A gray-box model for a probabilistic estimate of regional ground magnetic perturbations: Enhancing the NOAA operational Geospace model with machine learning

Enrico Camporeale

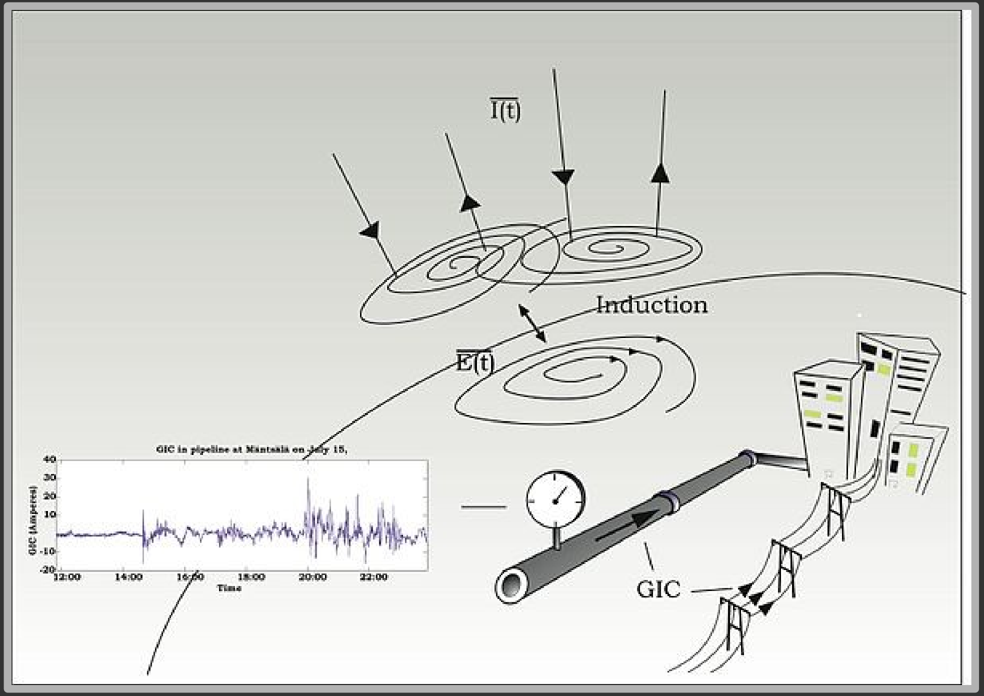
CIRES / CU Boulder & NOAA Space Weather Prediction Center
Center for Mathematics and Computer Science (CWI), Amsterdam, Netherlands

Collaborators: M. Cash, H. Singer, C. Balch (NOAA/SWPC)
Z. Huang, G. Toth (U. Michigan)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA)

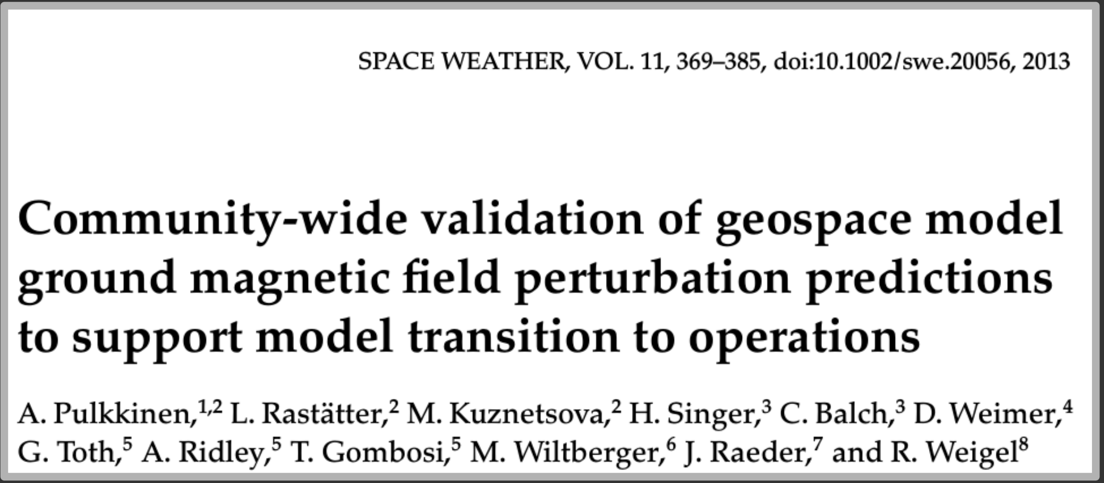
Why do we care about dB/dt?

Sudden changes in the magnetic field induce, via Faraday’s law, an electric field that can generate a Geomagnetically Induced Current (GIC) in long conductors such as electric power lines



Prediction of dB/dt

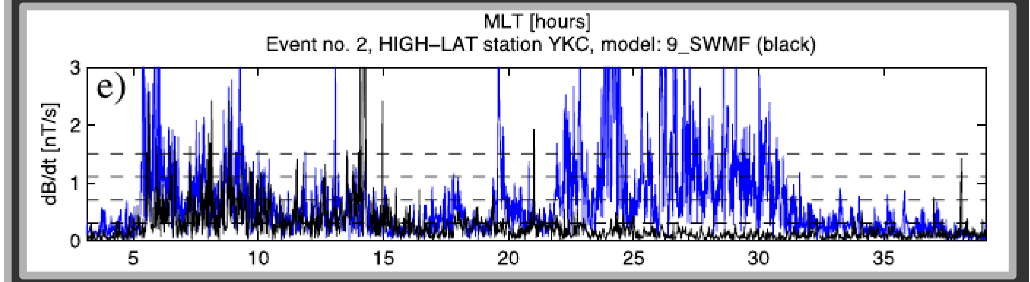
- Forecast is possible because of the propagation time between L1 (where solar wind is observed) and Earth
- However, an high-cadence (1 min) prediction of dB/dt is too difficult
- The *de-facto standard* for validating dB/dt predictions has been introduced in Pulkkinen et al. (2013)



SPACE WEATHER, VOL. 11, 369–385, doi:10.1002/swe.20056, 2013

Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations

A. Pulkkinen,[1,2] L. Rastätter,[2] M. Kuznetsova,[2] H. Singer,[3] C. Balch,[3] D. Weimer,[4] G. Toth,[5] A. Ridley,[5] T. Gombosi,[5] M. Wiltberger,[6] J. Raeder,[7] and R. Weigel[8]



Prediction of dB/dt: problem set-up

$$dB/dt = \max_{\{t,t+\Delta t\}} \sqrt{(dB_n/dt)^2 + (dB_e/dt)^2}$$

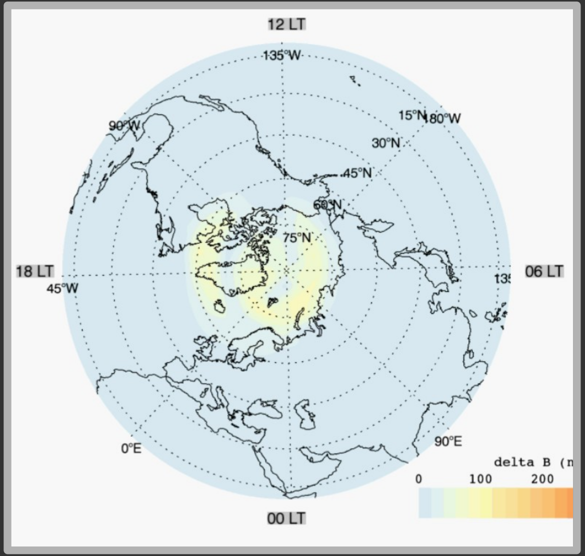
The focus is on predicting whether (the max of) dB/dt will exceed a given threshold in a 20-mins interval (binary classification problem)

