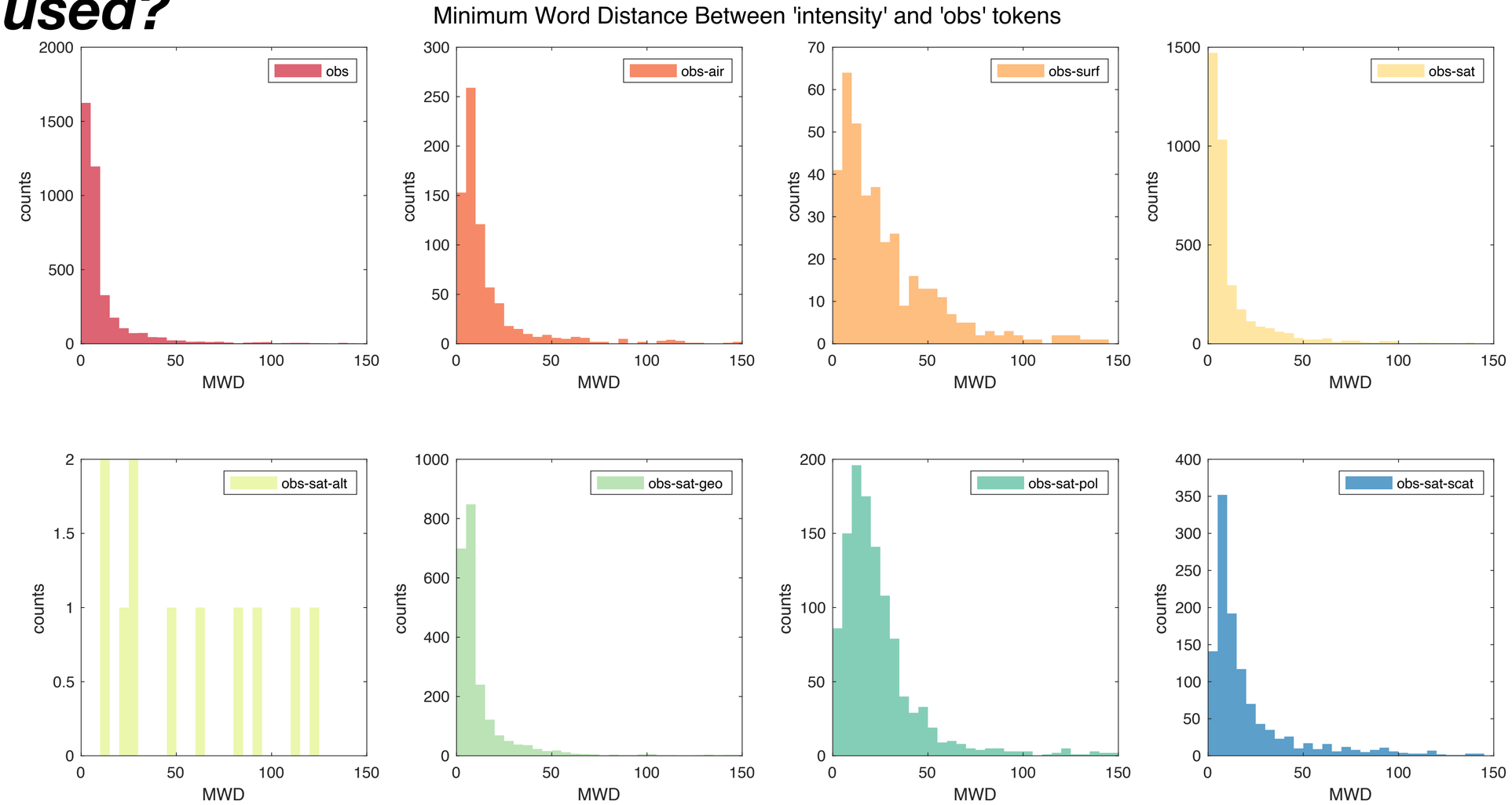




Proximity analysis

How are data used?

- *Strongest relationships*
  - *Aircraft data*
  - *Geostationary*
  - *Scatterometer*
- *Limited data*
  - *Altimeter*
- *Note the dynamic scales*



What conclusions can we draw?

And under what framework are they valid?

- **Forecast discussions can provide insight into the forecaster's thought processes and decisions**
  - Still sensitive to systemic biases (training documents, style guides, etc.) imposed during forecast generation
- **Matching bulk FD data to other forecast skill metrics *may* provide statistically-significant insights on how these decisions contribute to forecast skill**
  - Careful thought needs to put in to the construction of any classification schema and proximity model
- **We shouldn't overlook text products or other unstructured data**
  - It memorializes the decision-making process and should be archived and catalogued
- **What questions do you think this type of data can answer?**