IPSW 2022 IATA Final Presentation

Turbulence in the air: creating a heat map and building a seasonal diagram

Javier Almonacid (SFU), Sean Bohun (OntarioTech),
Douglas Bowen (WLU), Thomas Gkelsinis (Rouen),
Slim Ibrahim (UVic), Michael R Lindstrom (UTRGV),
Joy Liu (Dalhousie), Kyle Onghai (UCLA), Alberto Fornaci (IATA)

August 26, 2022

IATA

Agenda

Turbulence

What is the problem?

1. Each year, approximately 58 people in the United States are injured by turbulence while not wearing their seat belts [1]
2. Current turbulence forecasts may be inaccurate
3. Weather radar cannot predict clear air turbulence
4. Predicted that turbulence is going to increase by 149% in the future
5. Pilot reports are subjective

How is Turbulence Measured?

What is the Eddy Dissipation Rate (EDR)?

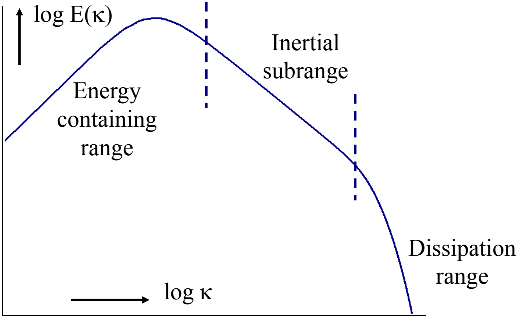
Aircrafts measure turbulence through the **eddy dissipation rate** (EDR). Some important characteristics of EDR are:

1. Greater values of EDR correspond to more significant turbulence
2. EDR can take any value in $[0, \infty)$, but its range practically speaking is $[0, 1]$ ($[\mathrm{EDR}] = \mathrm{m}^{2/3}\ \mathrm{s}^{-1}$)
3. EDR is insensitive to the size and weight of the aircraft that records the parameters fed to the algorithm as input

How is Turbulence Measured?

Eddy Dissipation Rate (EDR)

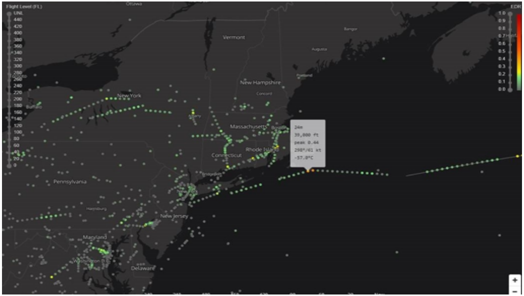
1. Ensure the the flight is within nominal characteristics
2. Sample the wind speed and compute the energy spectrum. In the inertial subrange, $E(k, \epsilon) = C\epsilon^{2/3}k^{-5/3}$
3. Compare $E(k, \epsilon)$ to a standard model, $E_0(k; 1)$, with $\epsilon = 1$ to estimate ϵ

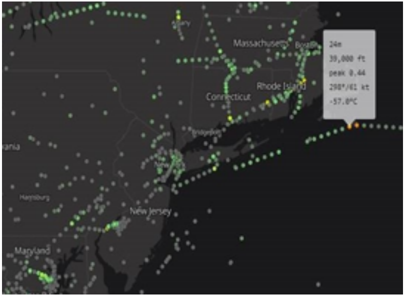


Current State

What has been accomplished already?

1. IATA has the ability to plot individual measurements of EDR for the past four hours
2. These plots provide too much detail
3. A visual representation of the clear air turbulence summarizing this information is needed





What Does IATA Want?

Problem Statement

How can we take live turbulence data and represent it as a heat map that is **quick and easy for pilots to interpret**?