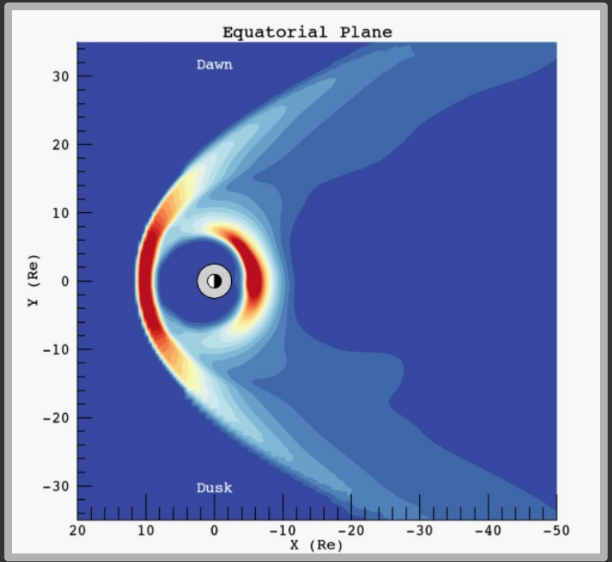
- P = total number of observed positives (event occurrences);
- N = total number of observed negatives (event non-occurences);
- TP = True Positives: number of predicted positives that are observed positives;
- FP = False Positives: number of predicted positives that are observed negatives;
- TN = True Negatives: number of predicted negatives that are observed negatives;
- FN = False Negatives: number of predicted negatives that are observed positives;

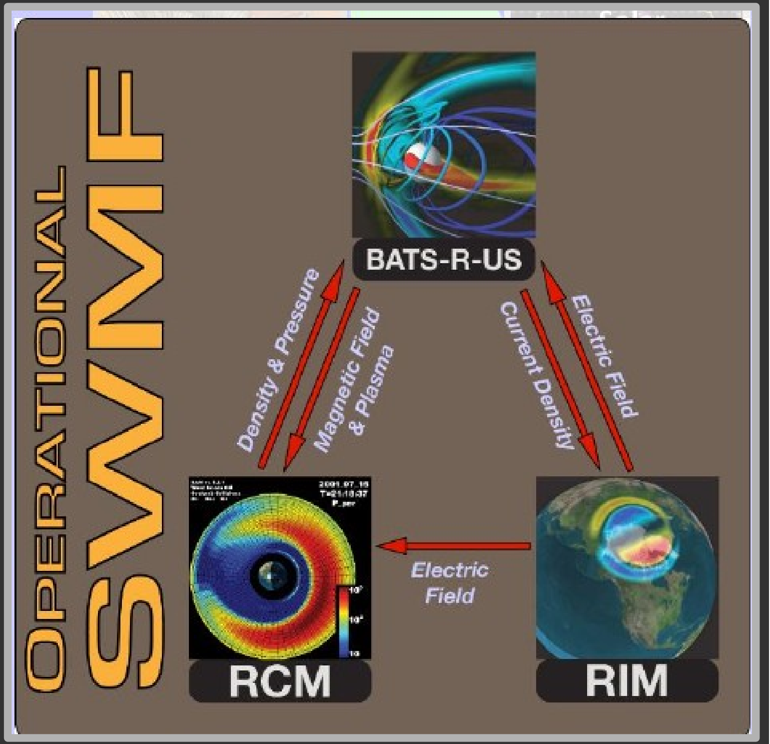
- $TPR = TP/P$ = True Positive Rate (also called Probability of Detection, Sensitivity, Hit Rate);
- $FPR = FP/N$ = False Positive Rate (also called Probability of False Detection);
- $TSS = TPR - FPR$ = True Skill Score.

The operational Geospace model

- MHD model of Earth’s magnetosphere
- 32 Re upstream to ~120 Re down tail
- U. Michigan’s Space Weather Modeling Framework (SWMF)
- Running every minute as long as solar wind data are available







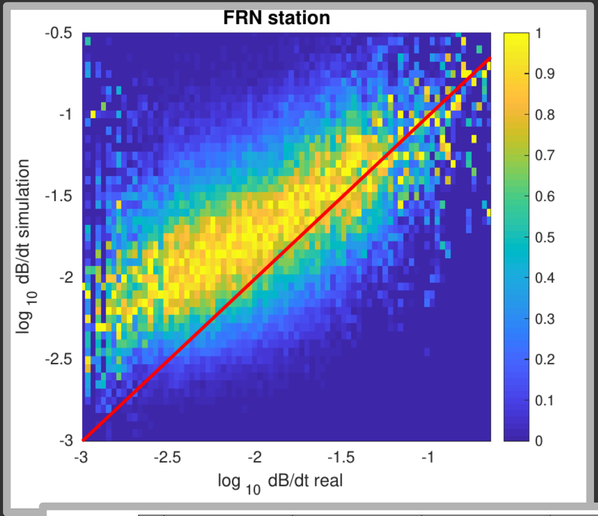
The data

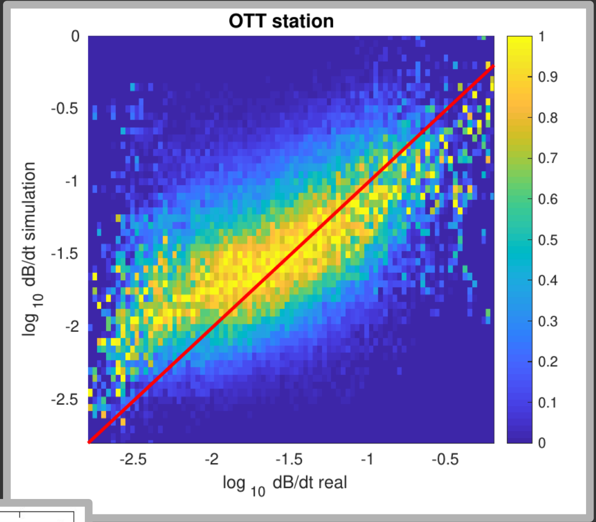
- In this work we focus on three stations:
  - Fresno (FRN, 43.12 N)
  - Ottawa (OTT, 54.88 N)
  - Iqaluit (IQA, 73.25 N)
- Real data: 1-min data for the period 2001-01-01 to 2019-05-05
  ~ 10M data points (INTERMAGNET)
- Archived simulations: 2017-05-28 to 2019-05-05
  ~ 1M data points

