JOHANNES GUTENBERG UNIVERSITÄT MAINZ

BINARY
Big Data in Atmospheric Physics

Turbulence Classification

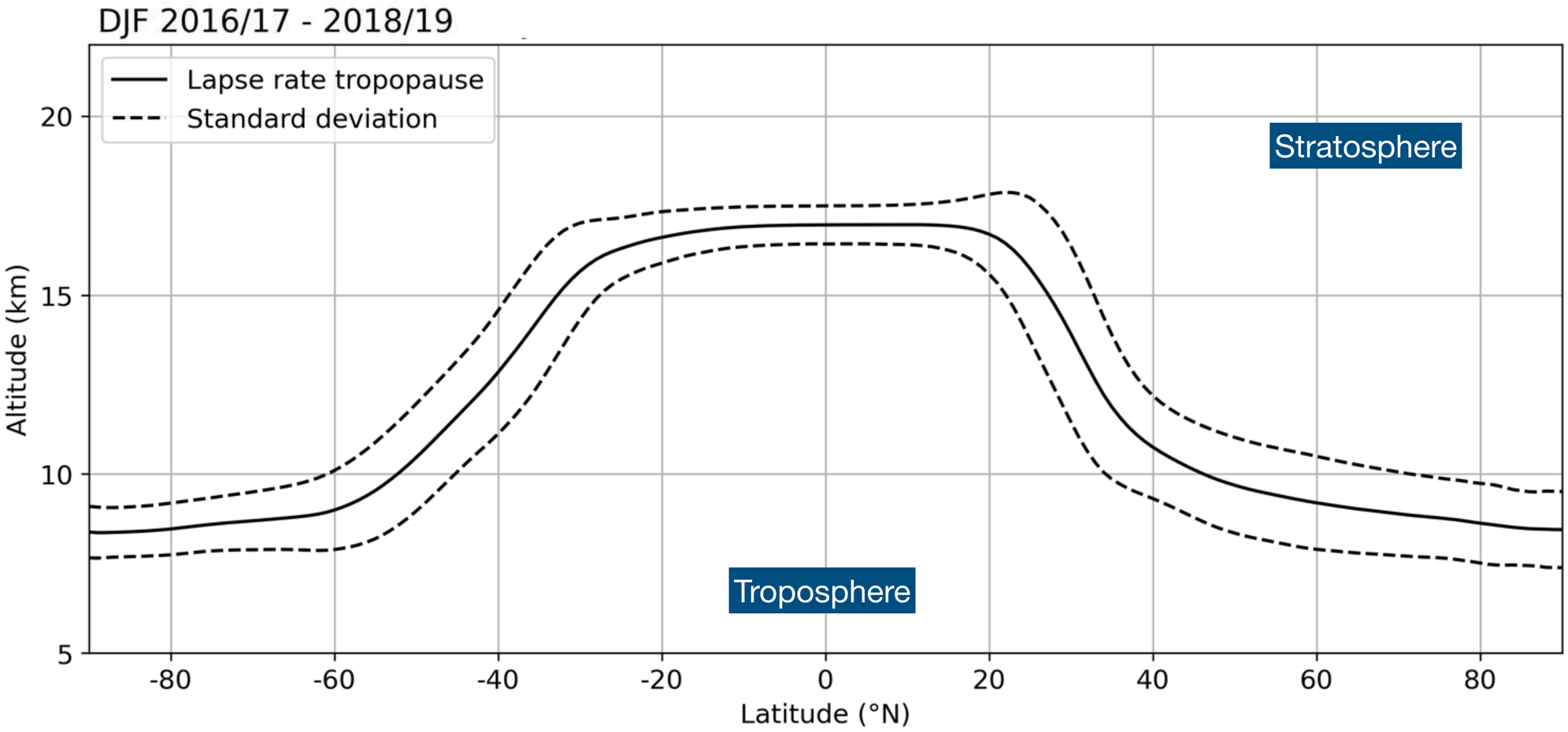
In the extratropical UTLS, using:

- **automated EDR reports**
- **ERA5 diagnostics**
- **a random forest classifier**

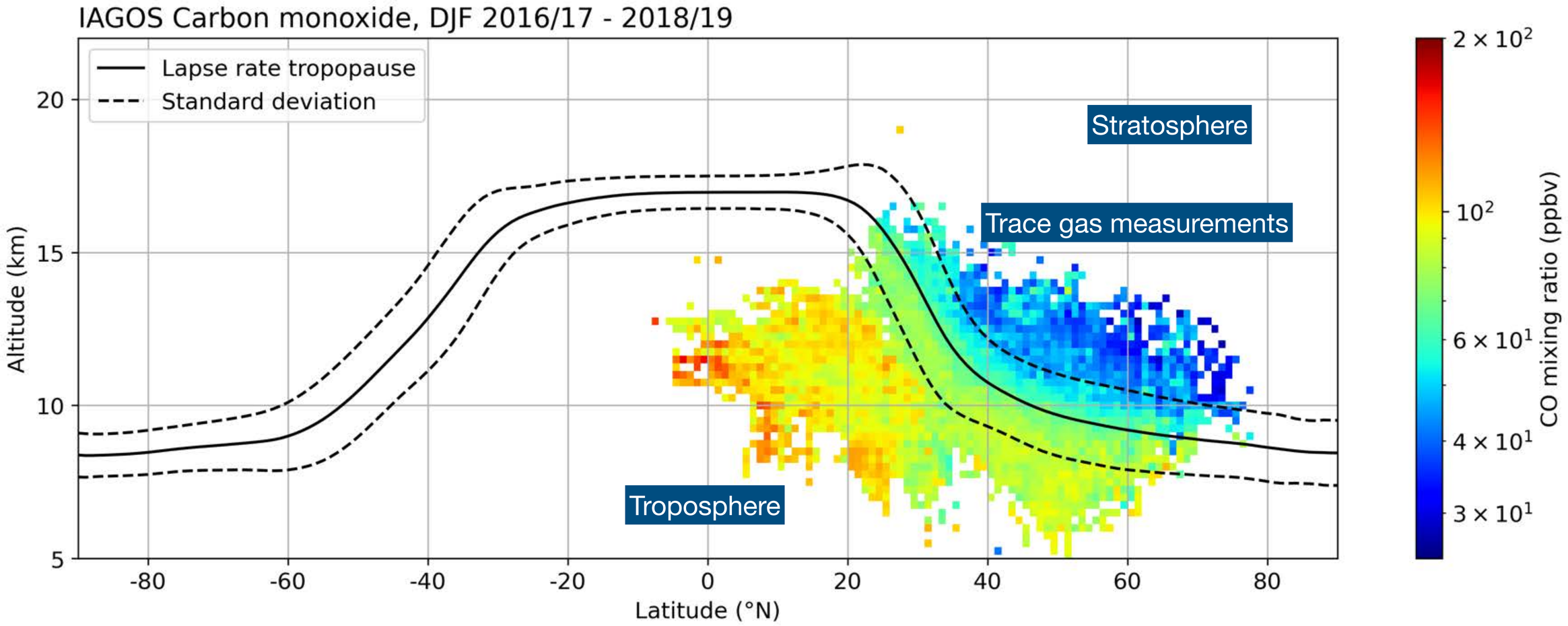
FISAPS Workshop
30 August - 1 September 2023
NCAR, Boulder Colorado