Important considerations

1. Large turbulence readings should be represented with high fidelity
2. Heat maps should be updated in real time based on the most recent EDR measurements
3. Regions of high EDR should not be distorted when the image is enhanced

Methods

Summary of our methods

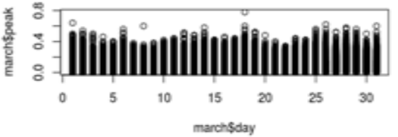
1. **Statistical Model:** We studied the monthly pattern of EDR via a linear GAM model
2. **Voronoi:** Initial Voronoi diagrams were satisfactory
3. **KDE Heat Map:** We extrapolated the EDR data by employing kernel density estimation to plot the data, achieving better results than Voronoi
4. **Clustering:** To remedy the issue of overlapping regions, we clustered spatially close and temporally close points
5. **Spatio-temporal Weighting:** We estimated EDR using the intuition that clustered points should be more heavily weighted than isolated ones

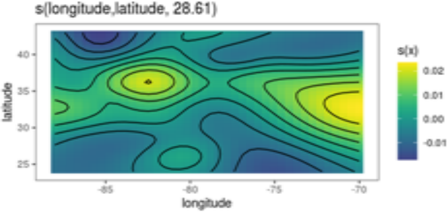
Summary Statistics

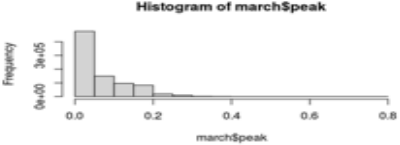
- In the data, we have the response variable (EDR), which is zero-inflated and heavily skewed
- We also have predictor variables:

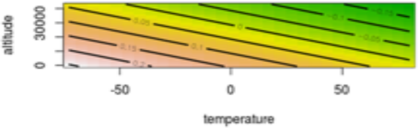
x: position (lon, lat), t: time, a: altitude

- There is a strong linear correlation between altitude and temperature









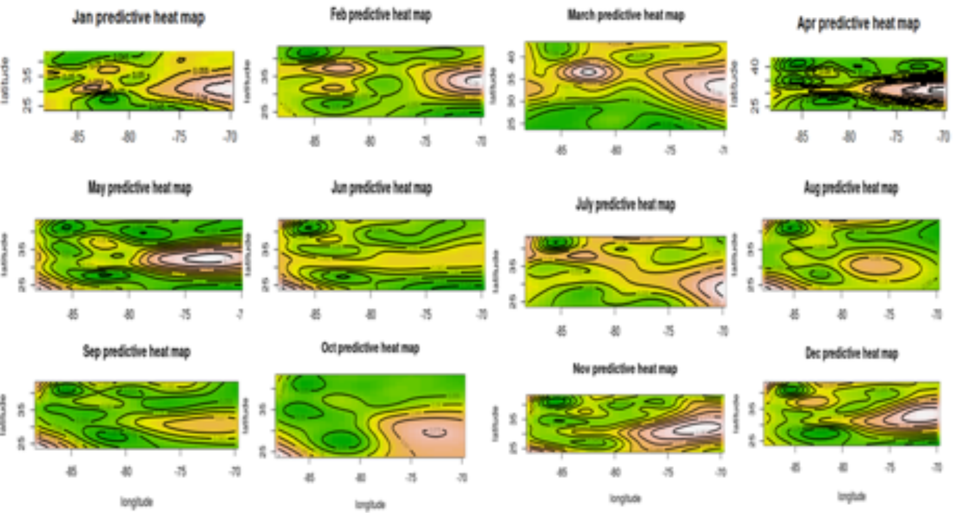
Statistical Model

1. To explore the monthly effect on EDR, we used the **generalized additive model** (GAM), in which the response variable (EDR) depends linearly on unknown smooth functions of the positional variables
2. The smooth functions are of interest
3. Use **5-fold cross validation method** to plot the residuals in space, indexed by month

The Model

1. $y_i = \log(\mathrm{EDR}_i + 1) \sim \mathcal{N}(u_i, \sigma^2)$, with $u_i = \beta_0 + \beta_1 t_i + \beta_2 a_i + s(x_i)$
2. t_i and a_i are the time and altitude variables, respectively, with fixed coefficients β_1 and β_2
3. $s(x_i)$ is a smooth function of the position x_i

Seasonal Heat Map



Results

- Altitude has a negative effect on EDR across the months

  ($\searrow$) Increasing altitude decreases EDR

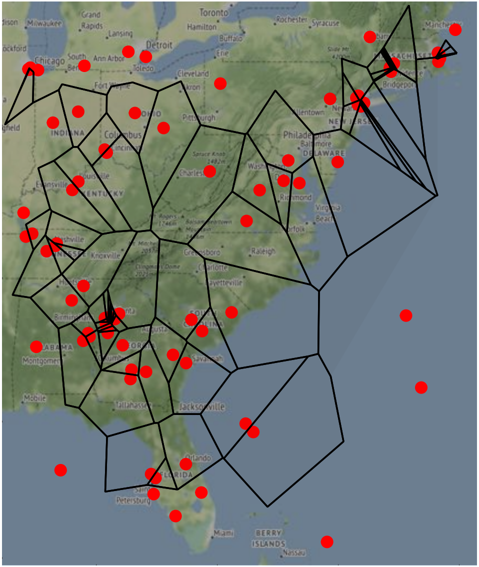
- Date has different effects on EDR across the months