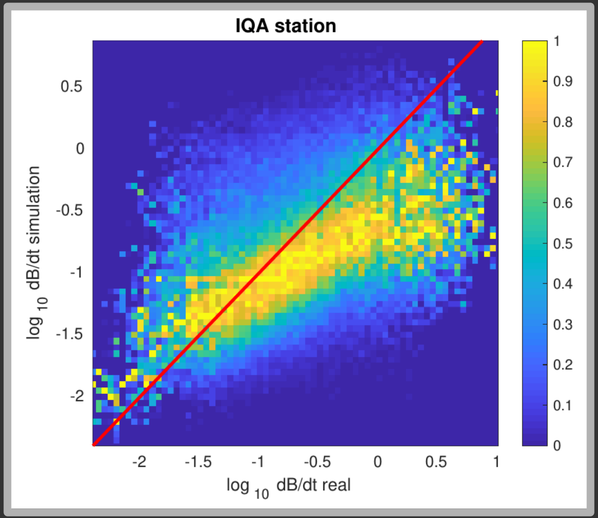
2001 – 2017 : training and validation

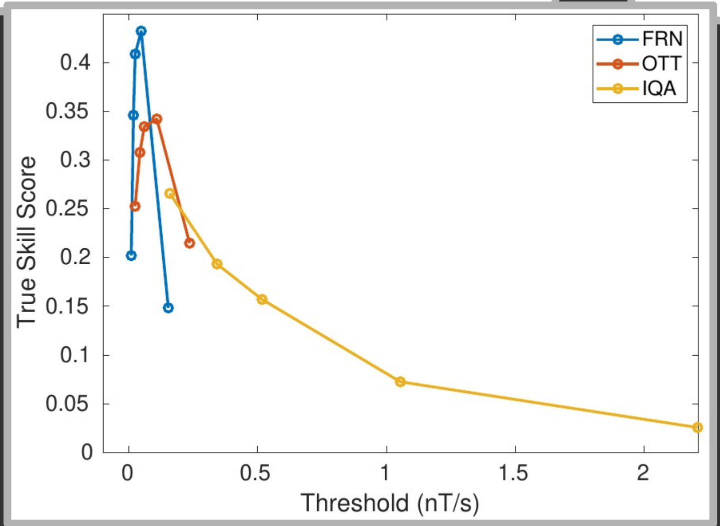
2018 – 2019 : testing

The baseline: TSS of Geospace model



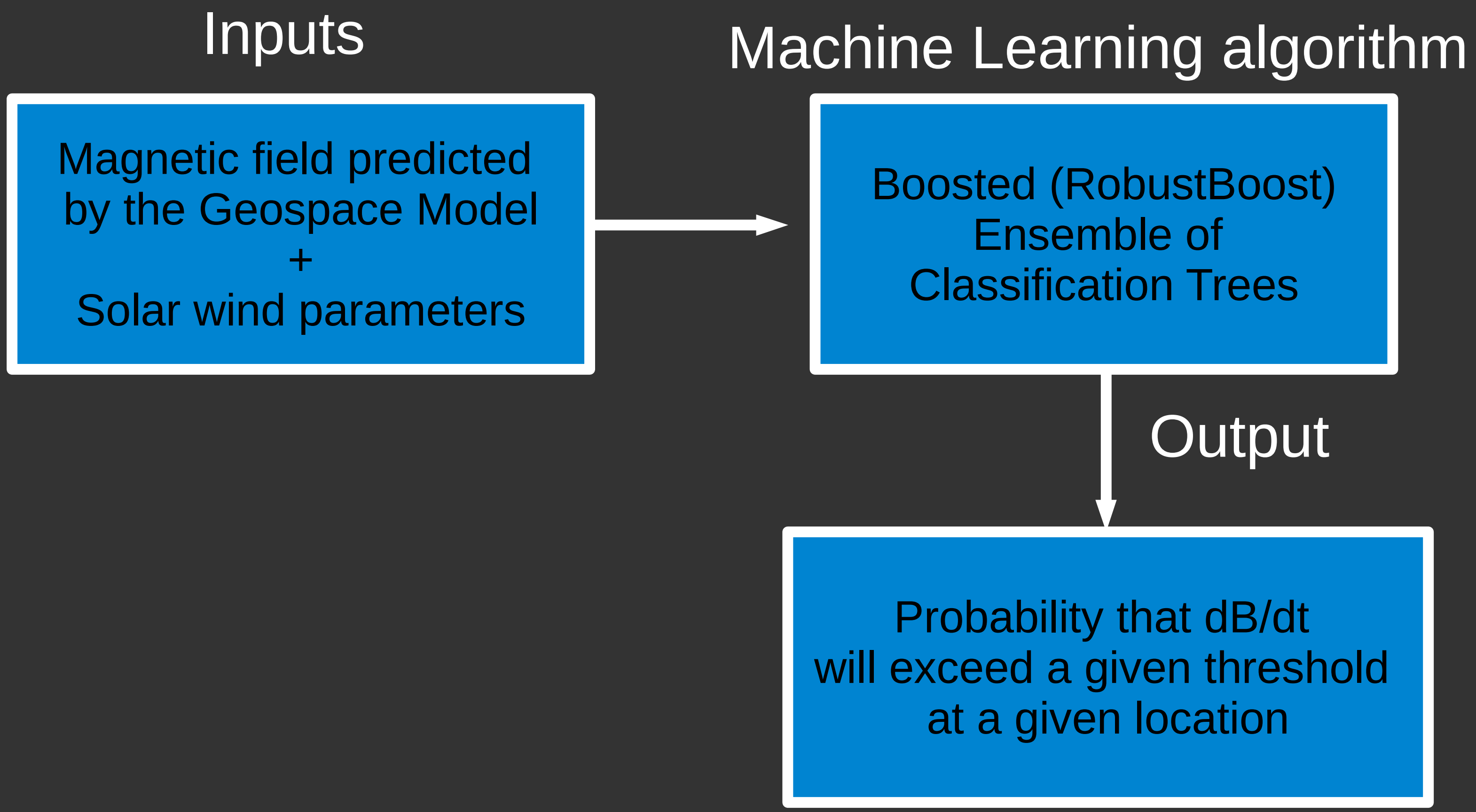






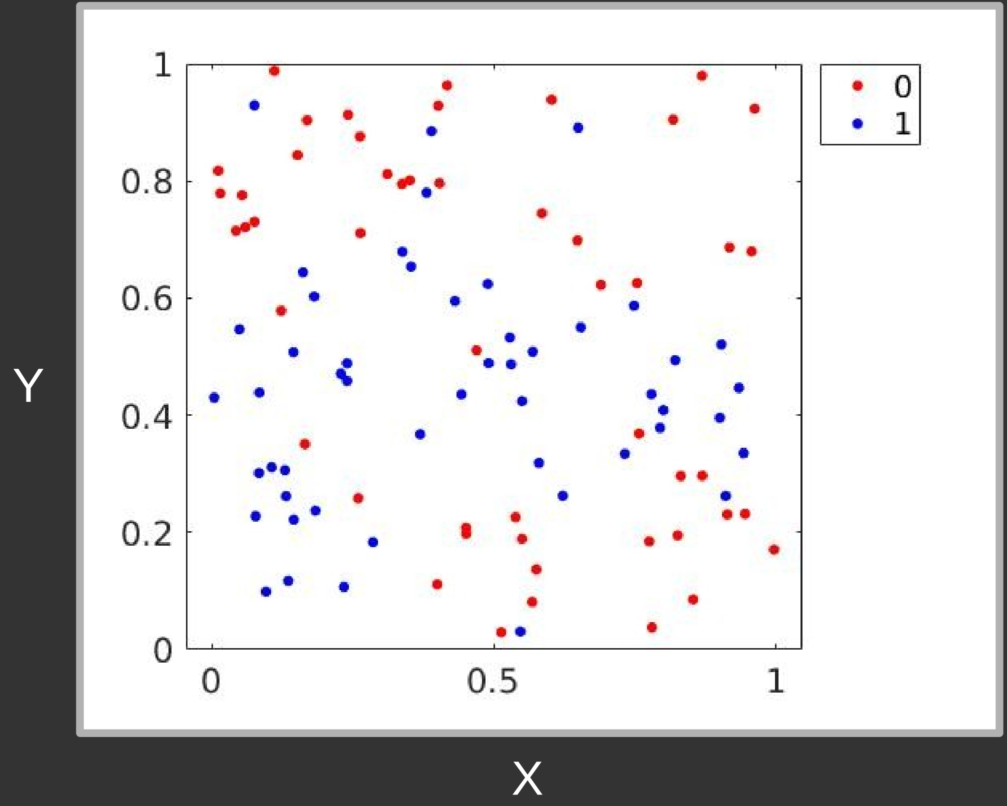
The goal of this work is to obtain a systematic improvement over this baseline for all stations and any threshold

Basic idea: Grey-box approach!