Thorsten Kaluza, Peter Hoor, Daniel Kunkel

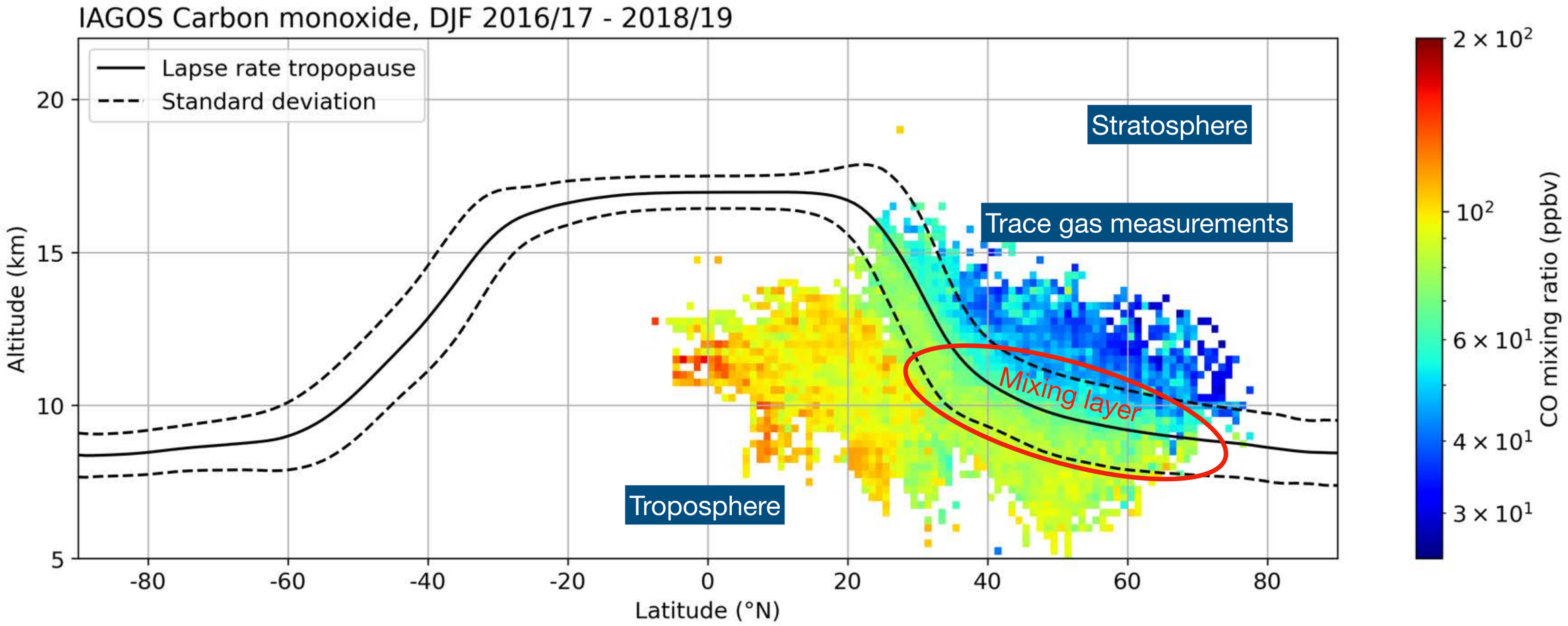
Motivation: UTLS



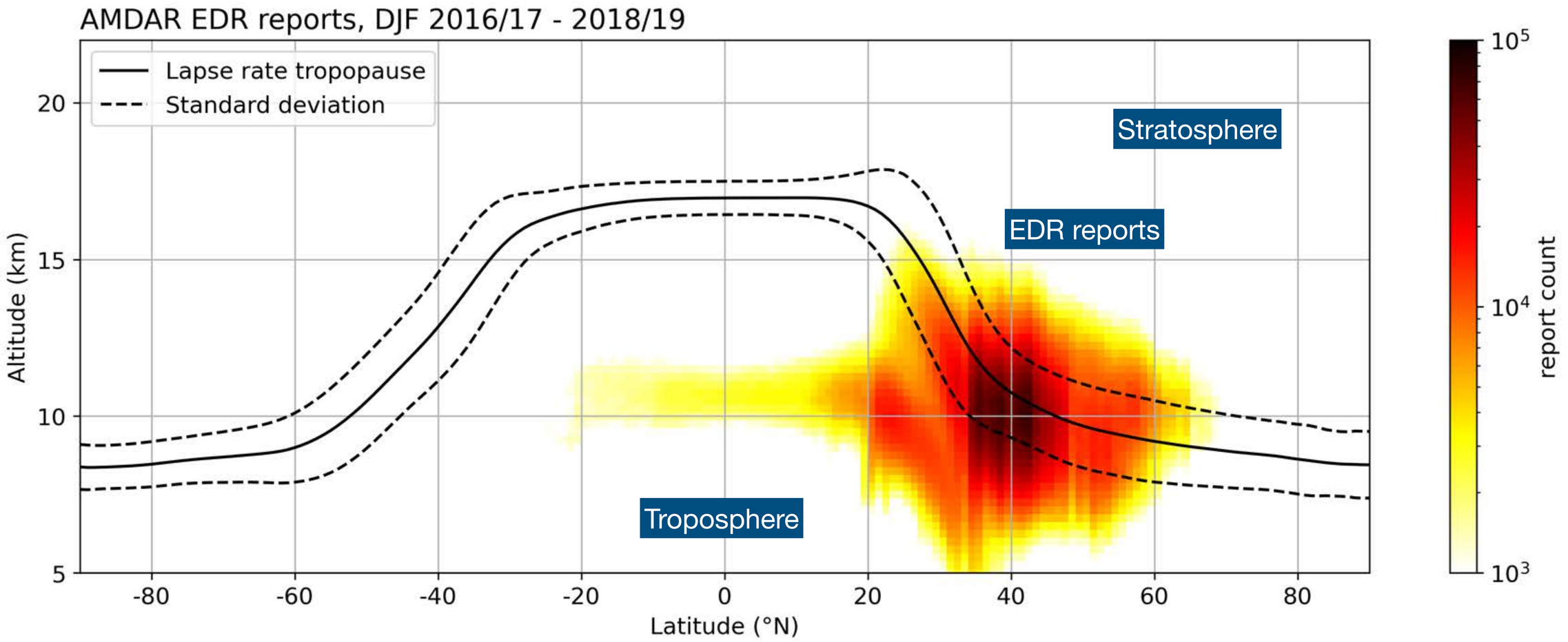
Motivation: Trace gas gradients



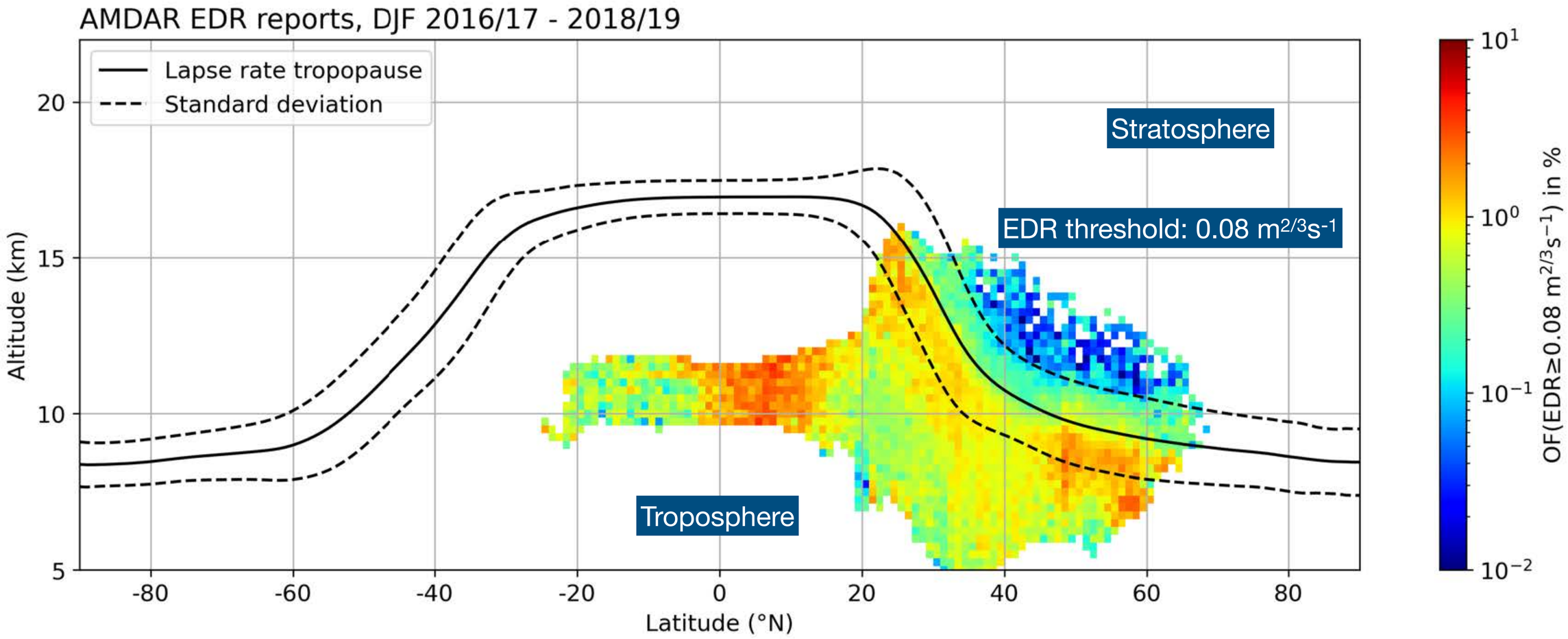
Motivation: Trace gas gradients



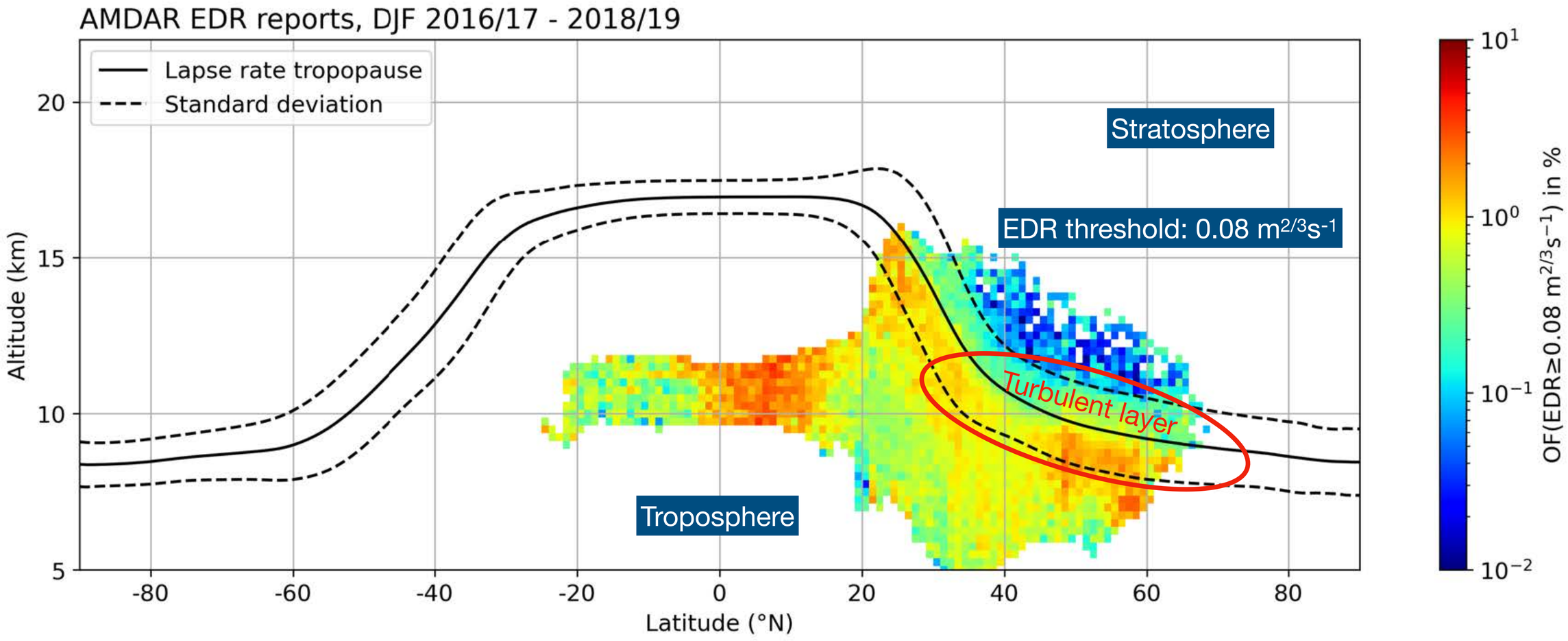
Motivation: Turbulent mixing



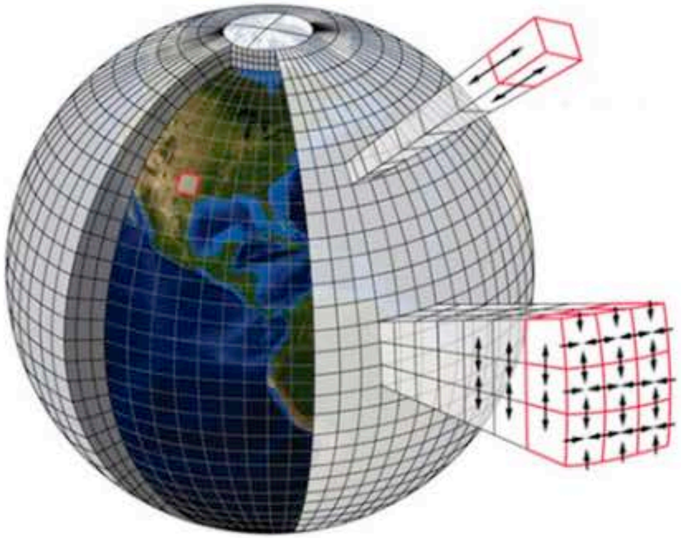
Motivation: Turbulent mixing



Motivation: Turbulent mixing



Data set



ERA5:

- Global, hourly
- Model level in vertical
- 0.25° horizontal

→ **Turbulence diagnostics**

2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020

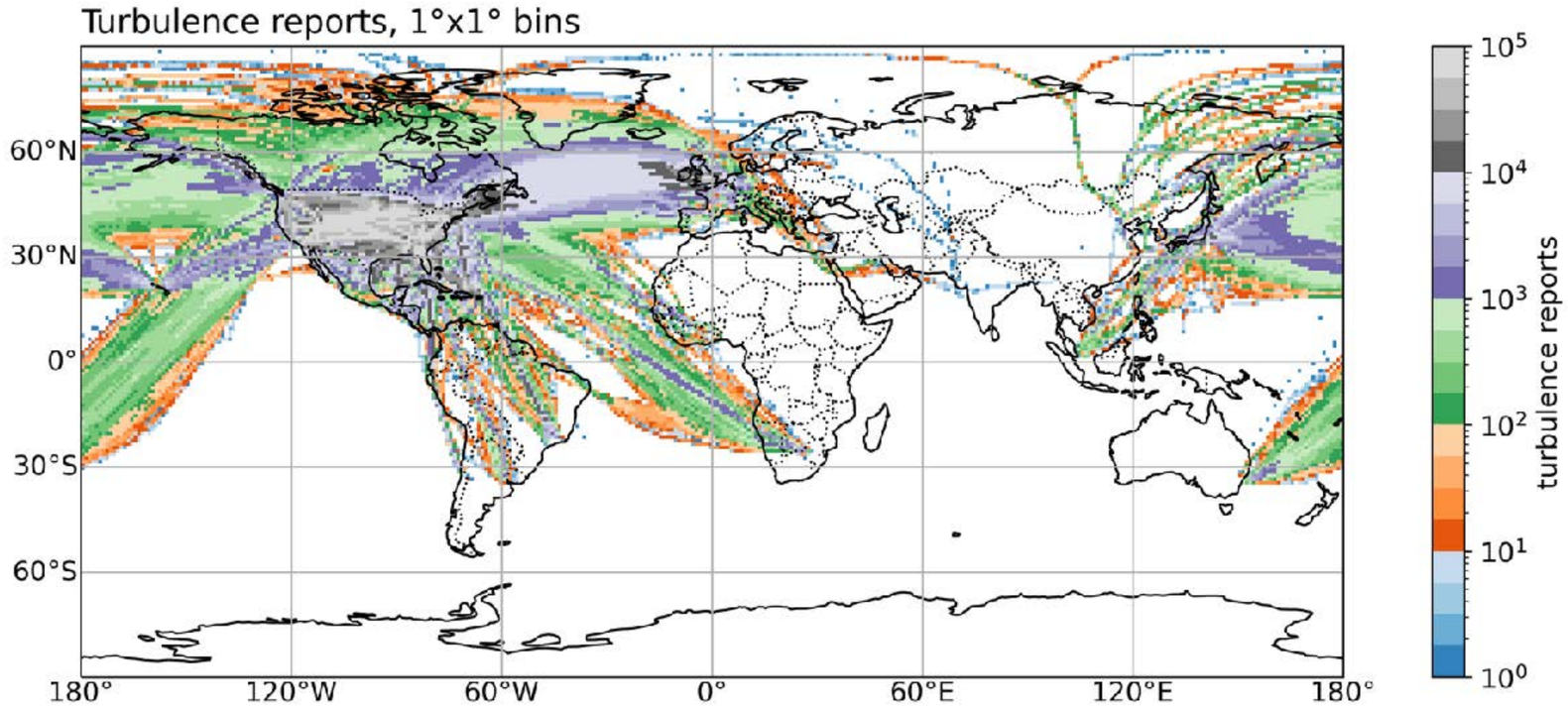
Ground truth ←

AMDAR EDR reports:

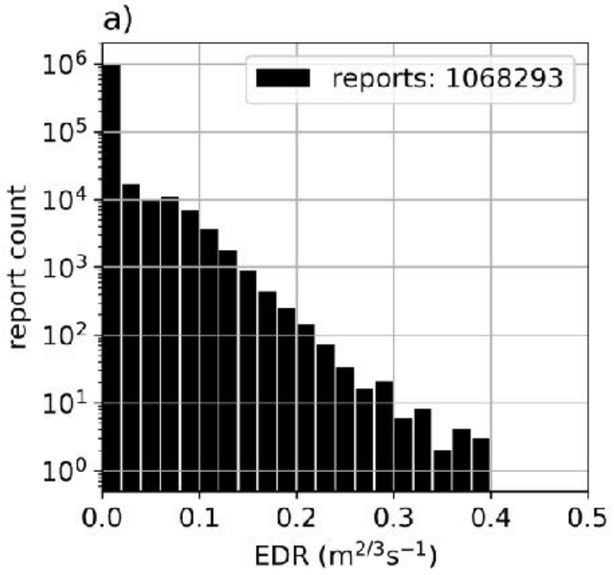
- 12.2016-12.2019
  → 63 million reports
- No JJA