  ($\searrow$) Negative: May–Sept, Nov   ($\nearrow$) Positive: Dec–April, Oct

- The smooth function has a temporal pattern
- Residuals are inconsistent with normal distribution prior and other distributions should be studied

Voronoi Diagram

Because the EDR measurements can be densely packed, the Voronoi diagram does not help visualize the data



Kernel Density Heat Map

- The Kernel Density Heat Map assigns a bi-variate kernel (e.g. Gaussian) in each data-point on the plane in order to estimate the probability density of the events
- The Bi-variate Kernel Density Estimate is defined to be

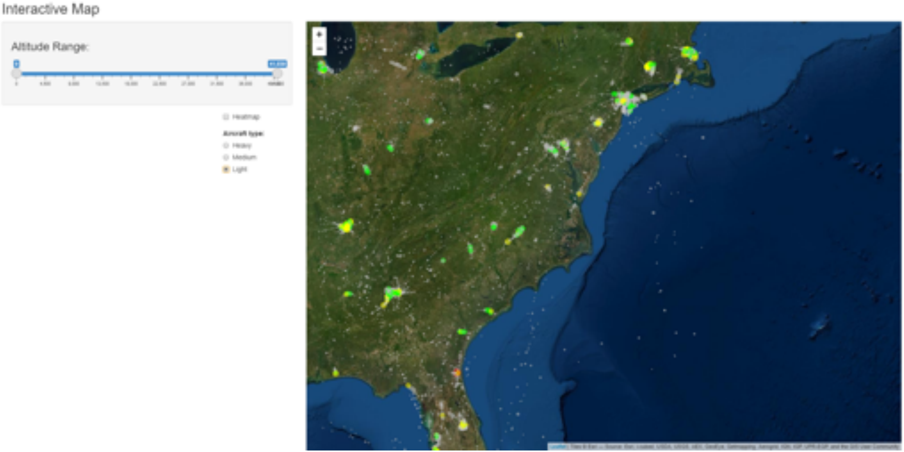
$$\hat{f}_{\boldsymbol{H}}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{x}_i),$$

where

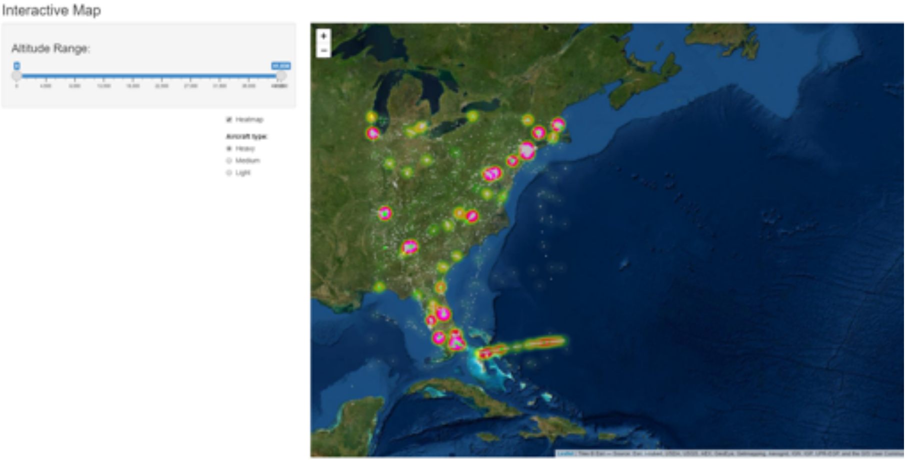
  - $\boldsymbol{x} = (x_1, x_2)^T,\ \boldsymbol{x}_i = (x_{i1}, x_{i2})^T,\ i = 1, 2, \ldots, n$
  - $\boldsymbol{H}$ is the bandwidth (or smoothing parameter) 2×2 matrix
  - K is the kernel function which is a symmetric bi-variate density
  - $K_{\boldsymbol{H}}(\boldsymbol{x}) = |\boldsymbol{H}|^{-\frac{1}{2}} K(\boldsymbol{H}^{-\frac{1}{2}} \boldsymbol{x})$