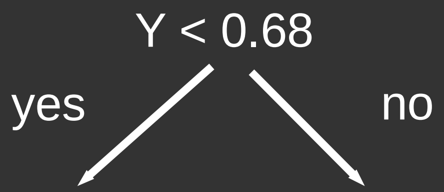
What is a decision tree for classification?

A decision tree partitions your input space in the optimal way to find cluster of labels

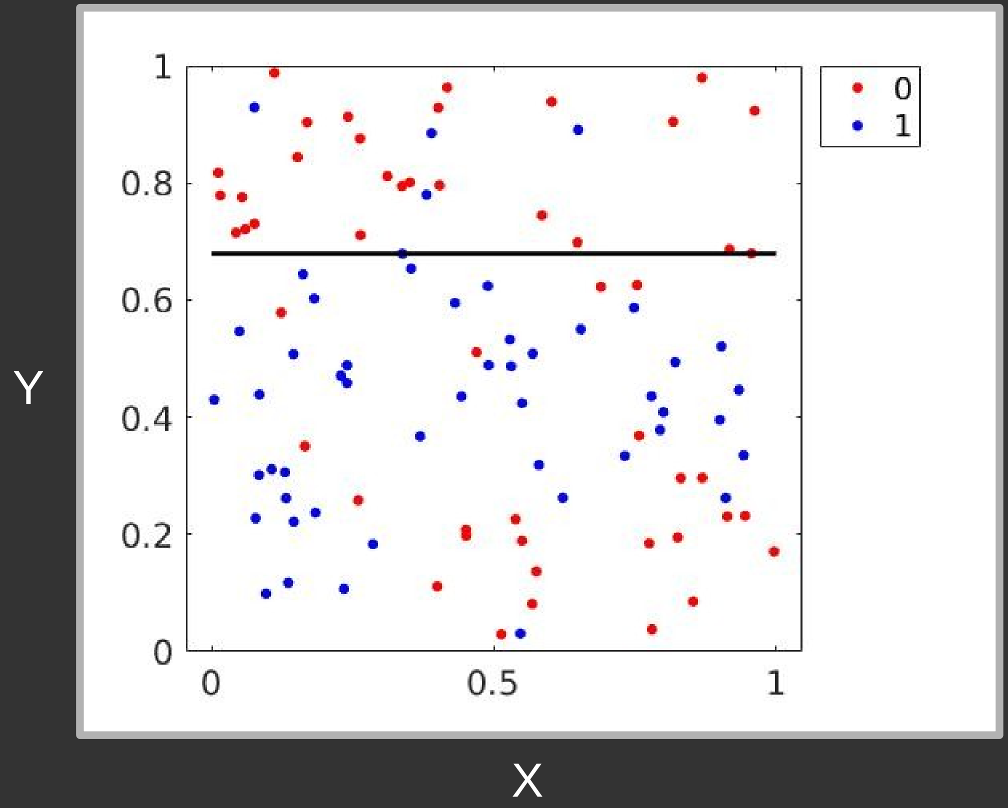


What is a decision tree for classification?

A decision tree partitions your input space in the optimal way to find cluster of labels

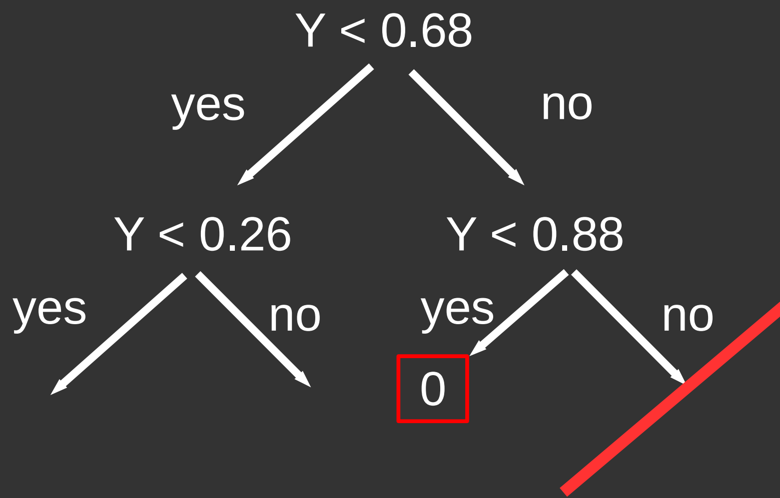


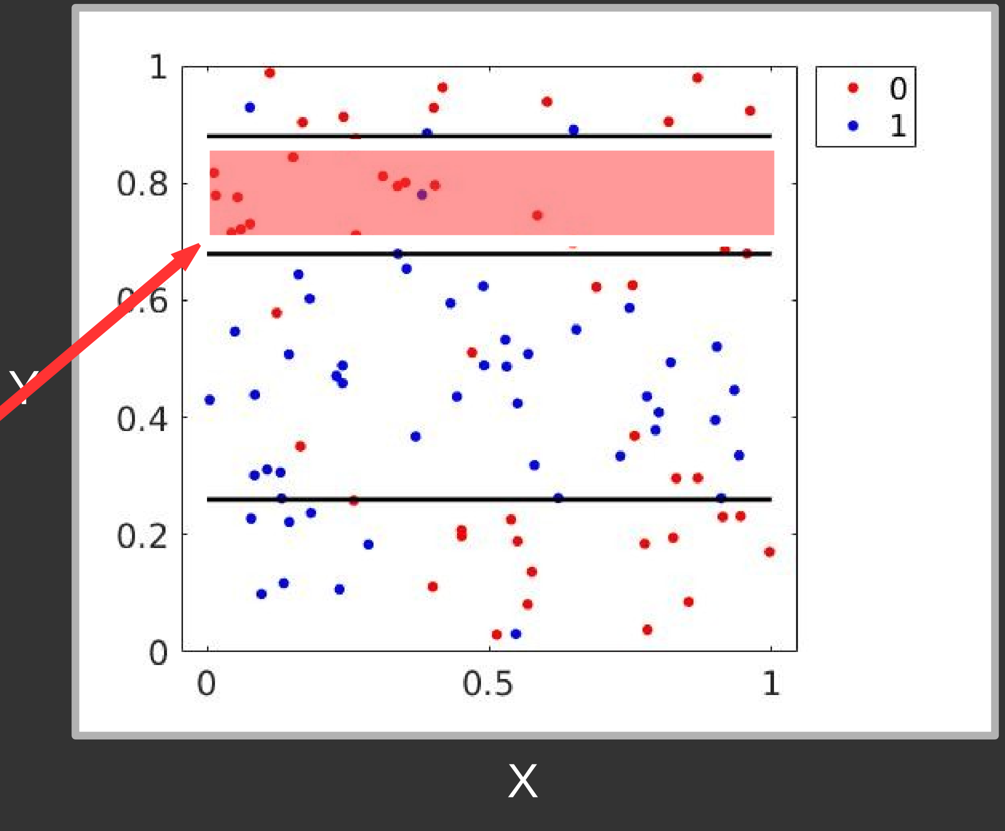
The cuts are always parallel to the axes and try to maximize the ratio between blue and red in each subset



What is a decision tree for classification?