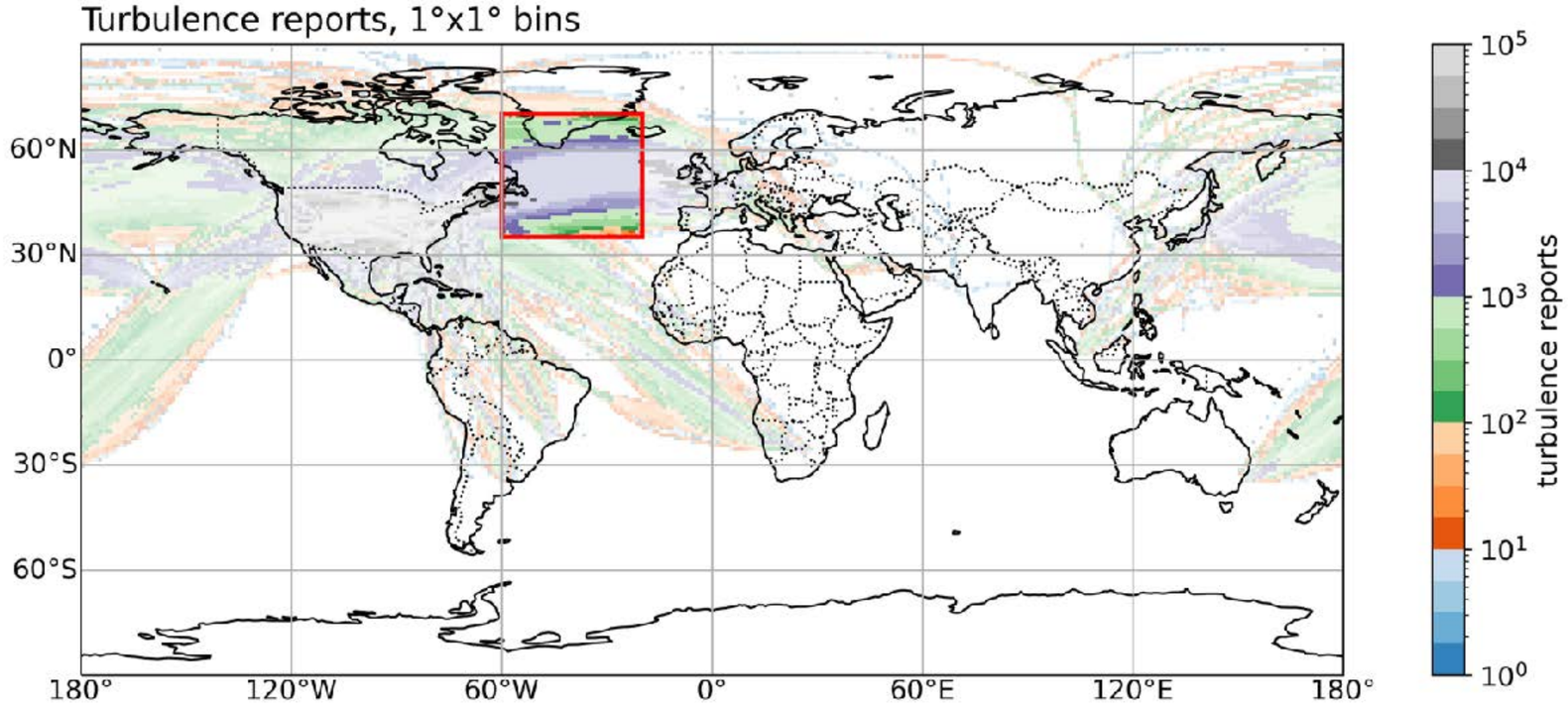
2016 2017 2018 2019 2020

ERA5 diagnostics and ROC curves

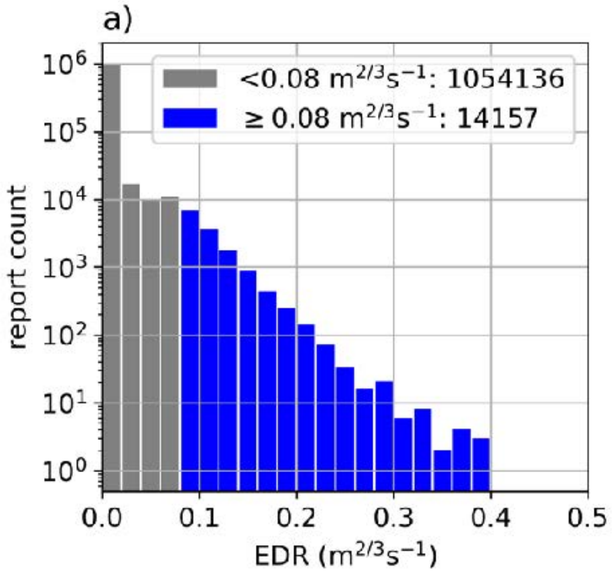


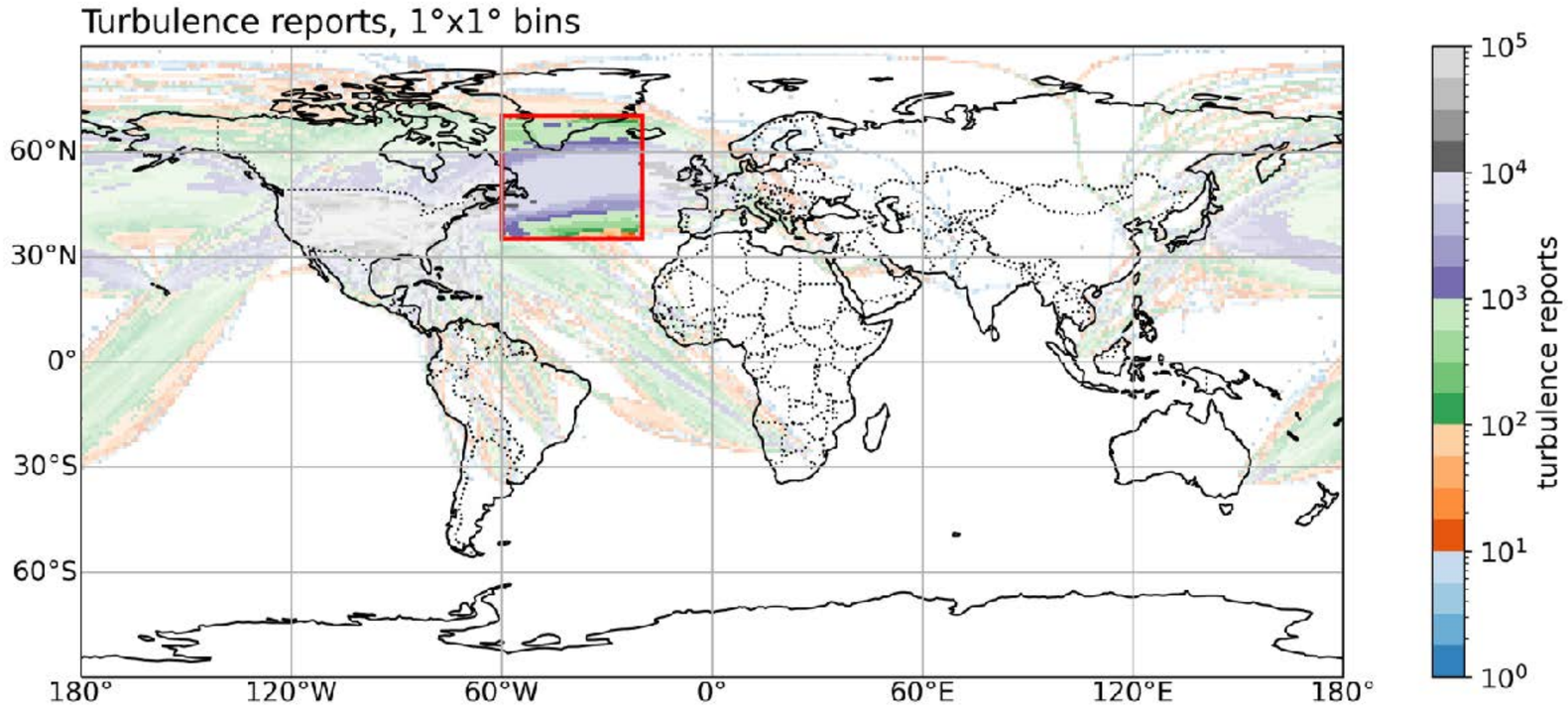
ERA5 diagnostics and ROC curves



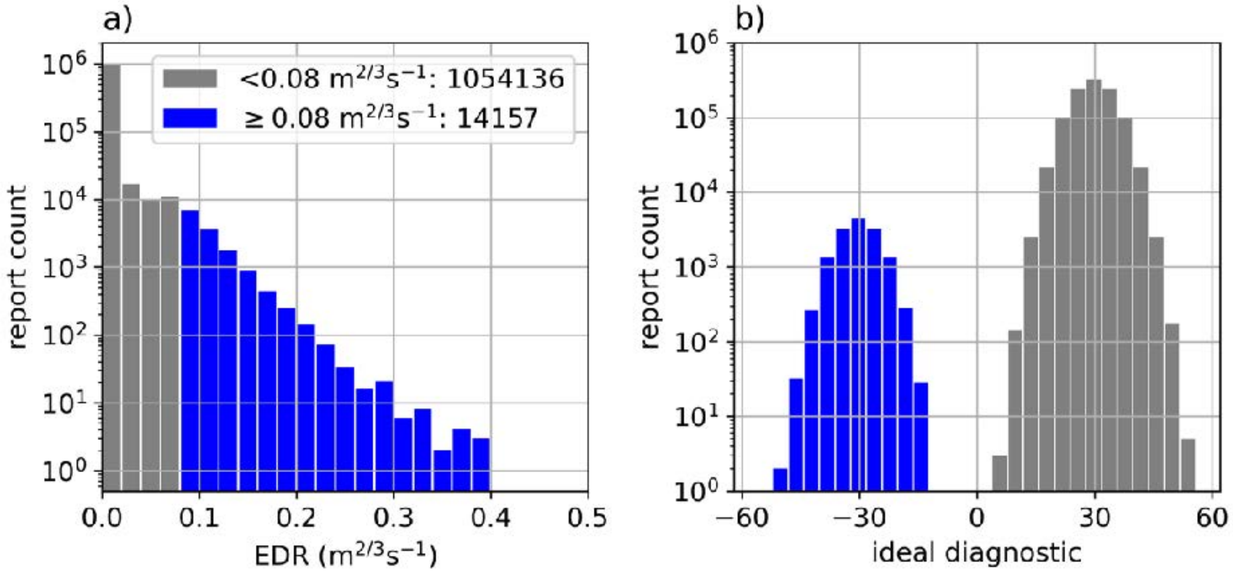


ERA5 diagnostics and ROC curves

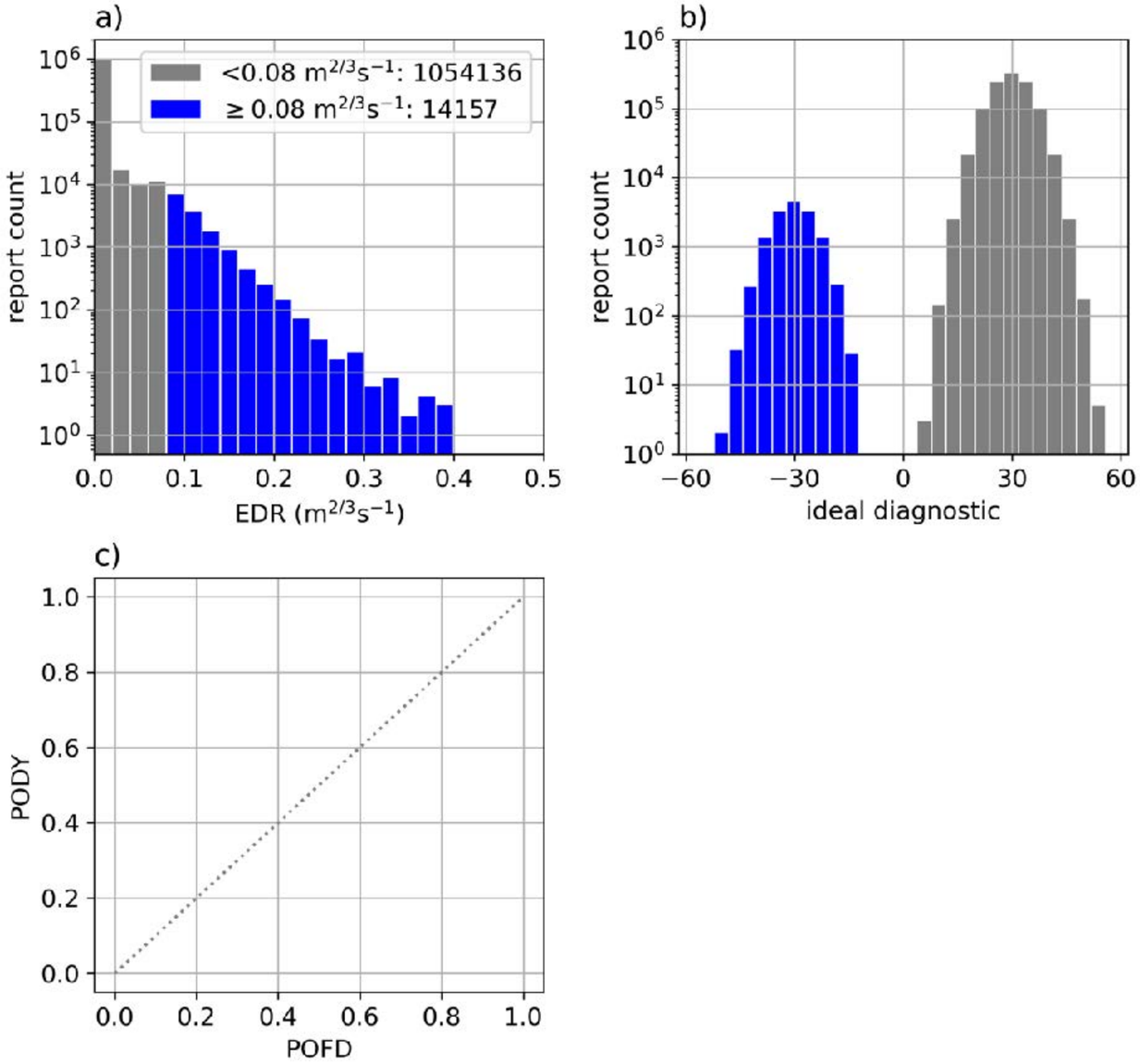




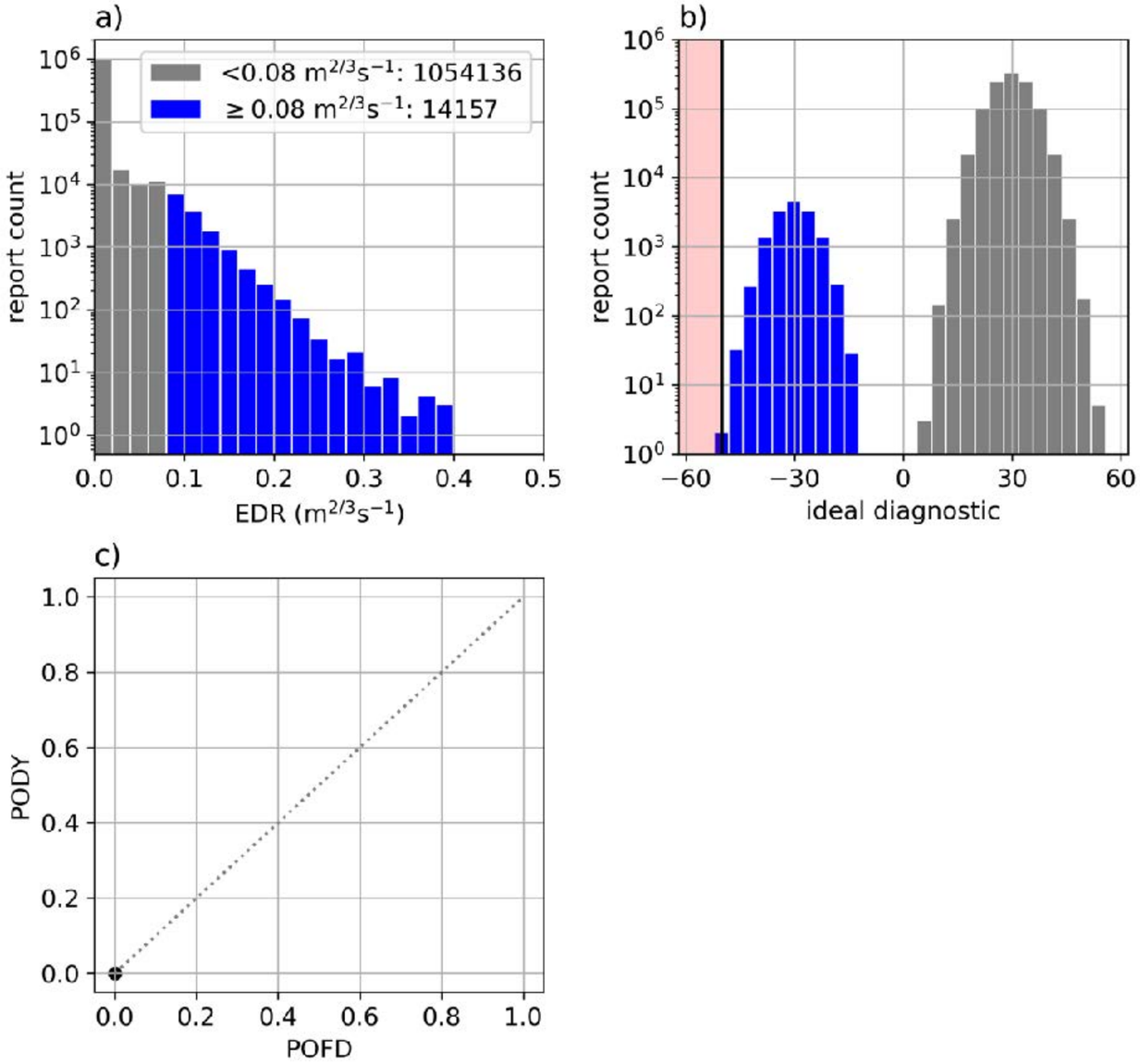
ERA5 diagnostics and ROC curves



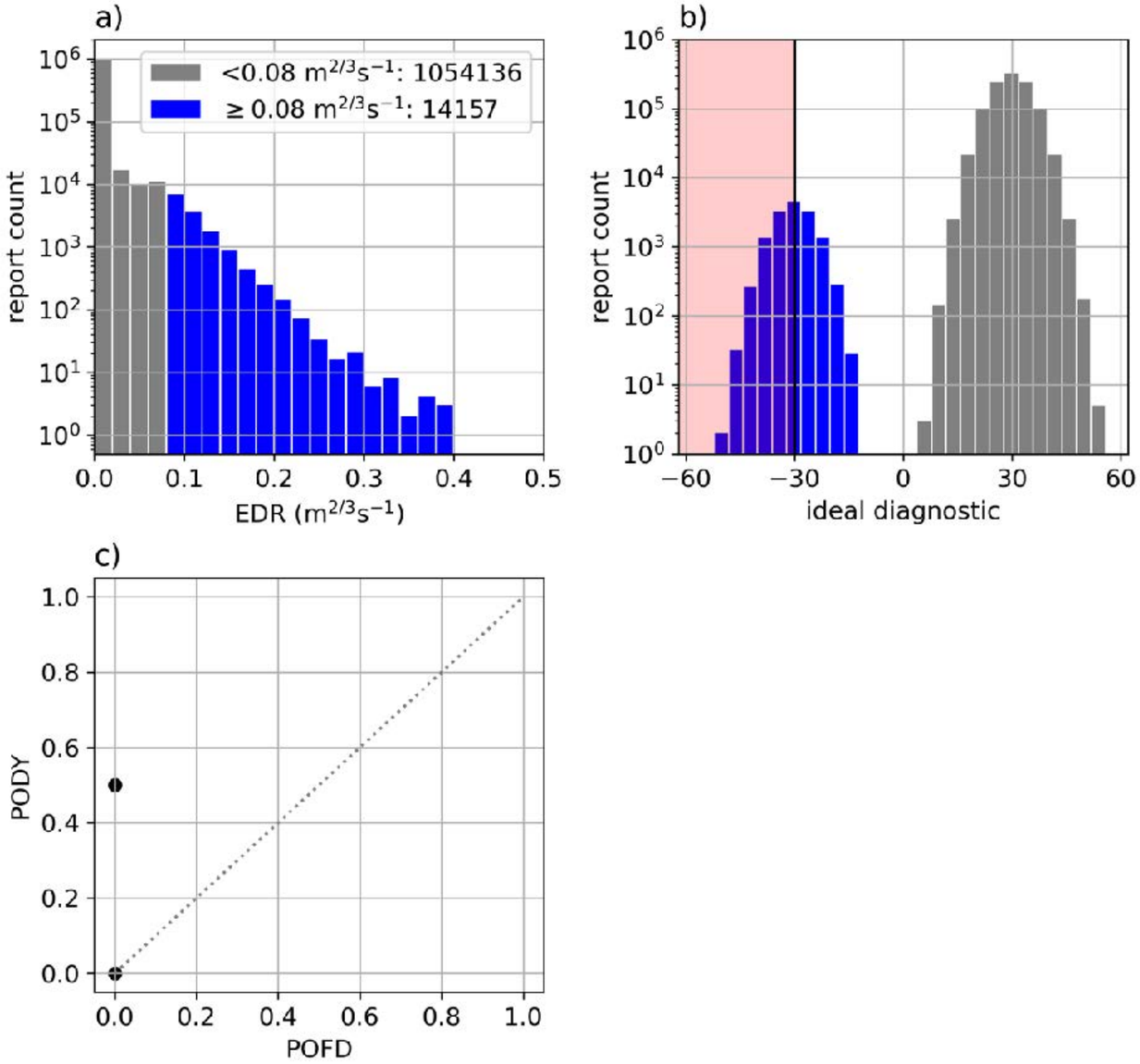
ERA5 diagnostics and ROC curves



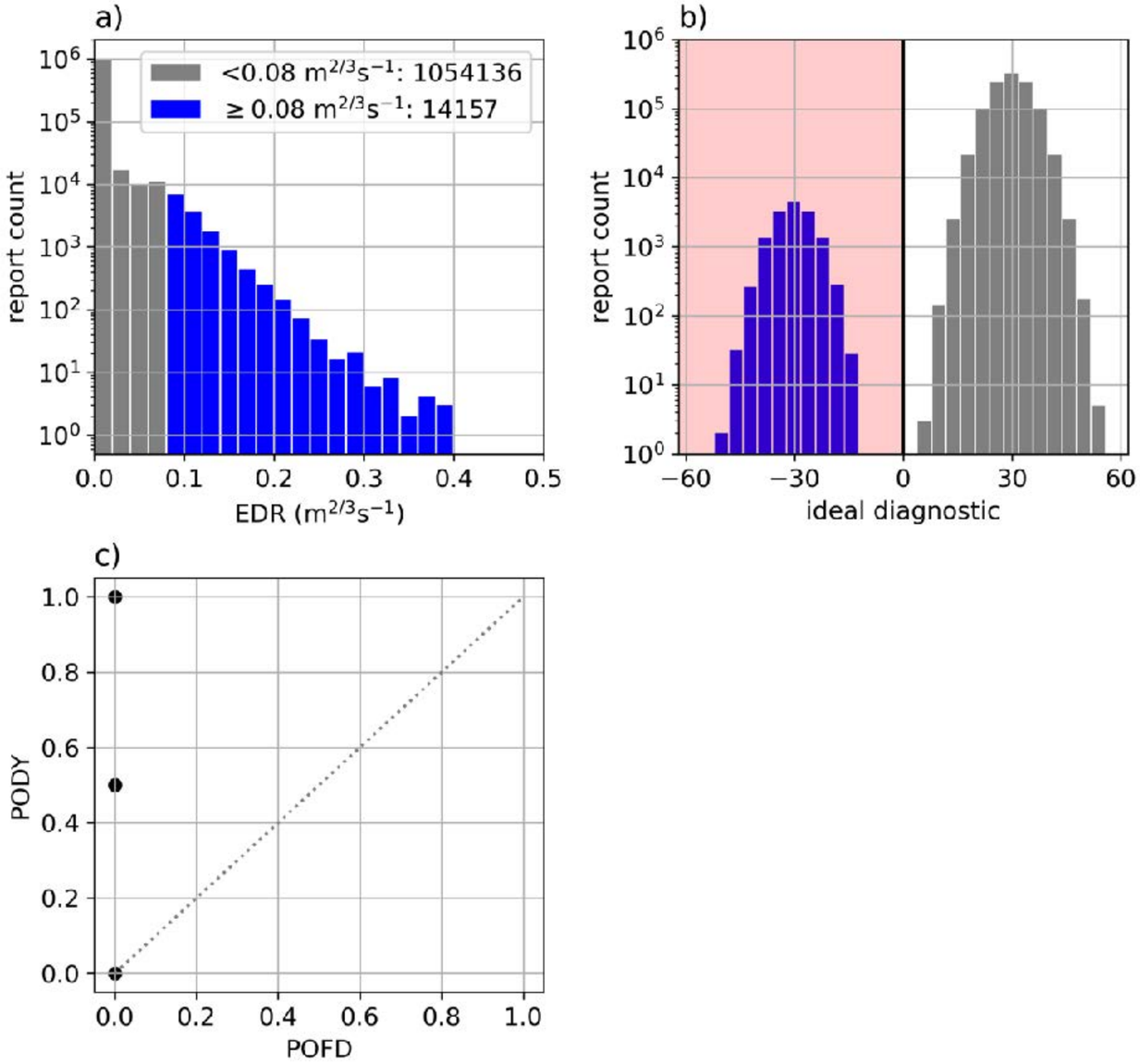
ERA5 diagnostics and ROC curves



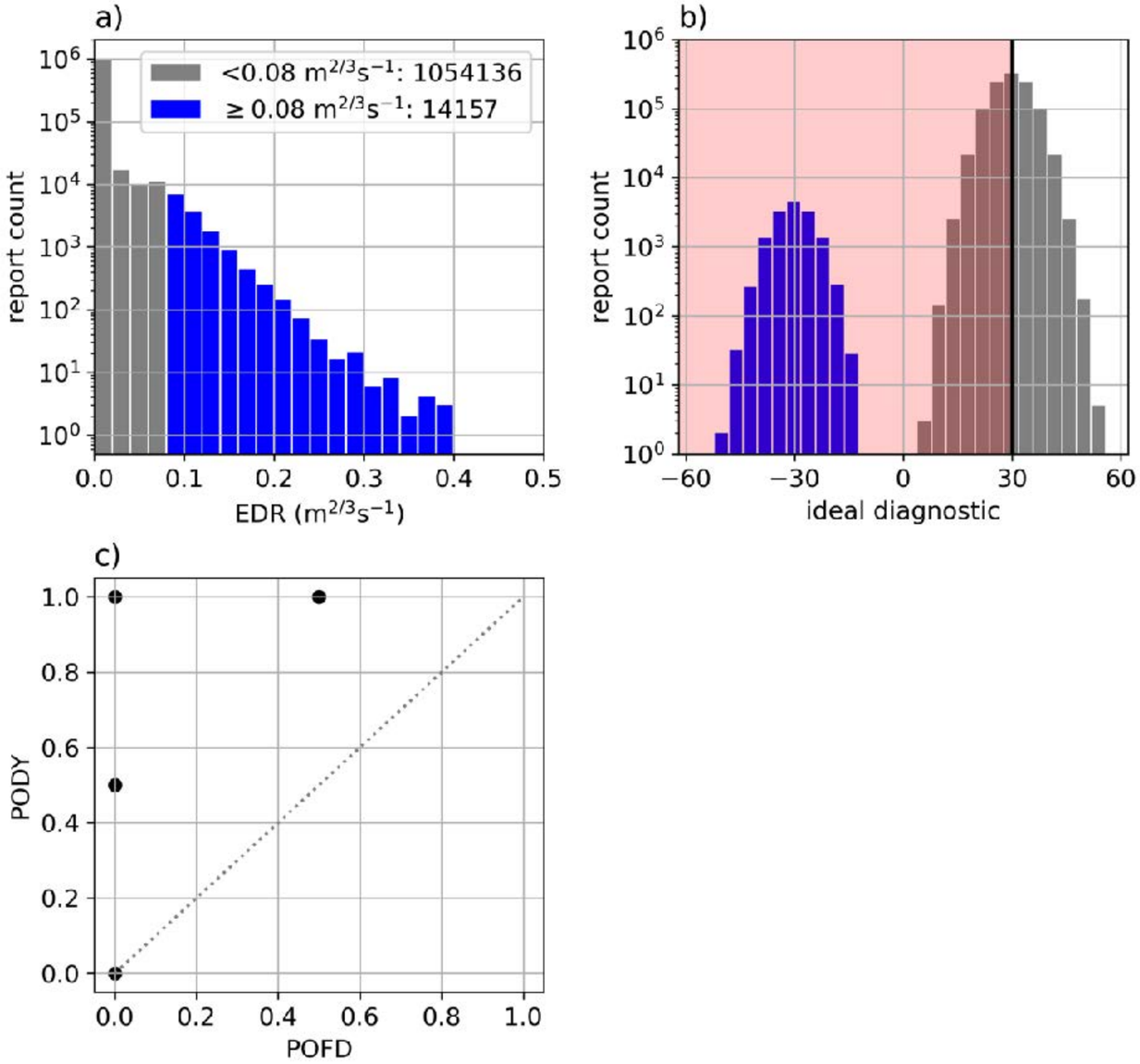