- In our case we use a 2-dimensional Gaussian Kernel where the bandwidth matrix $\boldsymbol{H}$ plays the role of the covariance matrix
- Implemented in R using `R-Shiny`

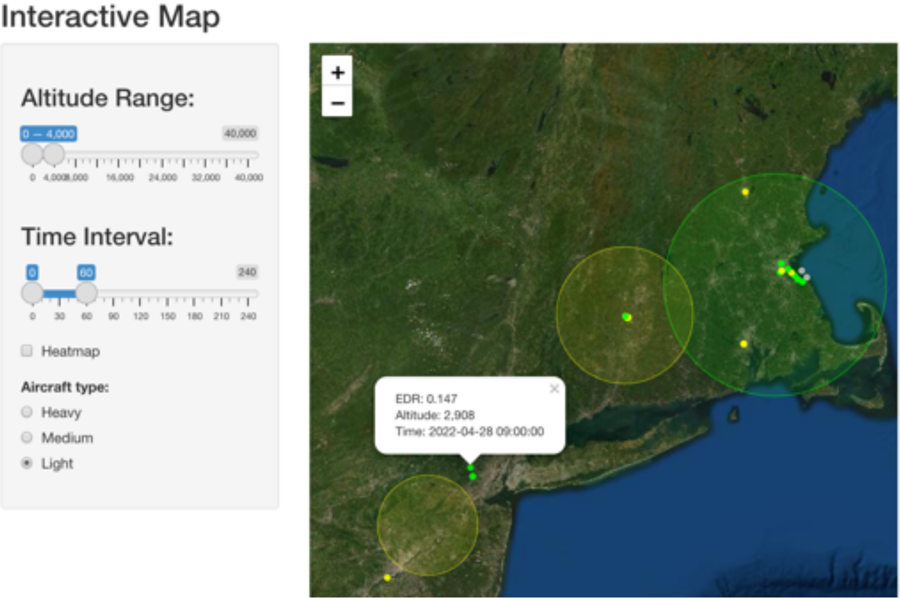
Kernel Density Heat Map



Kernel Density Heat Map

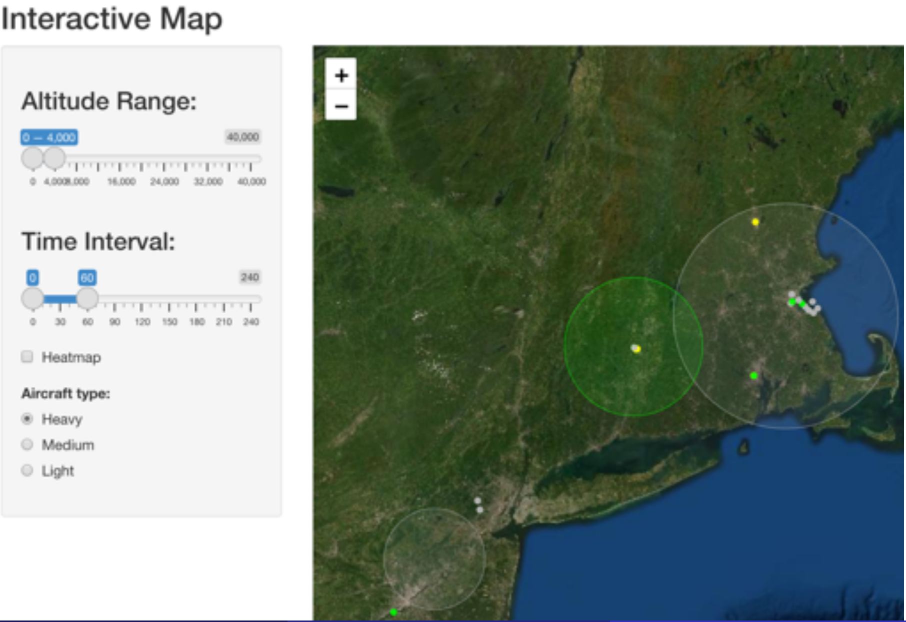


Clustering



Interactive Map
Altitude Range:
0 – 4,000
40,000
0 4,000 8,000 16,000 24,000 32,000 40,000
Time Interval:
0
60
240
0 30 60 90 120 150 180 210 240
Heatmap
Aircraft type:
Heavy
Medium
Light
+
−
EDR: 0.147
Altitude: 2,908
Time: 2022-04-28 09:00:00