A decision tree partitions your input space in the optimal way to find cluster of labels

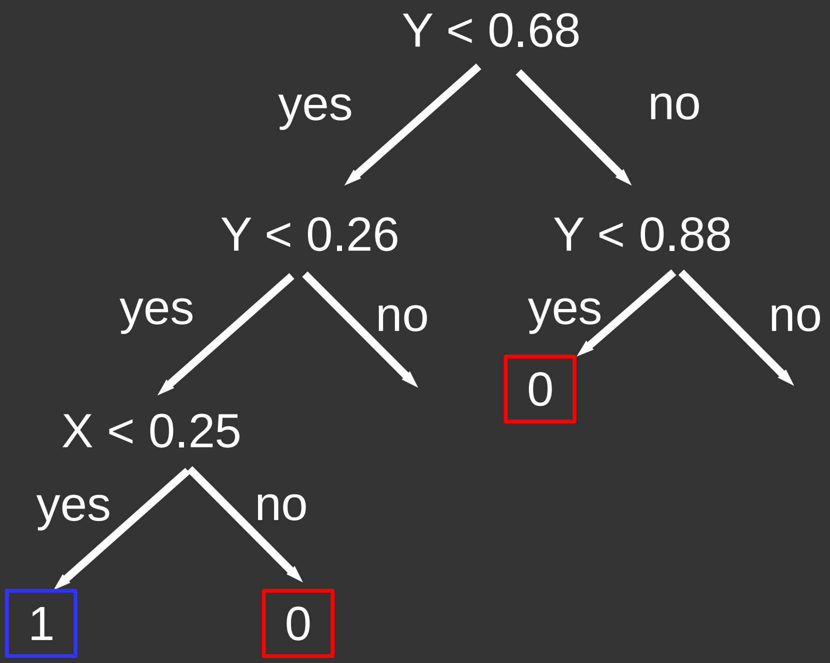


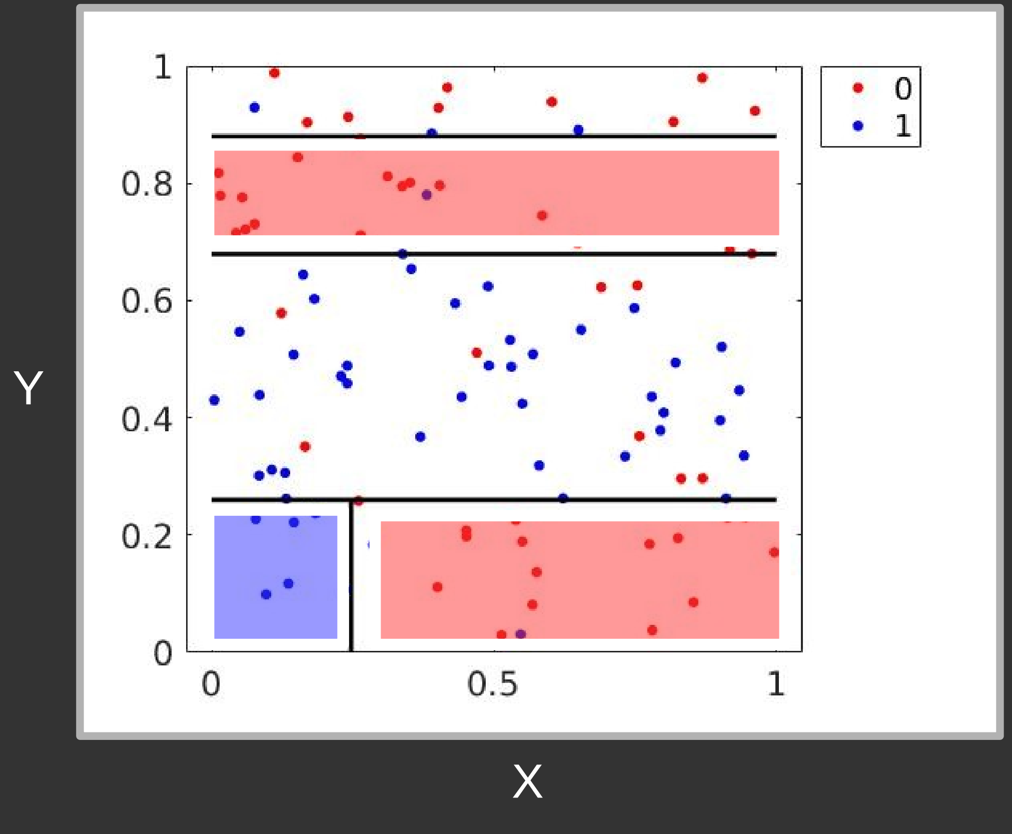


The ratio here is 2 blue vs 17 red.
No possible further cut can do better than this

What is a decision tree for classification?

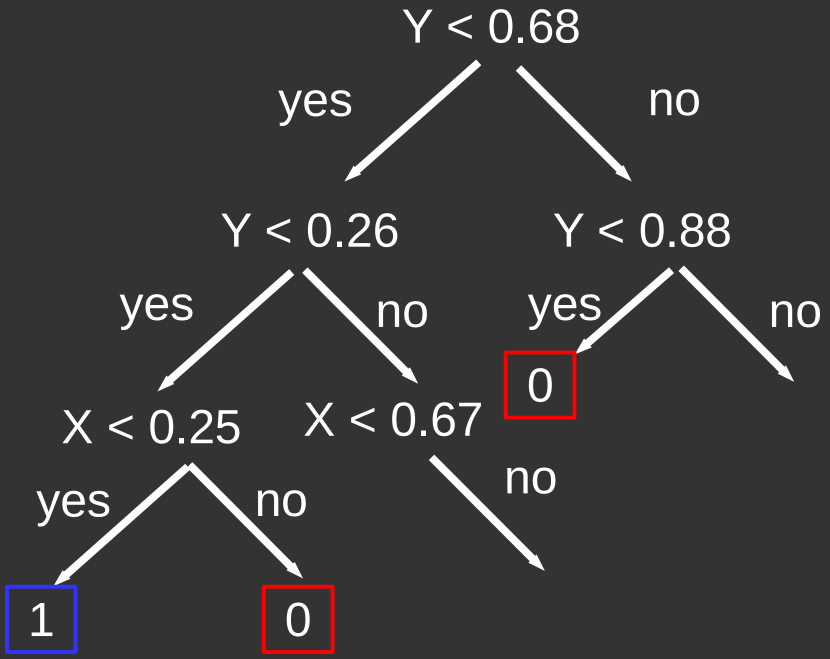
A decision tree partitions your input space in the optimal way to find cluster of labels