ERA5 diagnostics and ROC curves



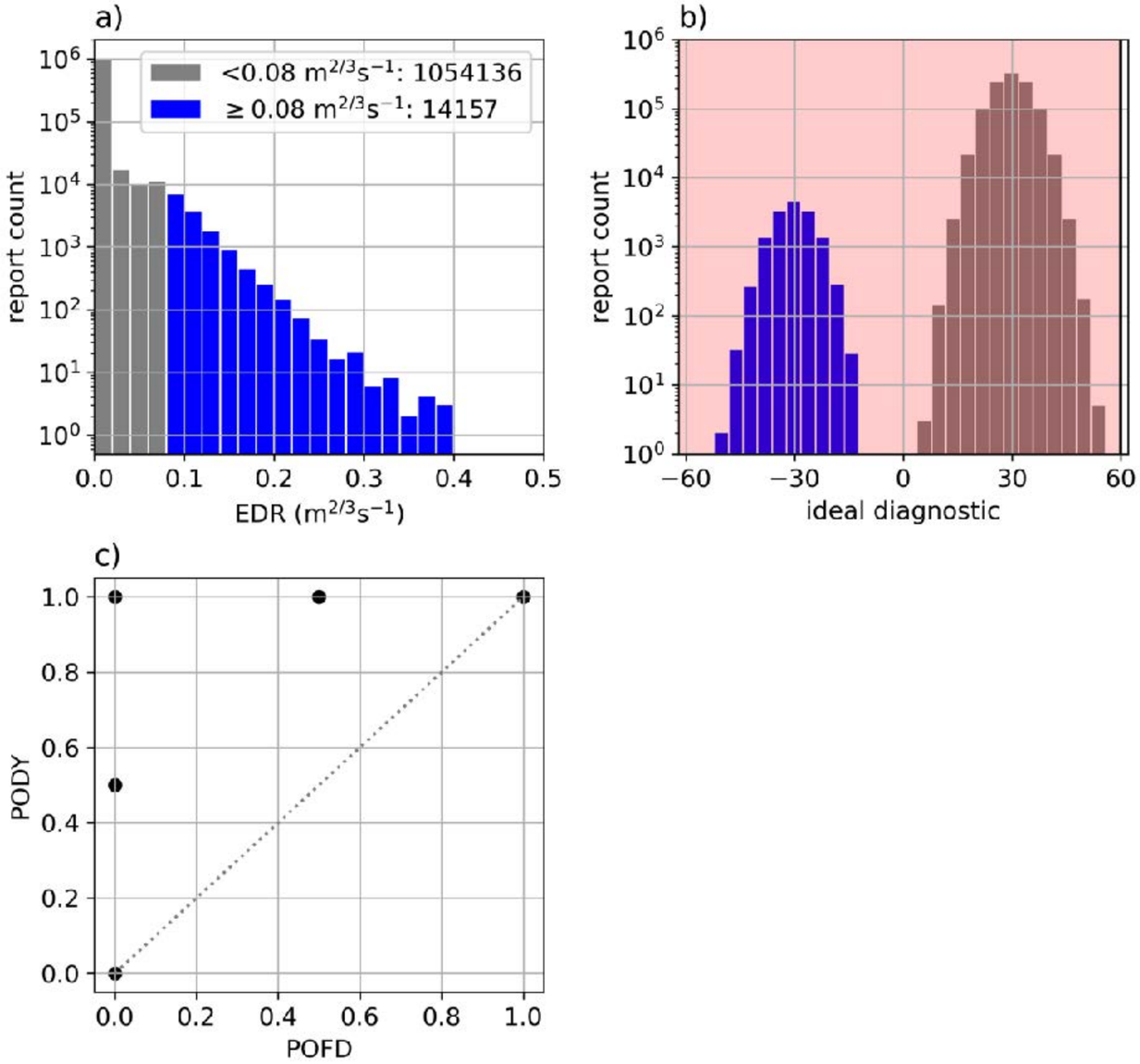
ERA5 diagnostics and ROC curves



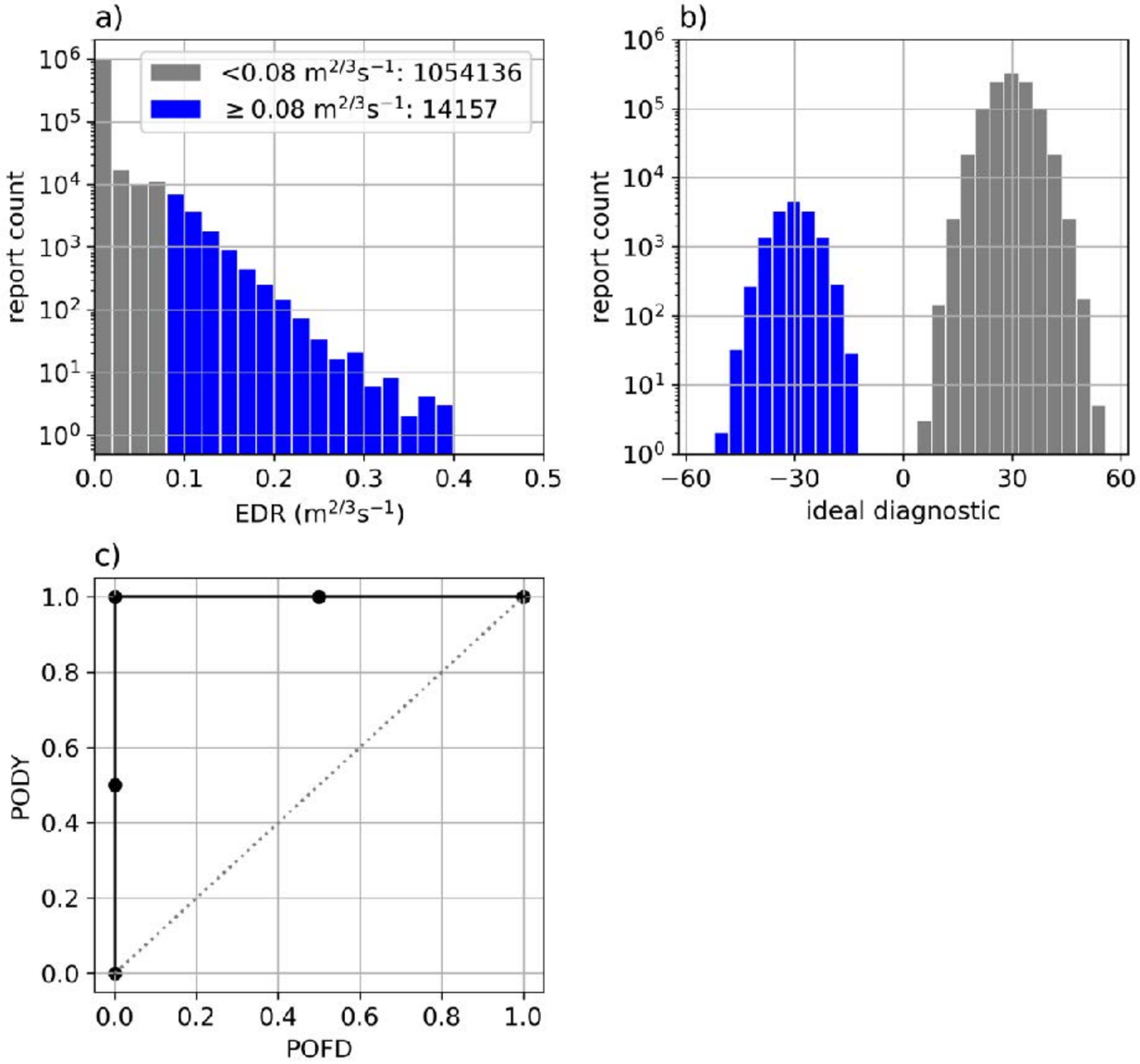
ERA5 diagnostics and ROC curves



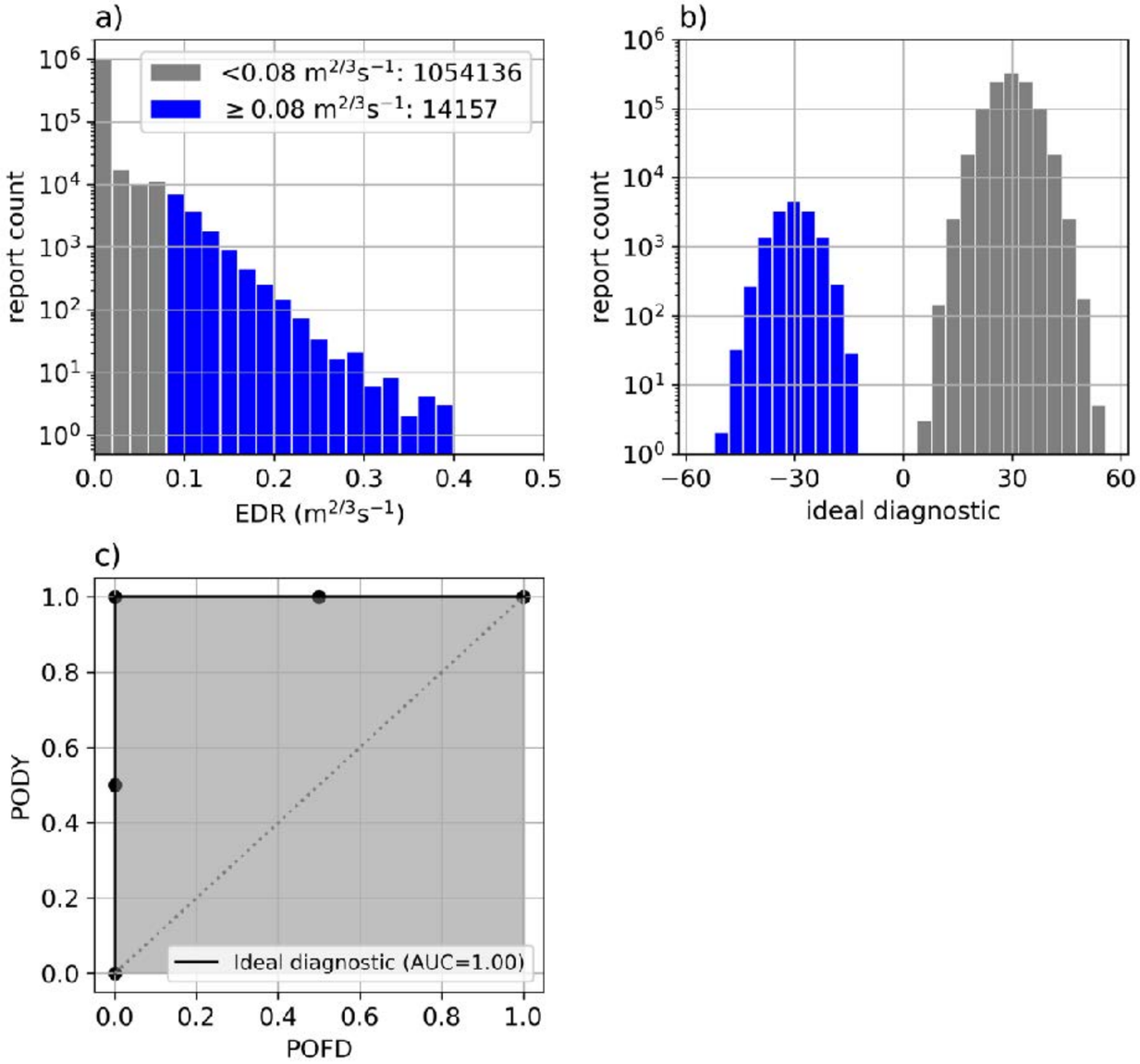
ERA5 diagnostics and ROC curves



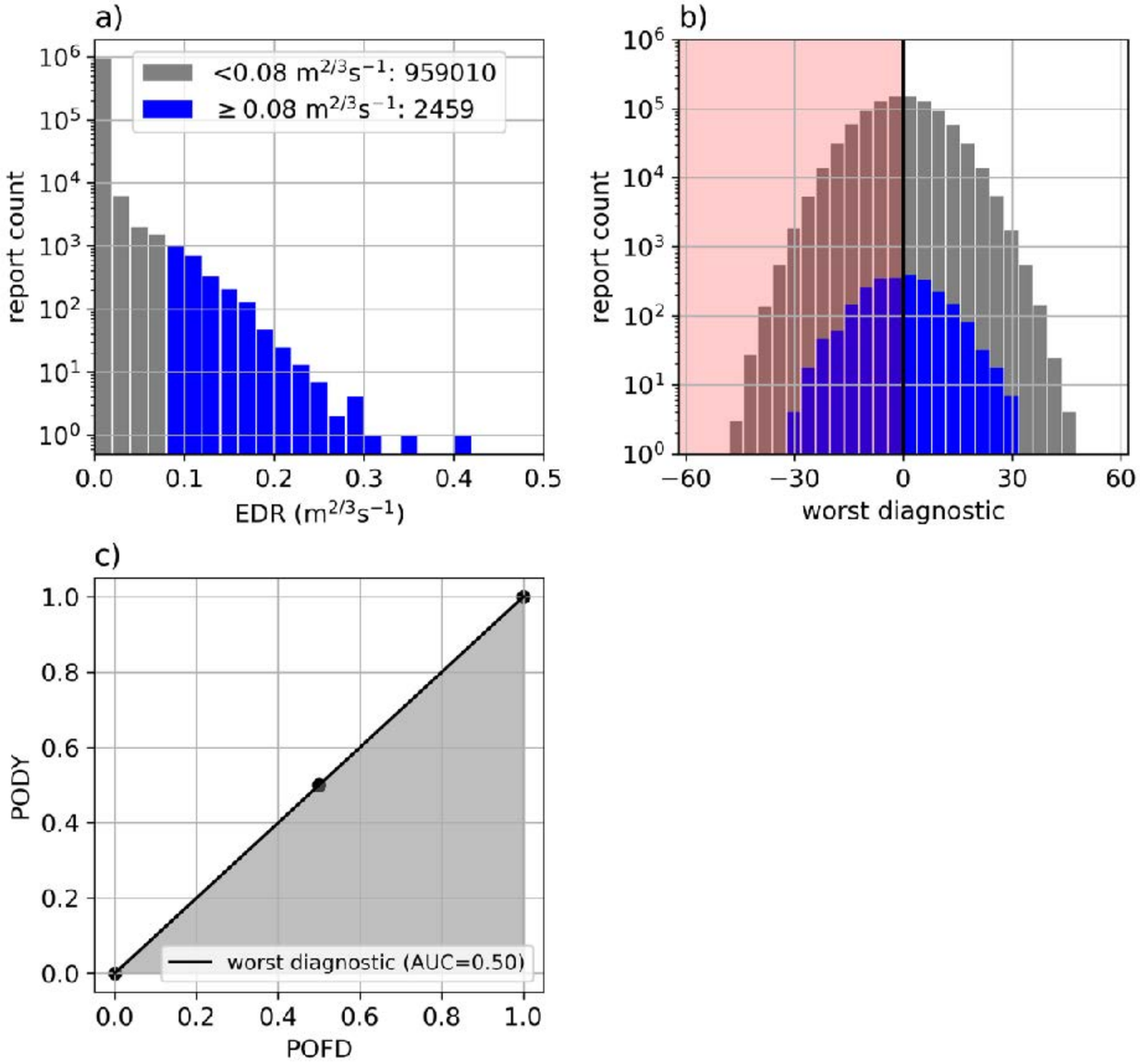
ERA5 diagnostics and ROC curves



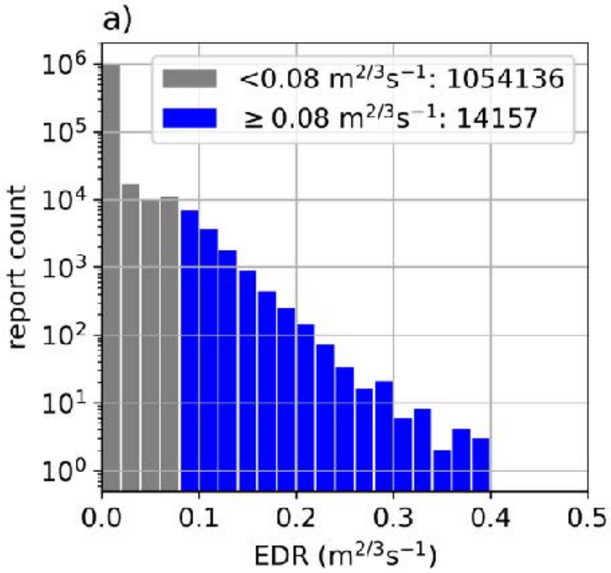
ERA5 diagnostics and ROC curves

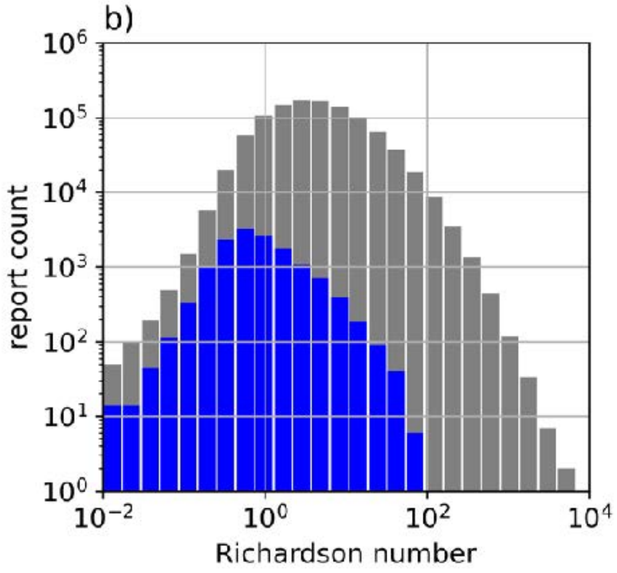


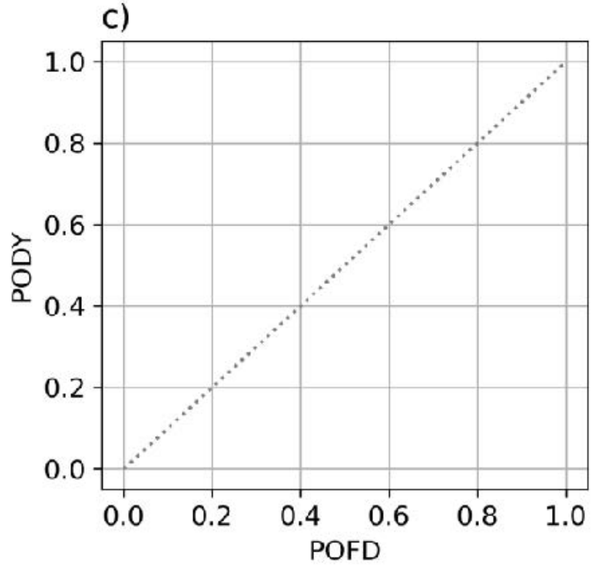
ERA5 diagnostics and ROC curves



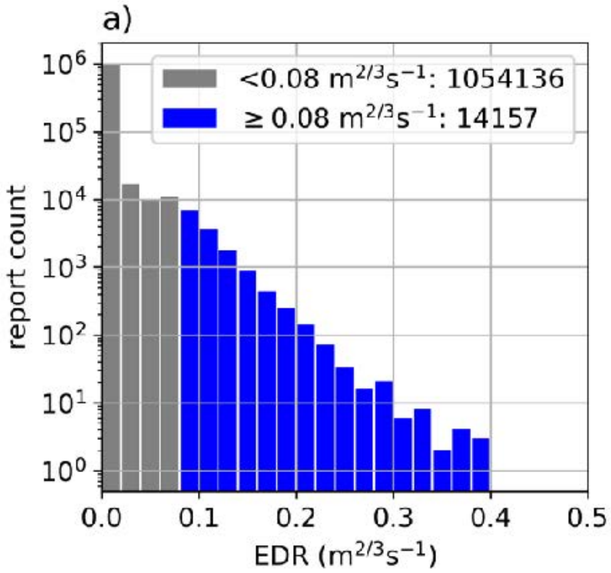
ERA5 diagnostics and ROC curves

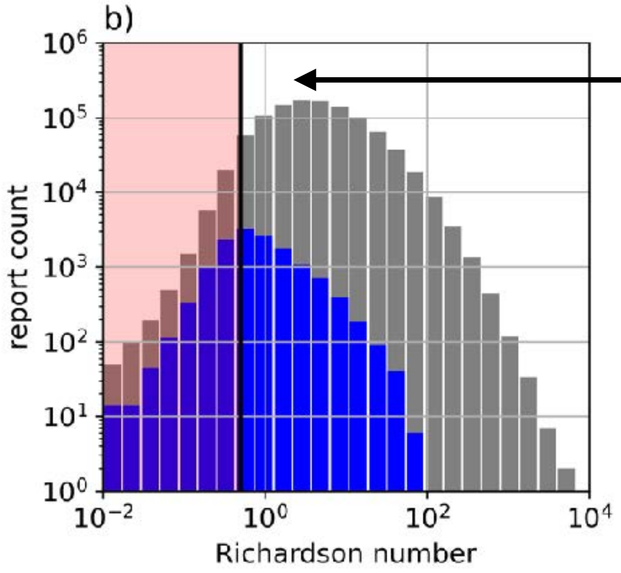