Clustering


Interactive Map
Altitude Range:
0 – 4,000
40,000
0
4,0008,000
16,000
24,000
32,000
40,000
Time Interval:
0
60
240
0
30
60
90
120
150
180
210
240
Heatmap
Aircraft type:
Heavy
Medium
Light
+
–

Spatio-temporal Weighting

Algorithm

For dataset $D = \{(x_i, t_i, \epsilon_i)\}_{i=1}^n$ of position-time-EDR values set

$$w_i(x,t) = \sum_{i=1}^{n} \mathrm{e}^{-(t-t_i)/\tau} \mathrm{e}^{-||x-x_i||/\lambda}$$

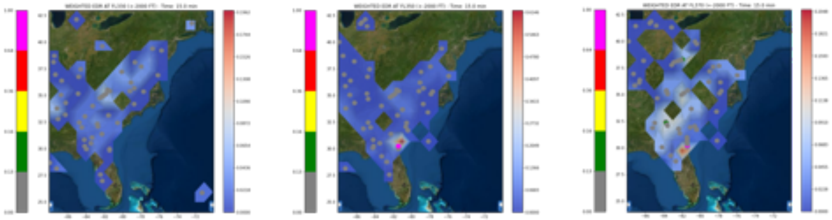
$$\hat{\epsilon}(x,t) = \begin{cases} \dfrac{\sum_{i=1}^n w_i(x,t)\epsilon_i}{\sum_{i=1}^n w_i(x,t)}, & \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} w_i(x,t) > w^* \\ \mathrm{NaN}, & \text{otherwise} \end{cases}$$

- Estimation of EDR is strongly influenced by nearby (in both space and time) measured values
- EDR values should not be estimated at locations too far from the measured points

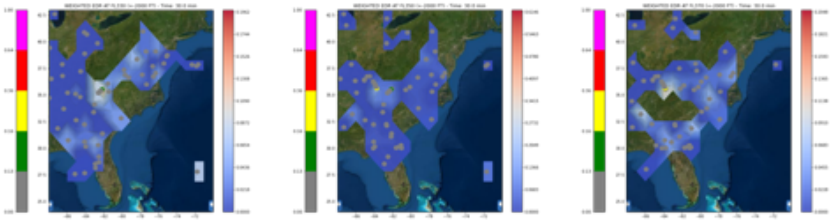
Spatio-temporal Weighting

- We consider a subset of 4 hours of data, grouped into chunks of 15-minute intervals
- We also consider three different flight levels ($\pm$ 2,000 ft): 33,000 ft (FL330), 35,000 ft (FL350), and 37,000 ft (FL370)
- Time and space scales used: $\tau = 4$ hr and $\lambda = 0.065^\circ$
- Threshold used $w^* = 10^{-10}$