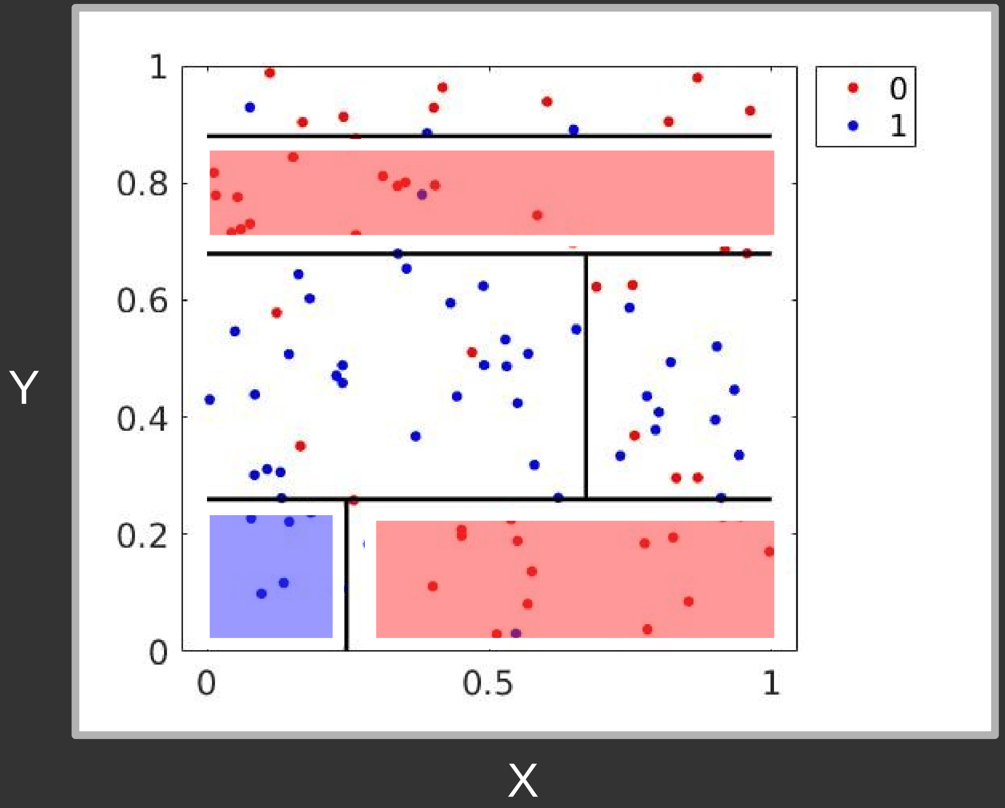




What is a decision tree for classification?

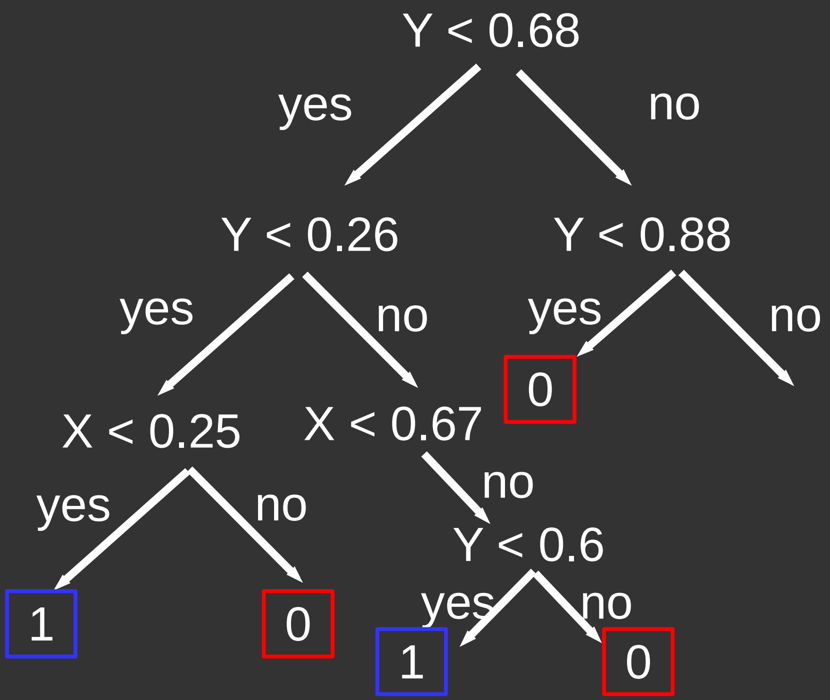
A decision tree partitions your input space in the optimal way to find cluster of labels

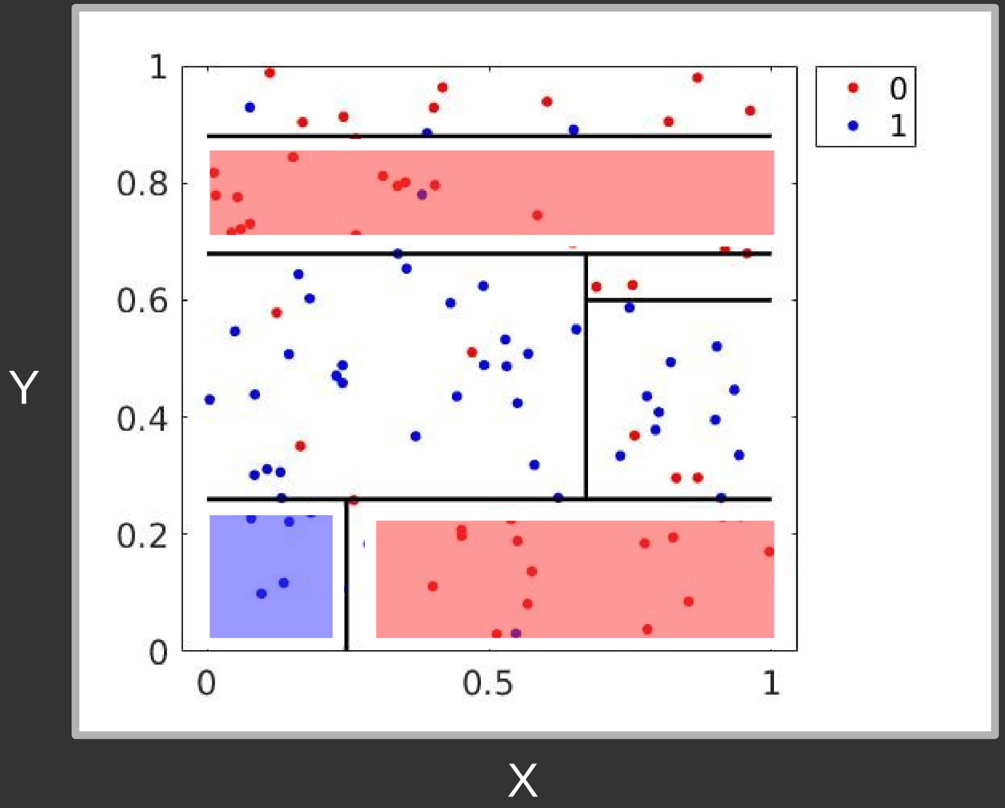




What is a decision tree for classification?