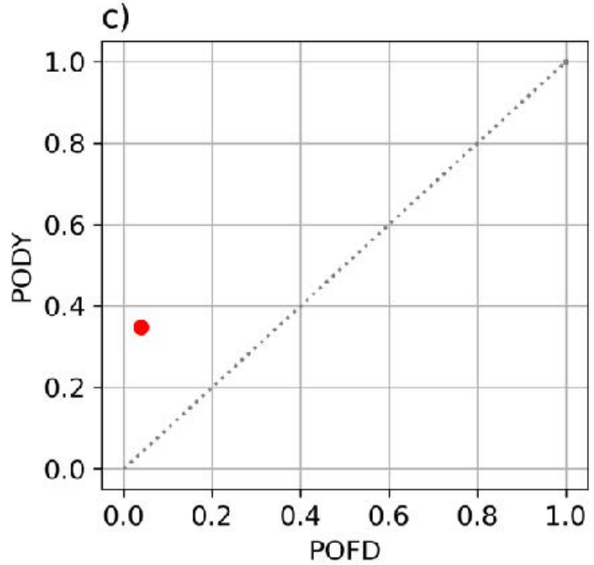


ERA5 diagnostics and ROC curves

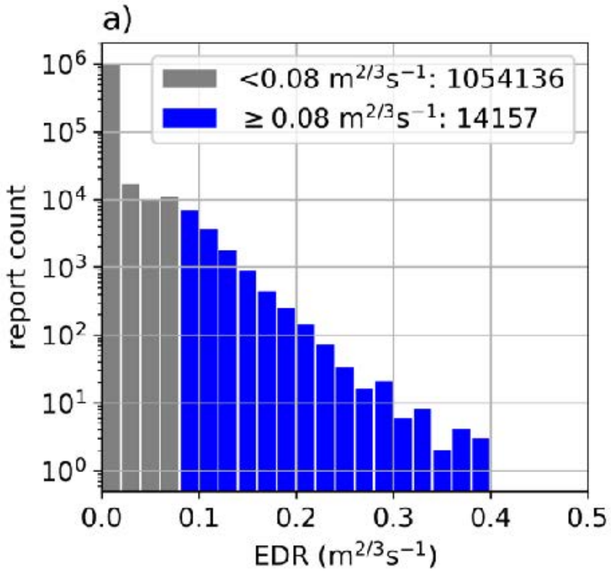


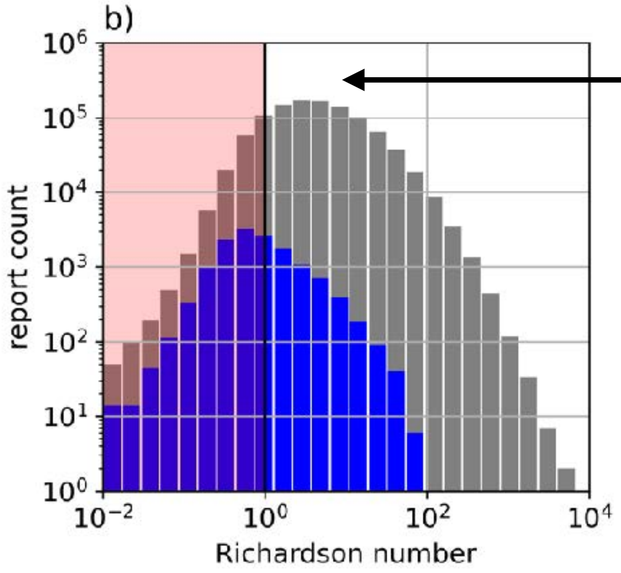


Richardson number threshold = 0.5

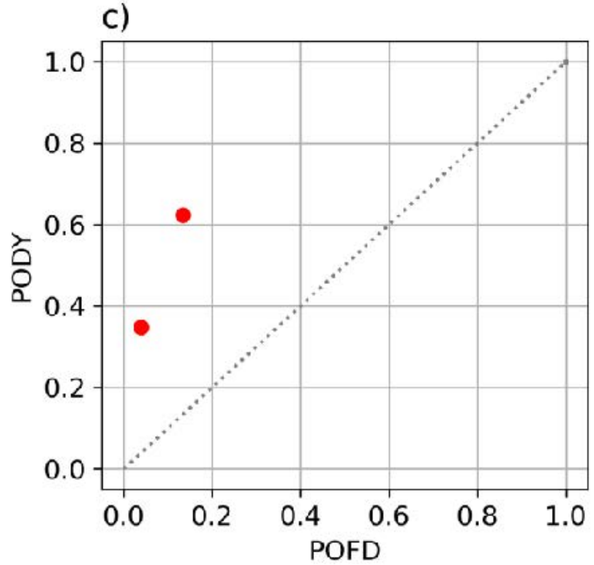


ERA5 diagnostics and ROC curves

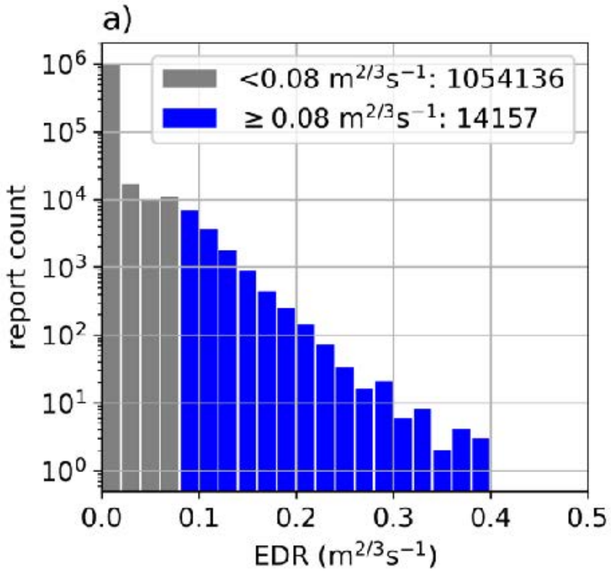


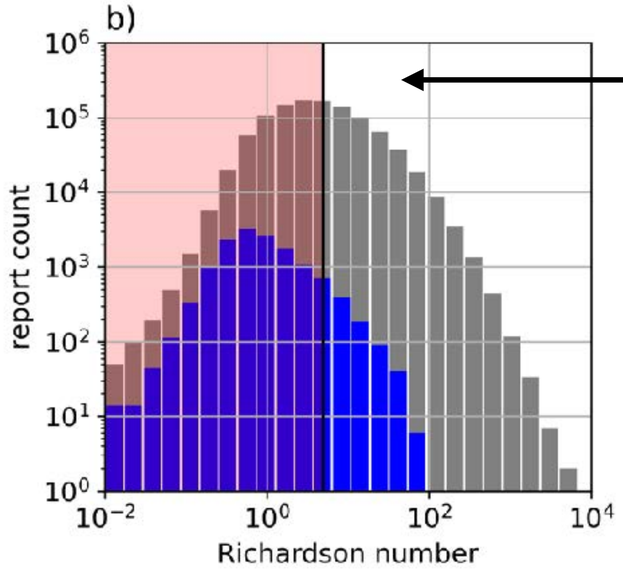


Richardson number threshold = 1

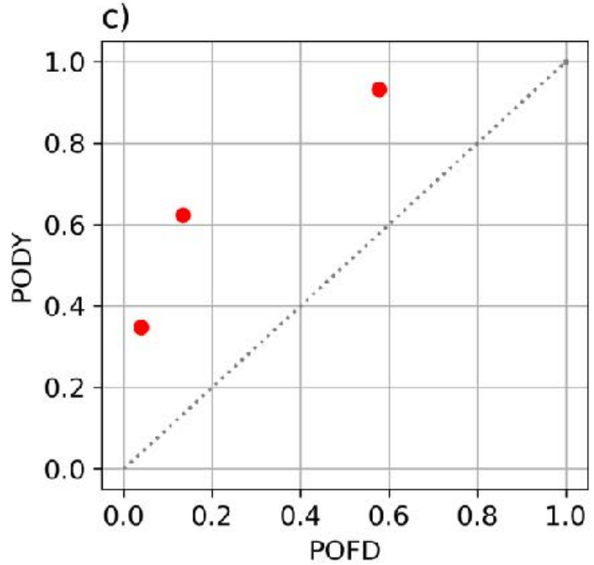


ERA5 diagnostics and ROC curves

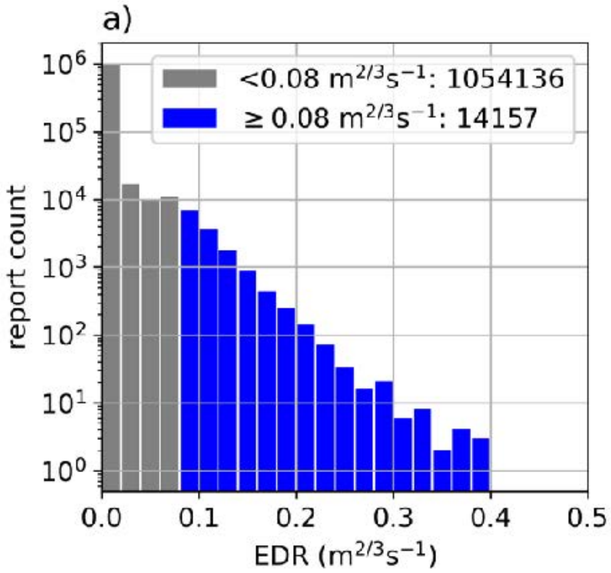


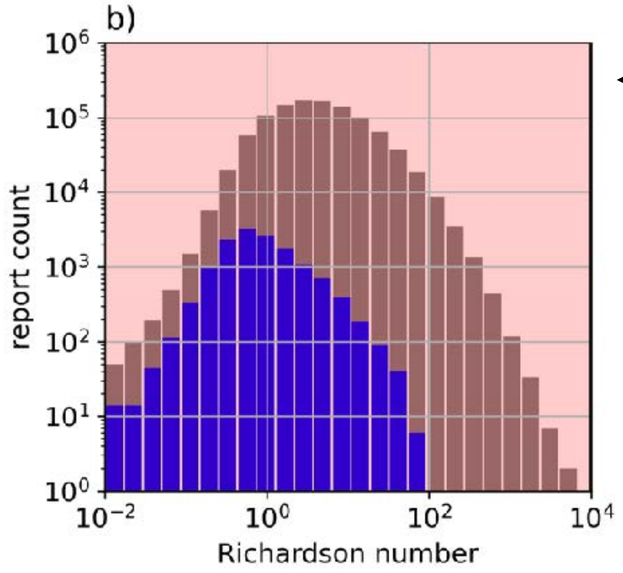


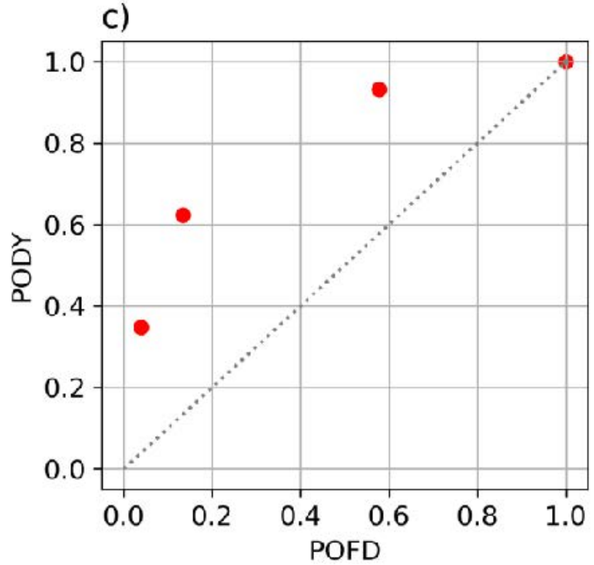