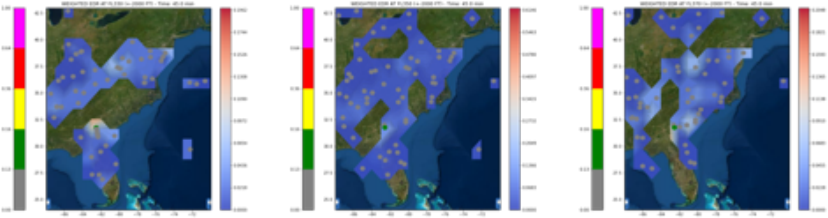
Spatio-temporal Weighting



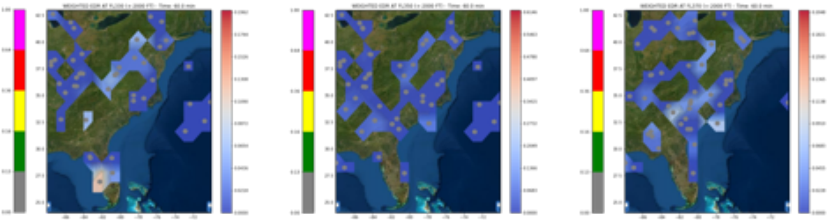
Spatio-temporal Weighting



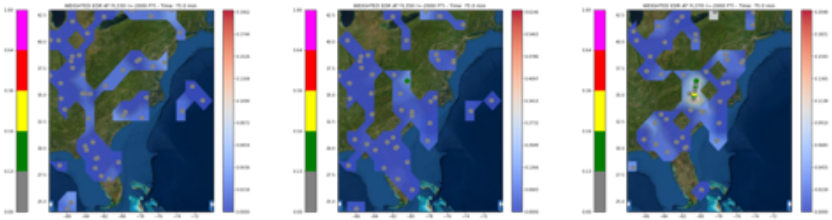
Spatio-temporal Weighting



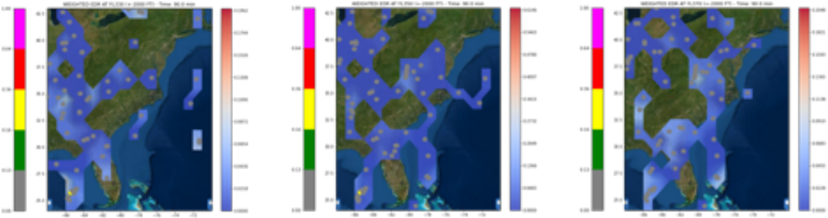
Spatio-temporal Weighting



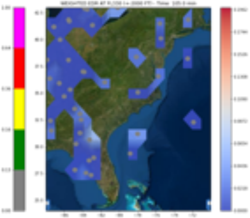
Spatio-temporal Weighting

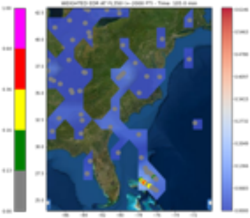


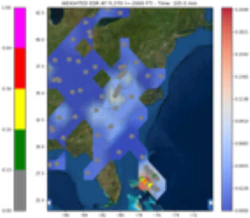
Spatio-temporal Weighting



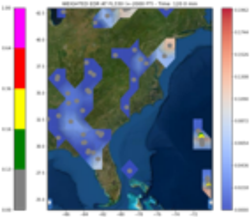
Spatio-temporal Weighting

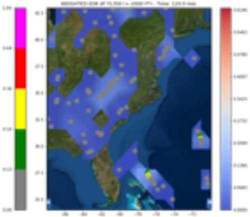


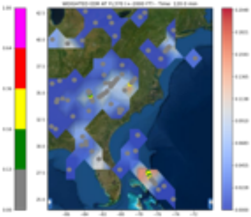




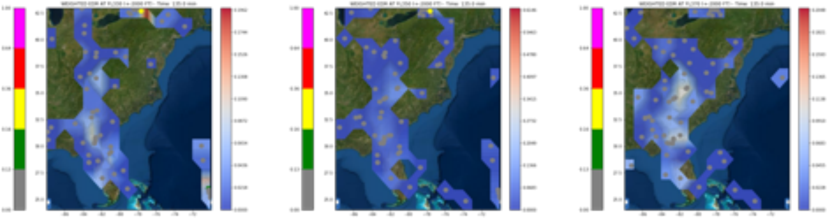
Spatio-temporal Weighting



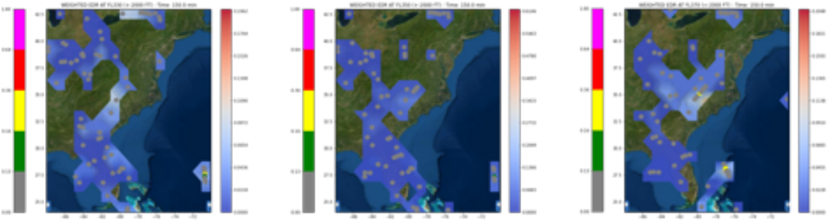




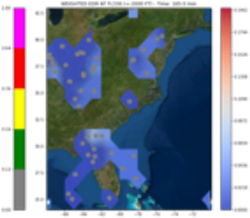
Spatio-temporal Weighting

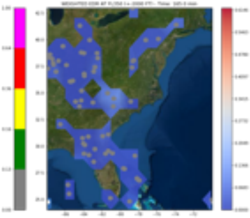


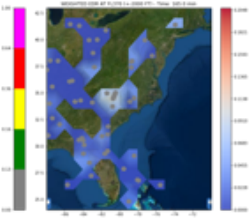
Spatio-temporal Weighting



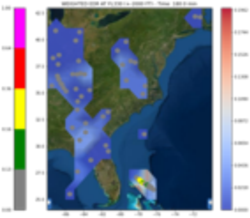
Spatio-temporal Weighting

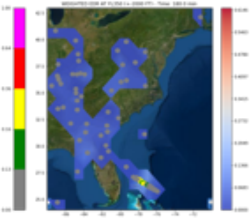


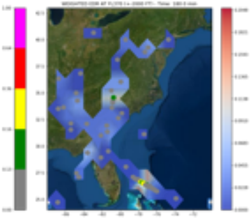




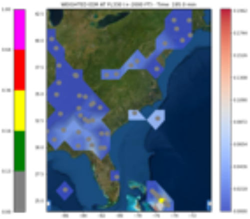
Spatio-temporal Weighting

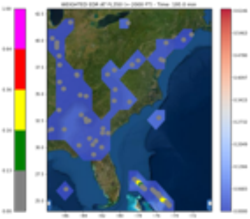


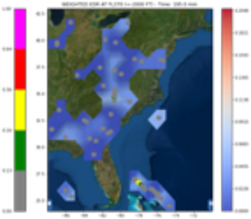




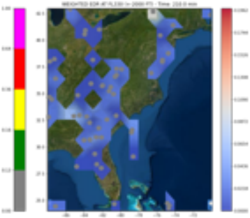
Spatio-temporal Weighting

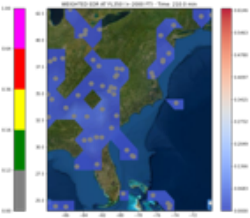


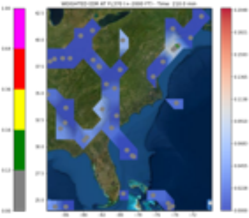




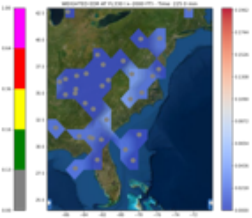
Spatio-temporal Weighting

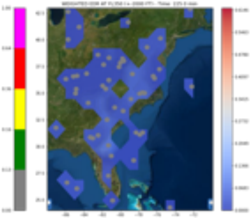