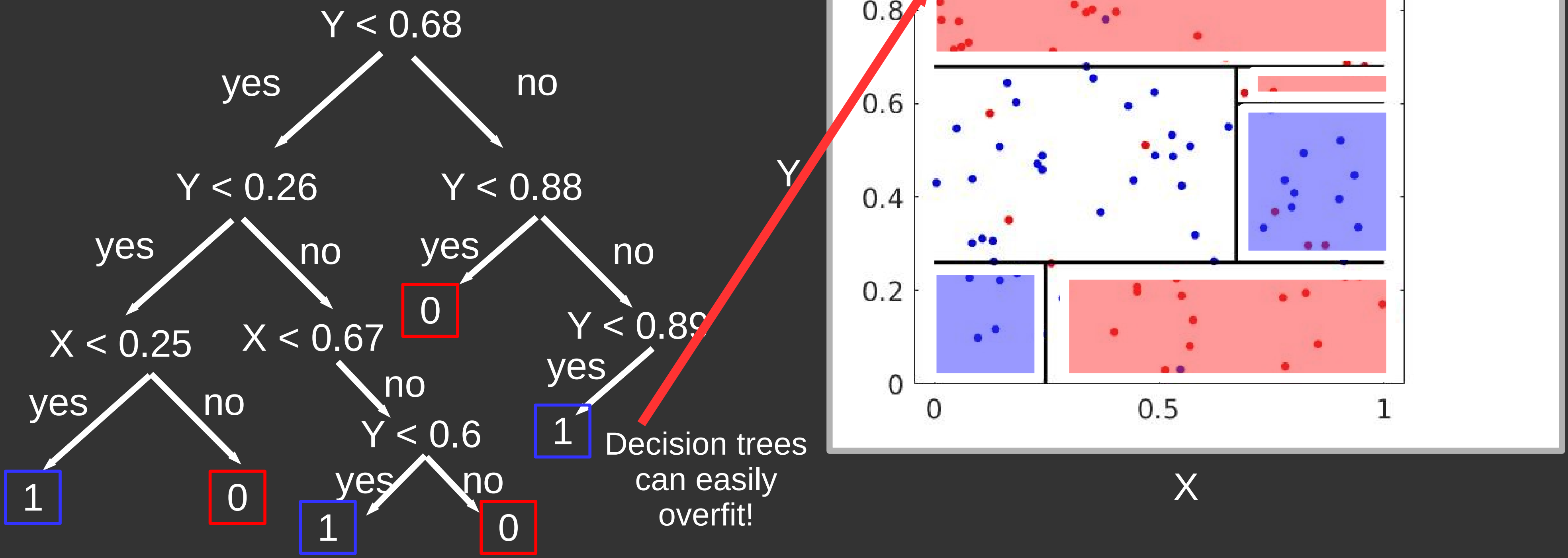
A decision tree partitions your input space in the optimal way to find cluster of labels





What is a decision tree for classification?

A decision tree partitions your input space in the optimal way to find cluster of labels



What is boosting?

- ‘Weak learners’ are trees that are grown only to a few levels
- Each weak learner performs only slightly better than random chance. However, an ensemble of weak learners can perform much better than a full-grown tree.
- An iterative algorithm boosts a weak learner increasingly focusing on the data that was mis-classified at the previous iteration

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of computer and system sciences*, 55 (1), 119–139.

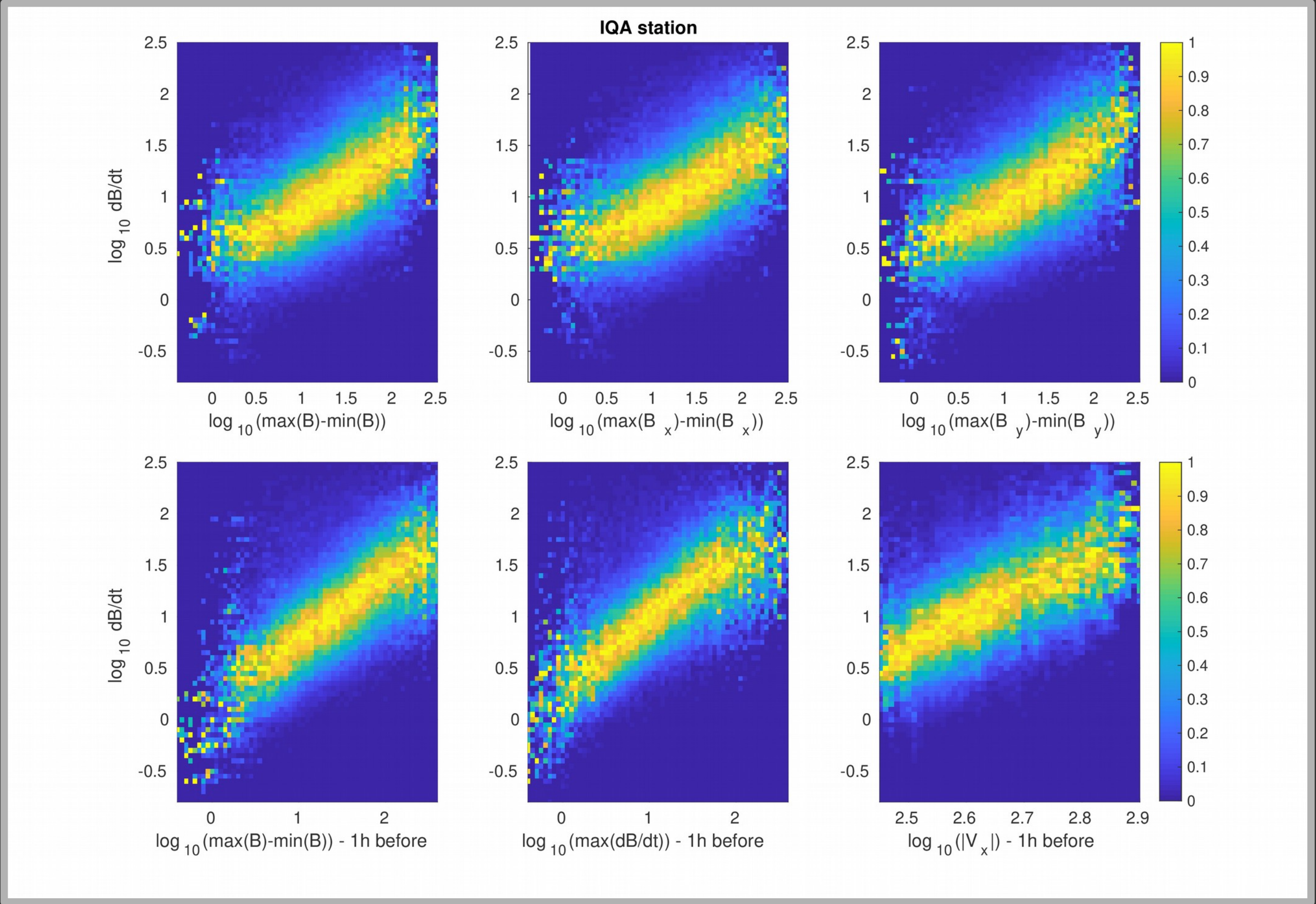
Freund, Y. (2009). A more robust boosting algorithm. *arXiv:0905.2138* .

Feature selection

1. log10 (max(B)−min(B)) (obtained from the Geospace simulation output)
2. log10 (max(Bx) − min(Bx)) (same as 1. but the x-component only)
3. log10 (max(By)−min(By)) (same as 1. but the y-component only)
4. log10 (max(B)−min(B)) - 1hr before (same as 1. but calculated in the time window 1 hour preceding the target window, obtained from magnetometer)
5. log10 (max(dB/dt)) - 1hr before (the same quantity as the target, but calculated in the time window 1 hour preceding the target window, obtained from magnetometer data)
6. log10 (|Vx |) - 1hr before (the x-component of the solar wind speed, measured one hour before the target window)

Tóth, G.,et al. (2014). Predicting the time derivative of local magnetic perturbations. *Journal of Geophysical Research*, 119 (1), 310–321.