Richardson number threshold = 5



ERA5 diagnostics and ROC curves

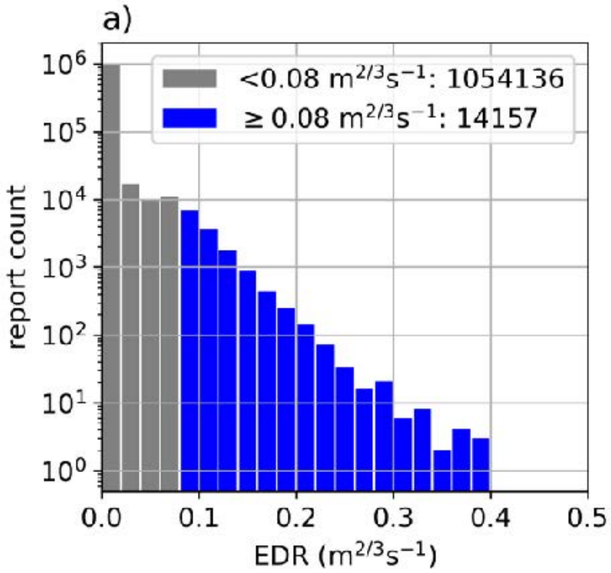


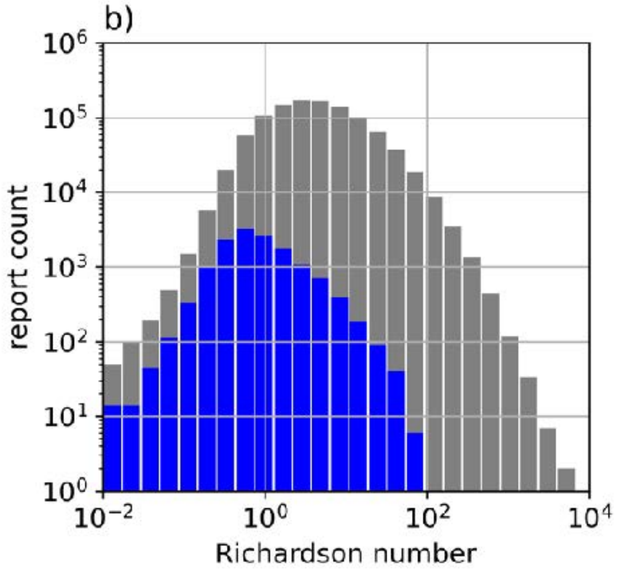


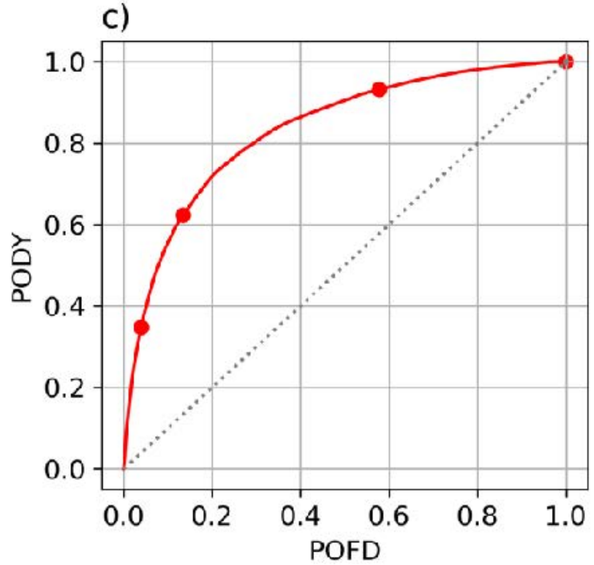


Richardson number threshold = 10000

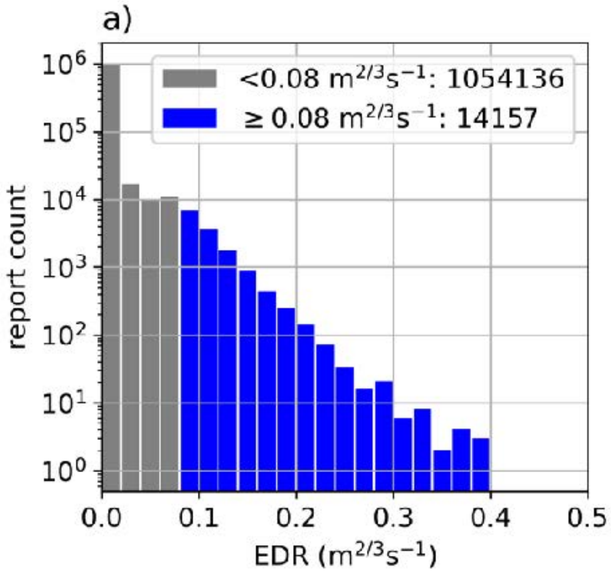
ERA5 diagnostics and ROC curves

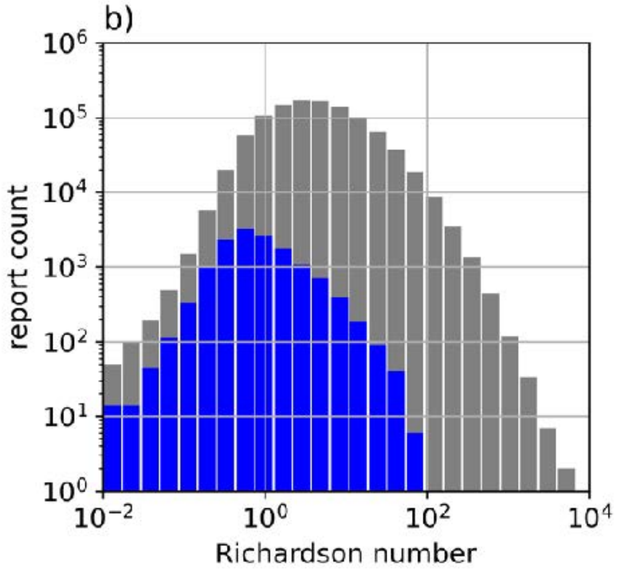


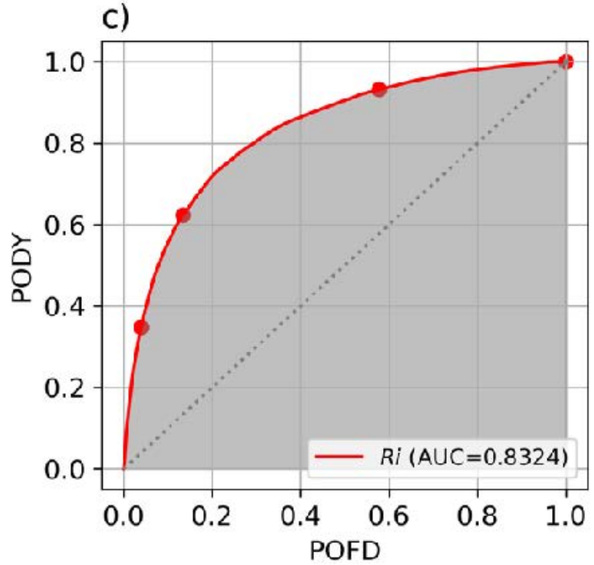




ERA5 diagnostics and ROC curves

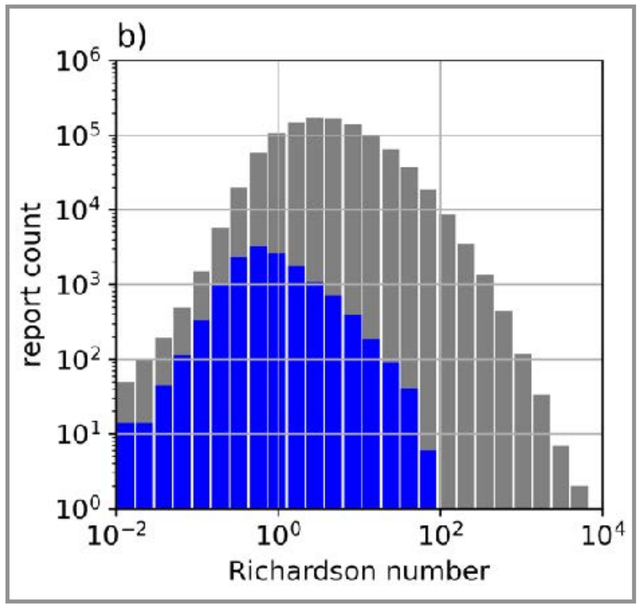


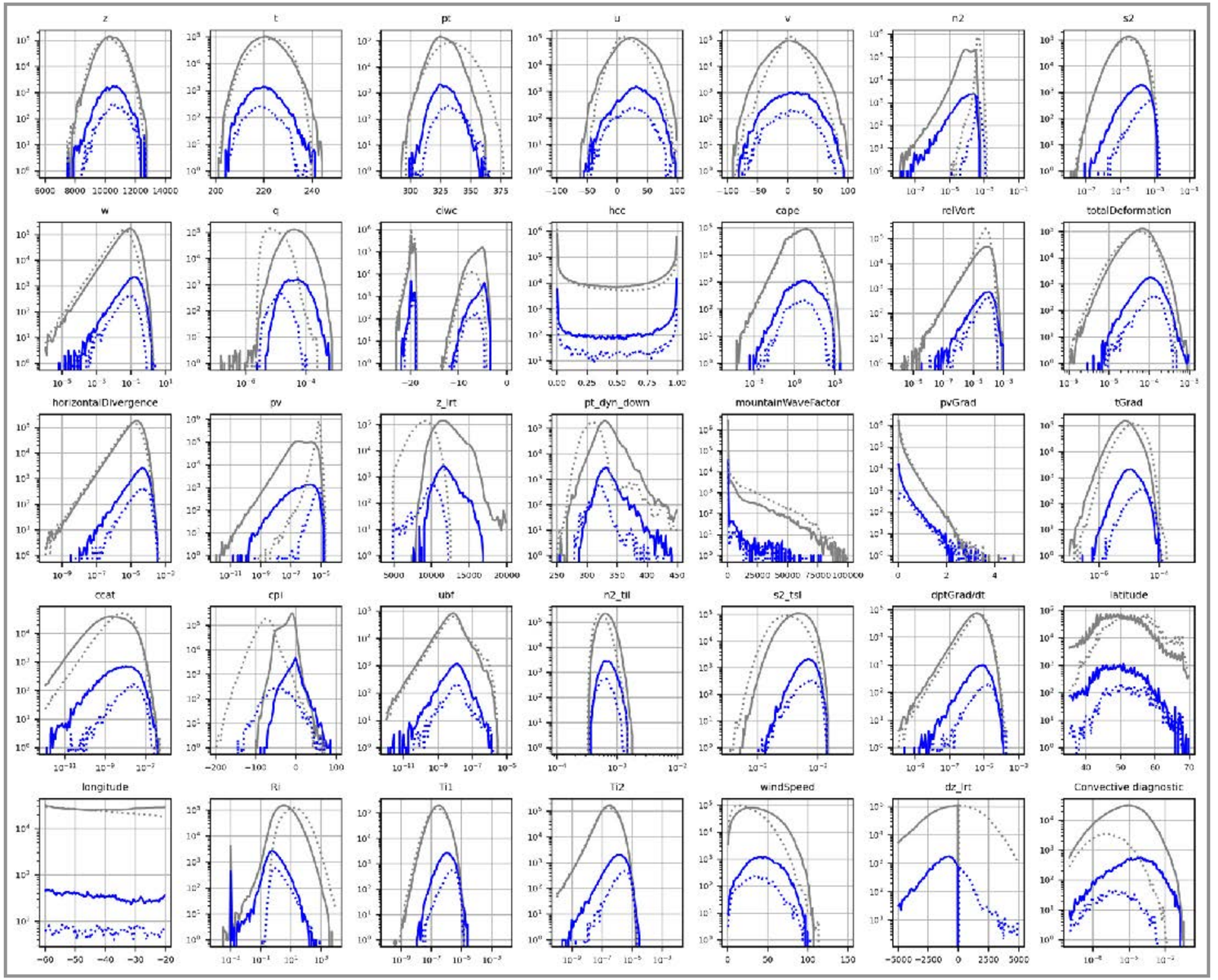




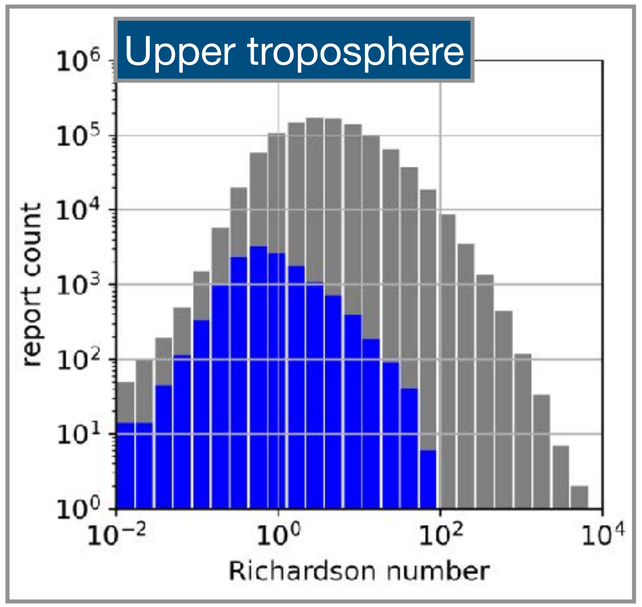
Richardson number as benchmark diagnostic:
AUC as metric: 0.83