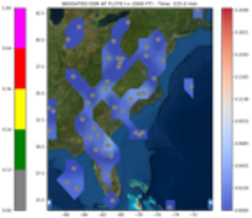




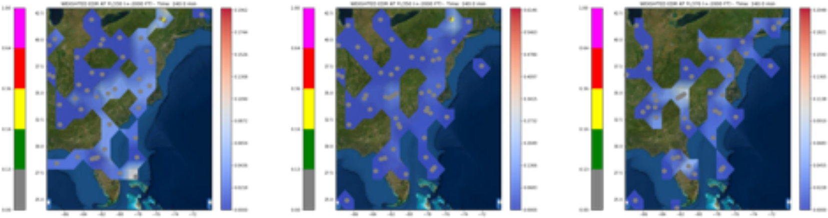
Spatio-temporal Weighting







Spatio-temporal Weighting



Conclusions

Statistical approach

- skewed distribution of EDR instances
- seasonal variations detected

Interactive KDE heat map

- allows dynamic changes of the altitude slice and the time interval
- method strongly depends on the density of the points

Spatio-temporal weighted heat map

- heuristic revealing of the neighbourhood of the observed EDR values
- method does not reveal regions of low confidence
- the neighbourhood of a predicted point is dominated by the influence of the closest measured point

Future Work

Different statistical model

- use a more realistic distribution than normal

Measuring distances more accurately

- consider geometry of the earth for spatial separations where[1] 1° latitude $\neq 1^\circ$ longitude

Tuning of hyper-parameters

- λ, τ, and w^* are tunable
- use statistical and physical heuristics to identify best choices

Mapping software

- further investigate cluster colouring, radius, and positioning choices
- combine the interactive map with assignments from the spatio-temporally weighted method

[1] One degree of latitude equals approximately 364,000 feet (69 miles). One degree of longitude $= \frac{\pi}{180} R \cos\phi$, depends on the latitude ϕ, and R, the radius of the earth.

Bibliography I

Federal Aviation Administration
https://www.faa.gov/travelers/fly_safe/turbulence.

Appendix

Extra movie