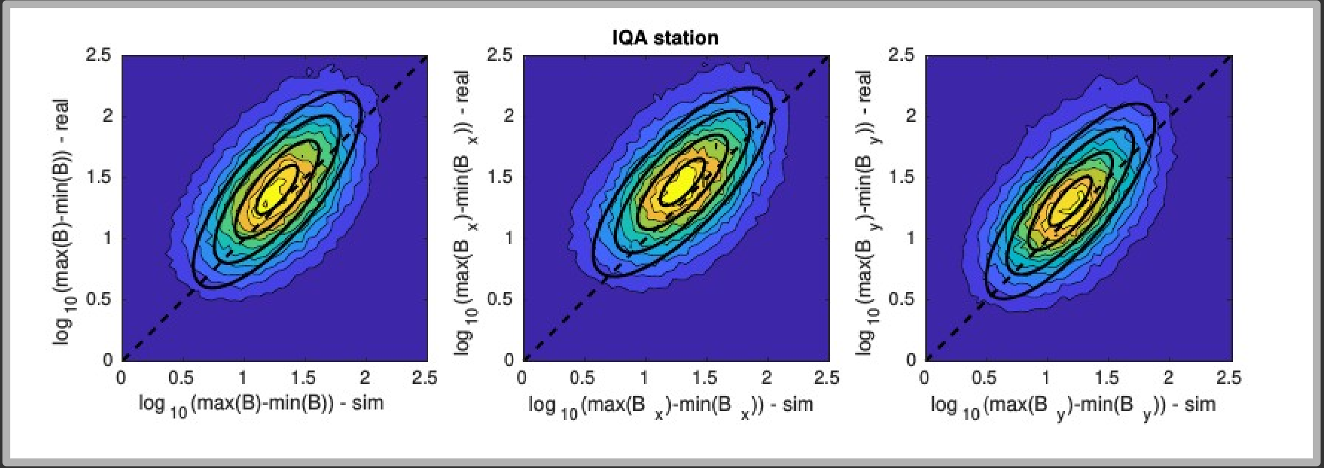
Feature selection

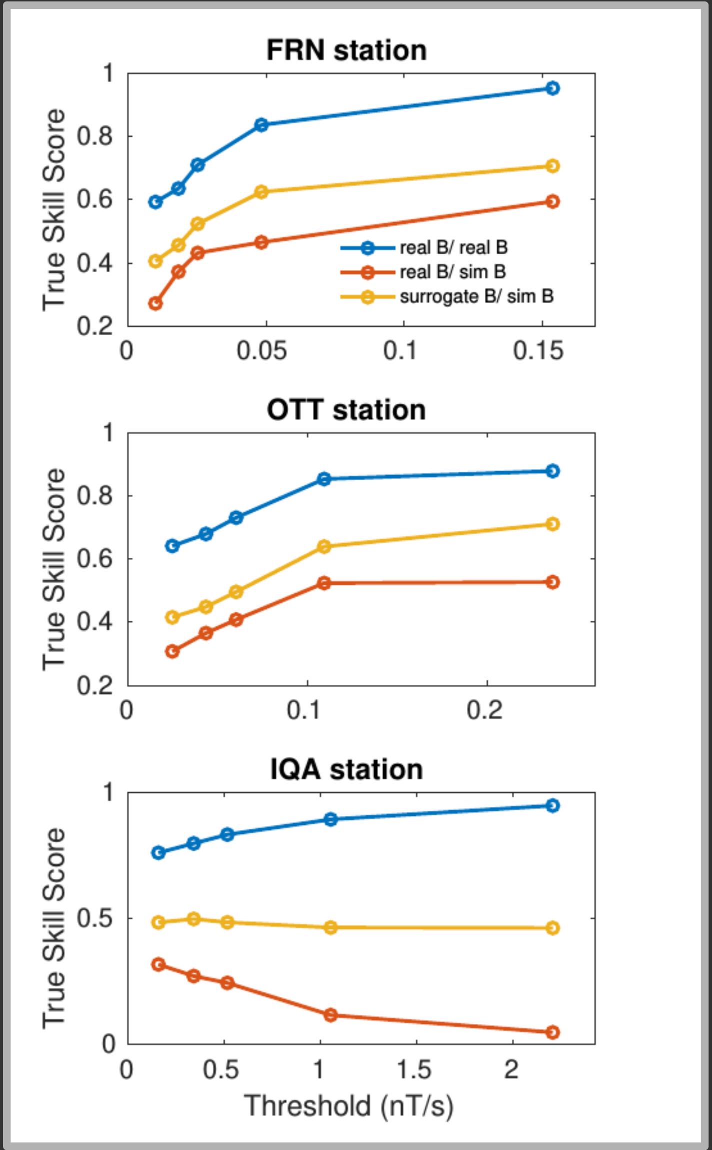


Surrogate Geospace data
(a small technicality...)

2001 – 2017 : training and validation

- We want to train our model on a time period for which we do not have the runs of the operational Geospace model
- We produce surrogate Geospace output by assuming a multivariate Gaussian distribution in the input-output space





Results: True Skill Score

Blue line → Theoretical upper limit in the case the Geospace model would produce perfect predictions

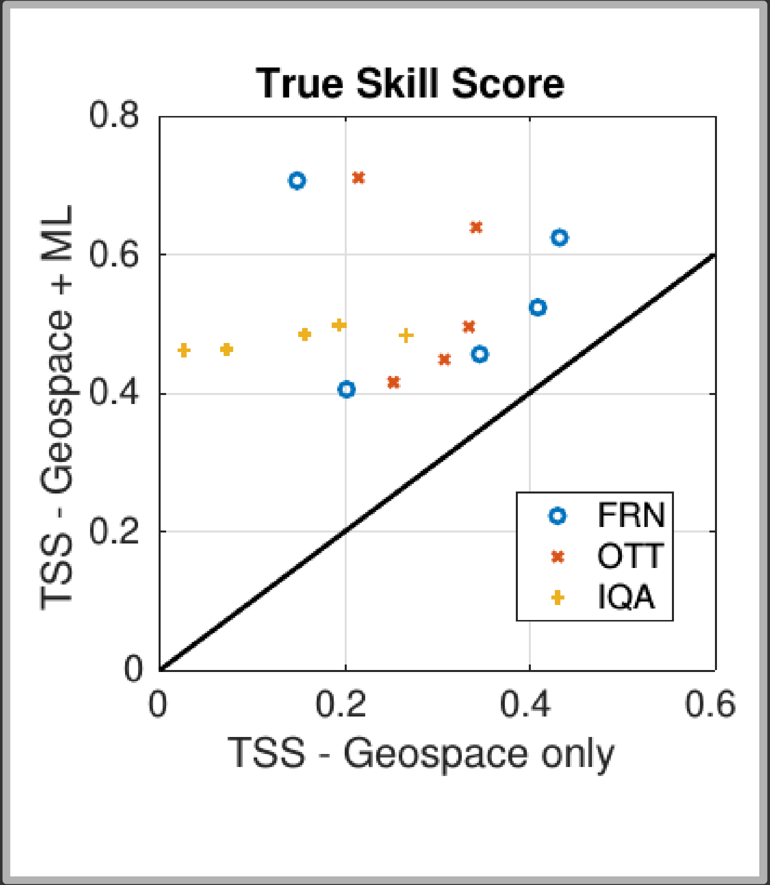
Red line → The result without producing surrogate model (train using real data but predict using Geospace)

Yellow line → The results of this work!

Comparing white vs grey-box models

- The Geospace + ML method consistently enhances the score for all three stations, for any threshold
- The method will be extended to more stations
- We plan to make it operational

The paper will be on arXiv by AGU
(~ 2 weeks !!)



e.camporeale@noaa.gov