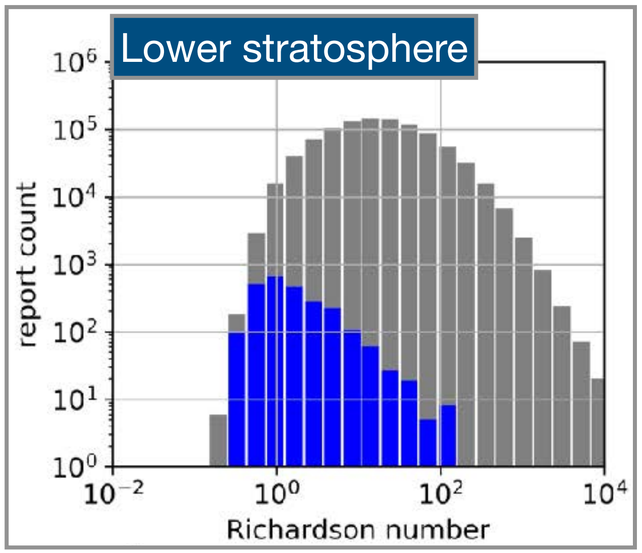
Random Forest Classifier

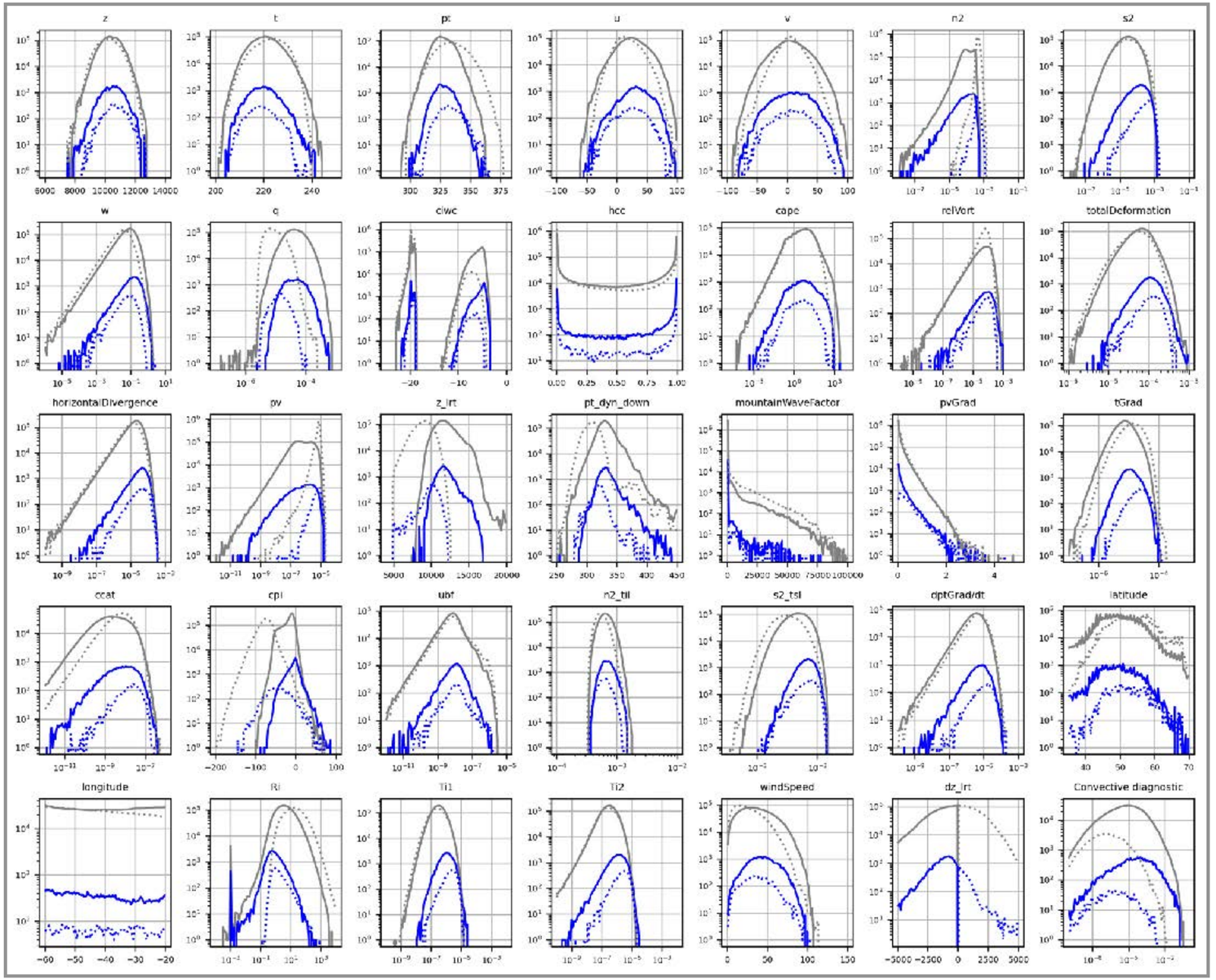




Random Forest Classifier







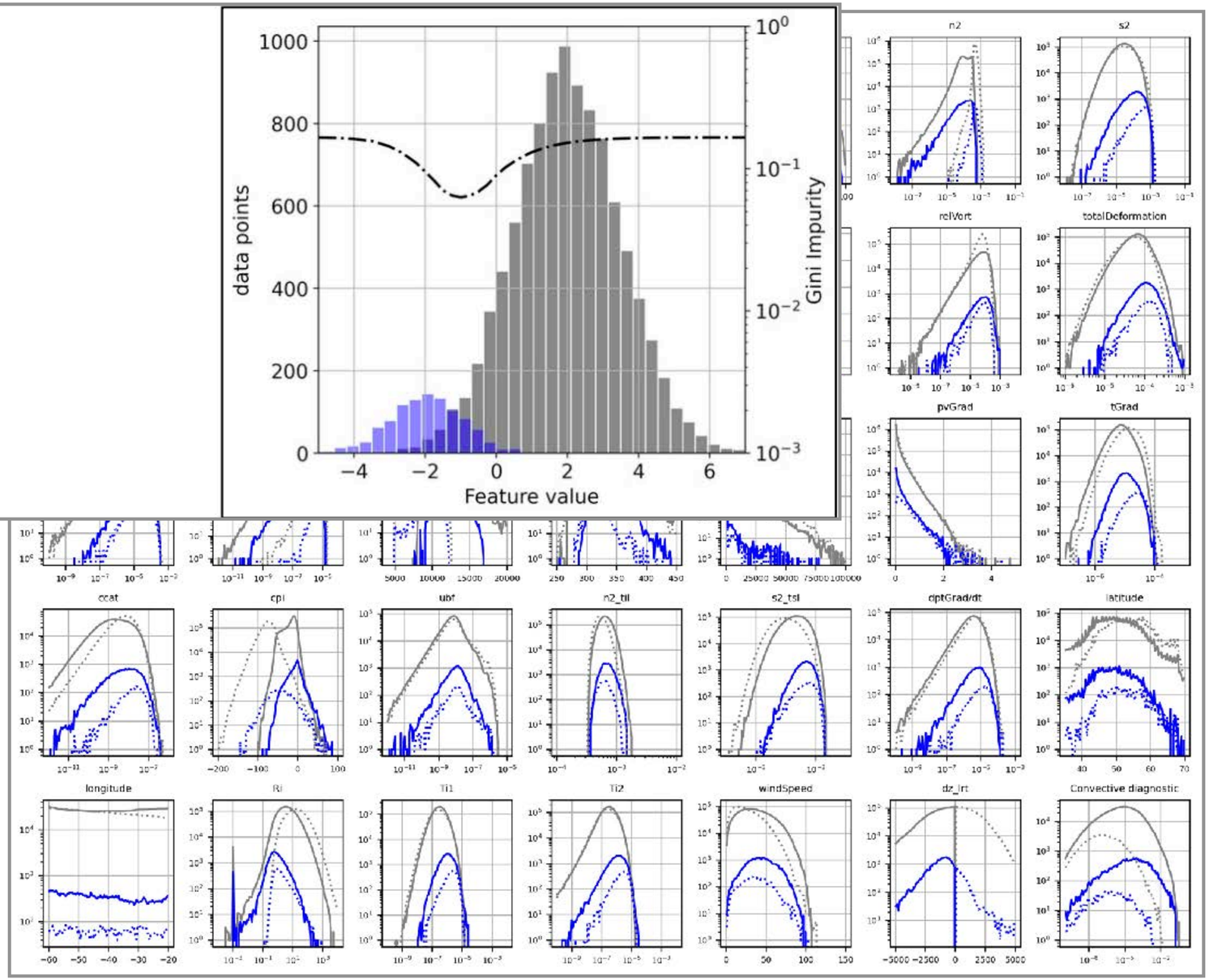
https://doi.org/10.1029/2022GL100036

Random Forest Classifier

Minimize loss: $\Theta^* = \text{argmin}_\Theta H(\Theta)$

Define loss as weighted Gini impurity after split: $H(\Theta) = \sum_{b=1}^{2} \frac{n^b}{n} G^b(\Theta)$

Gini ipmurity: $G^b(\Theta) = \sum_k \frac{n_k^b}{n^b}(1 - \frac{n_k^b}{n^b})$

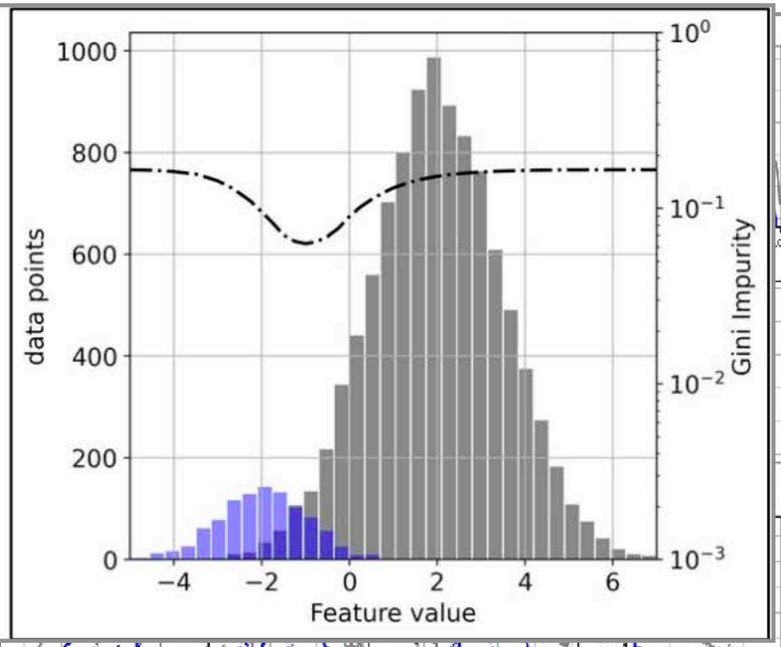


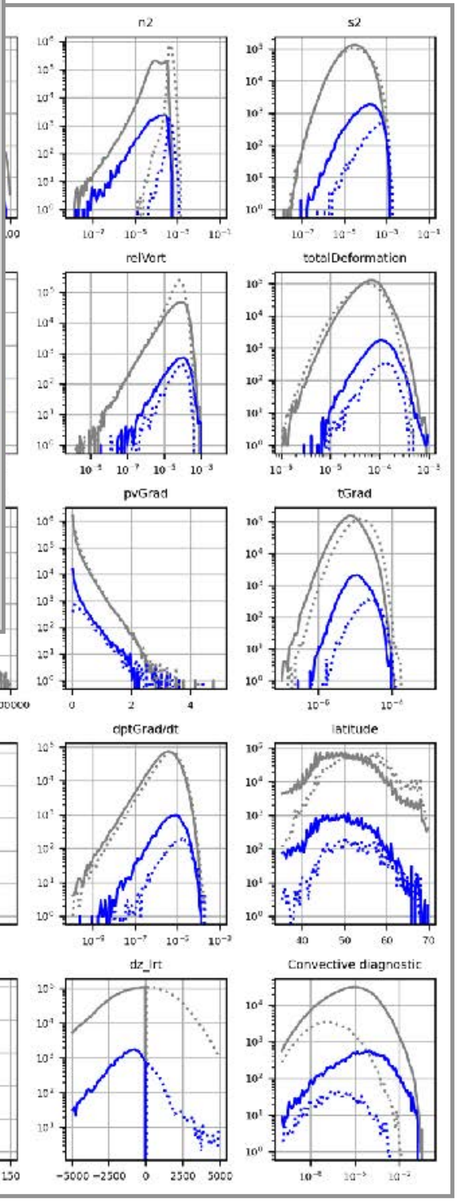
Random Forest Classifier

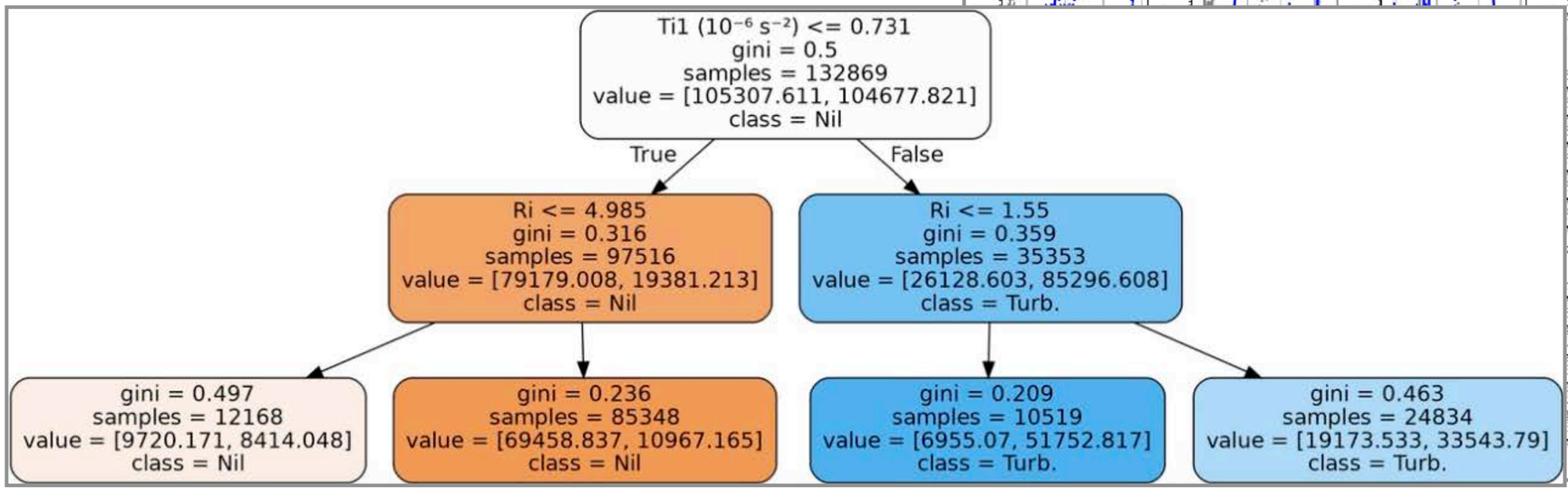
Minimize loss: $\Theta^* = \text{argmin}_\Theta H(\Theta)$

Define loss as weighted Gini impurity after split: $H(\Theta) = \sum_{b=1}^{2} \frac{n^b}{n} G^b(\Theta)$

Gini ipmurity: $G^b(\Theta) = \sum_k \frac{n_k^b}{n^b}(1 - \frac{n_k^b}{n^b})$





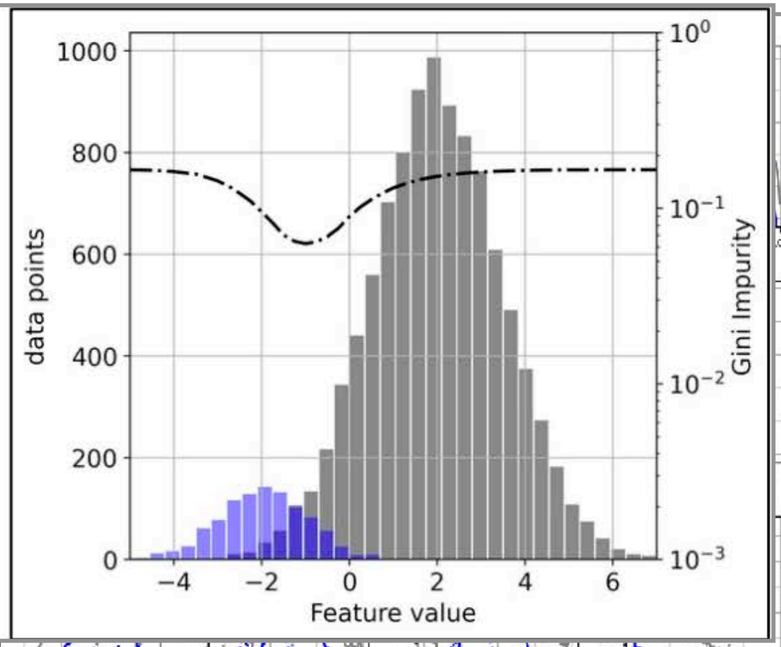


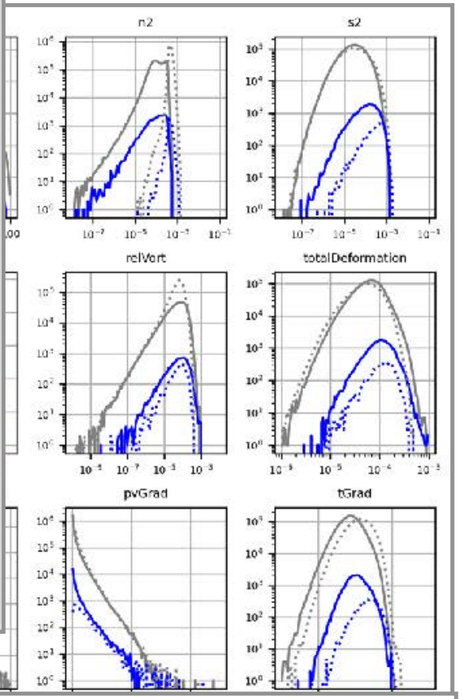
Random Forest Classifier

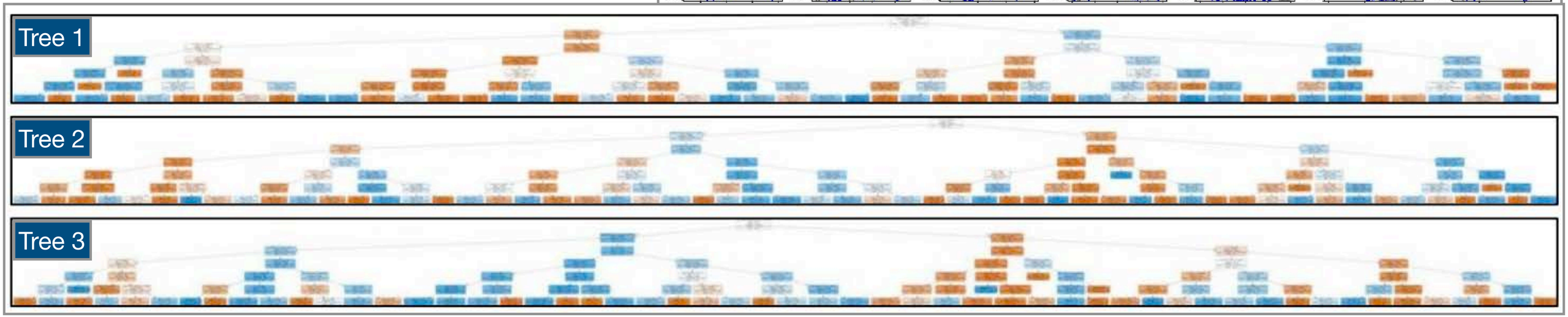
Minimize loss: $\Theta^* = \text{argmin}_\Theta H(\Theta)$

Define loss as weighted Gini impurity after split: $H(\Theta) = \sum_{b=1}^{2} \frac{n^b}{n} G^b(\Theta)$

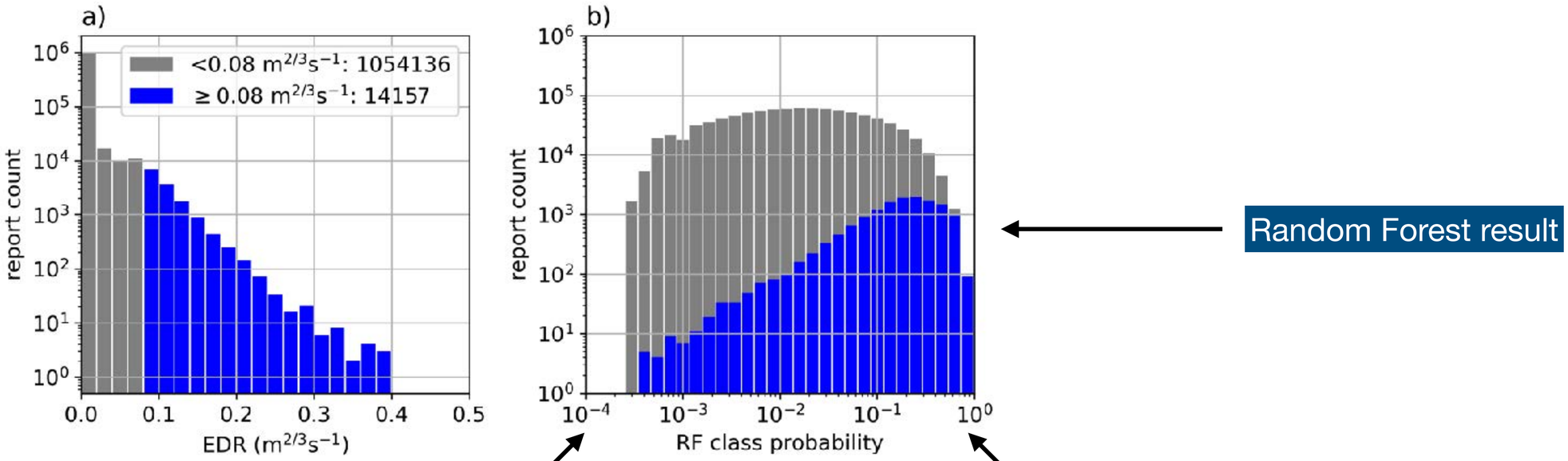
Gini ipmurity: $G^b(\Theta) = \sum_k \frac{n_k^b}{n^b}(1 - \frac{n_k^b}{n^b})$







Random Forest Classifier

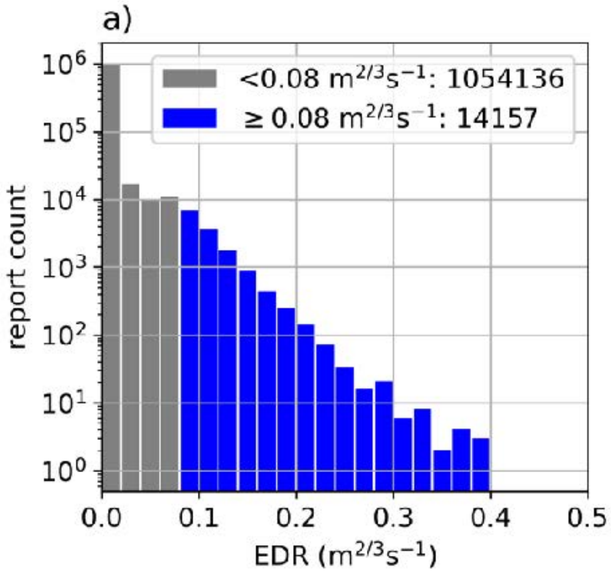


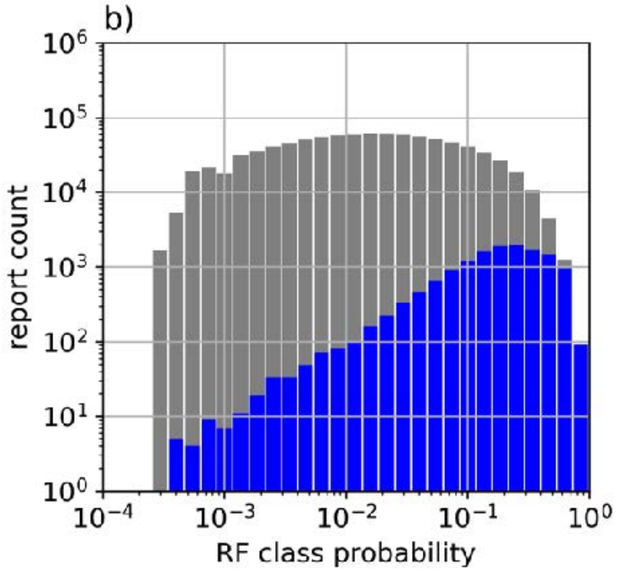
Random Forest result

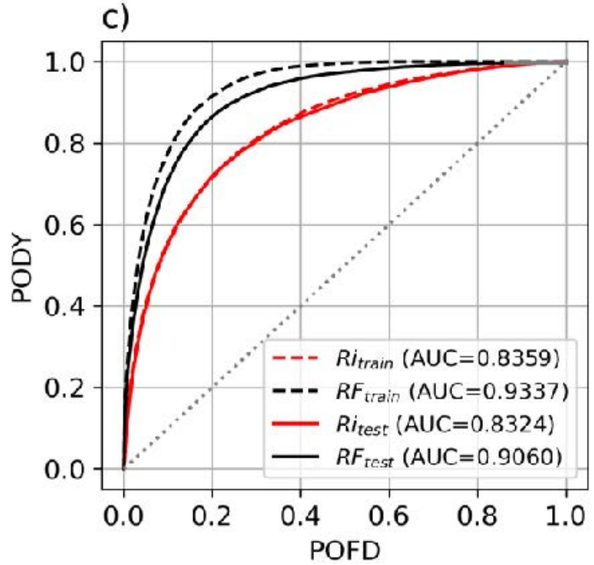
non-turbulent

turbulent

Random Forest Classifier

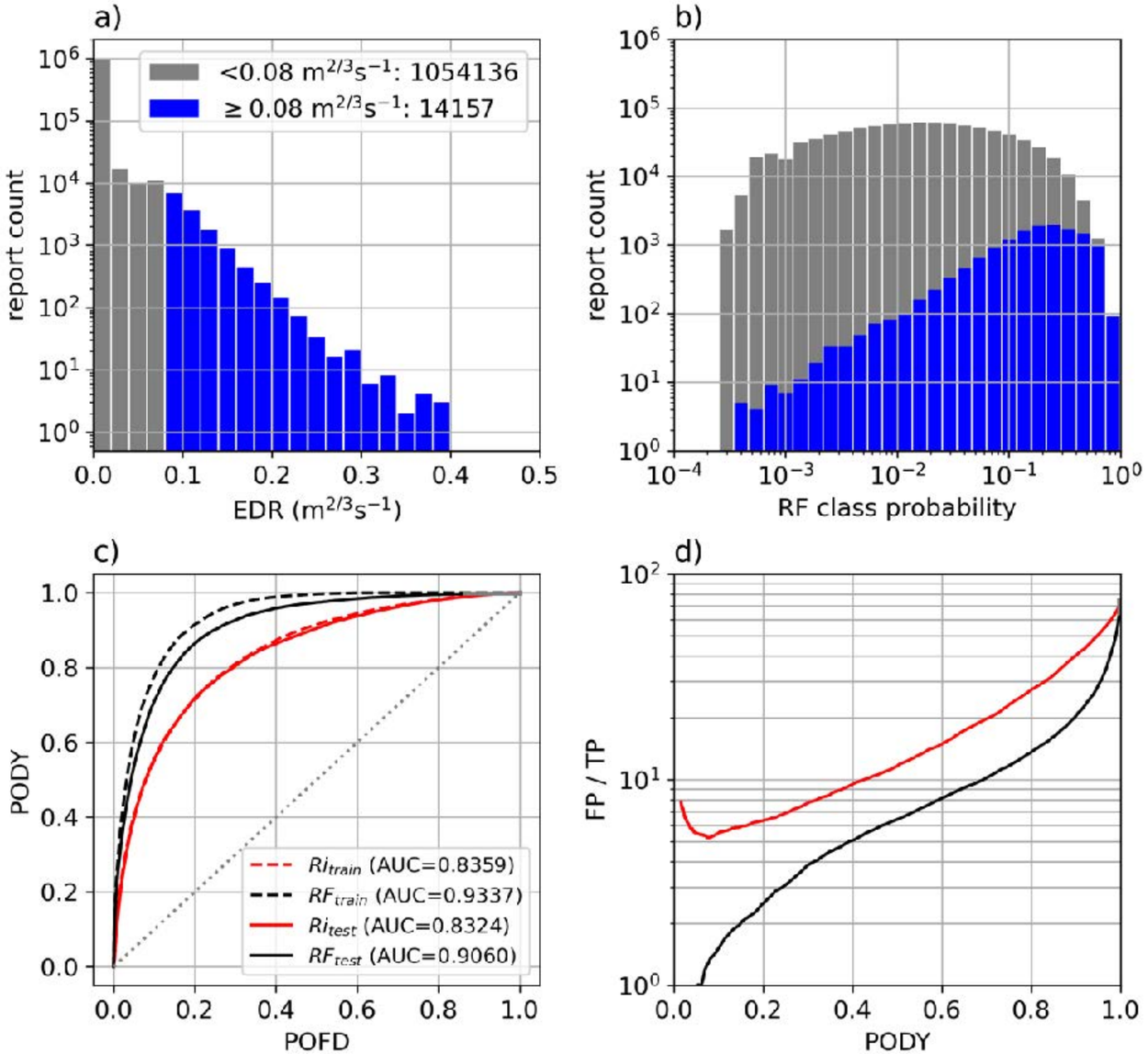






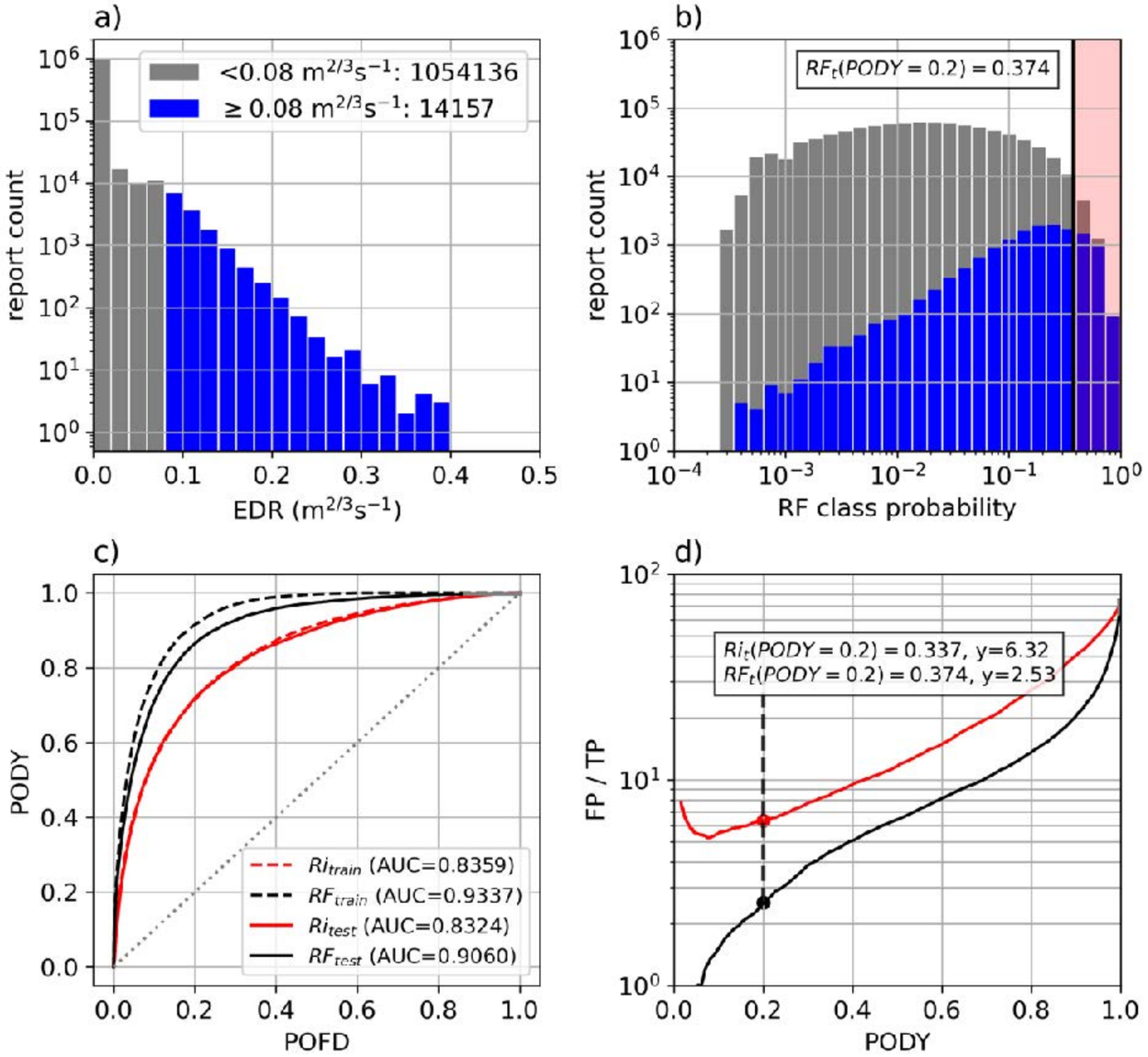
Random Forest result

Random Forest Classifier



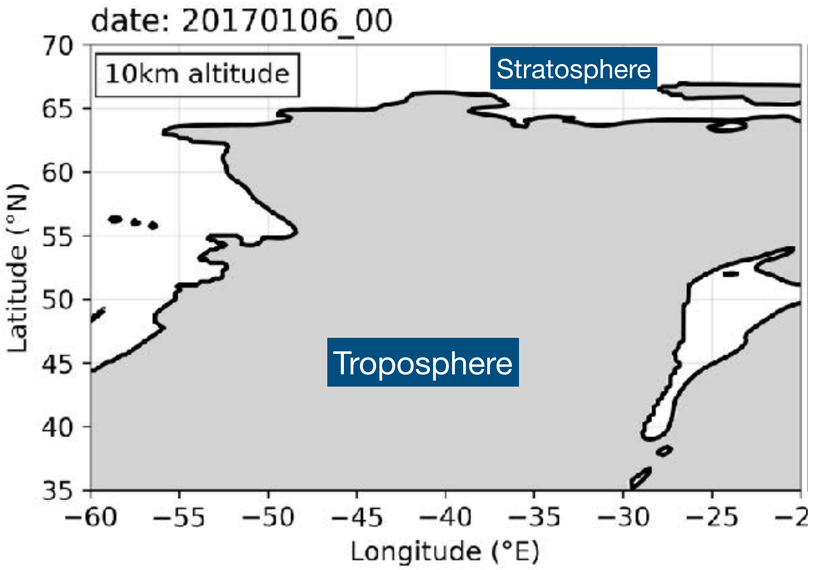
Random Forest result

Random Forest Classifier

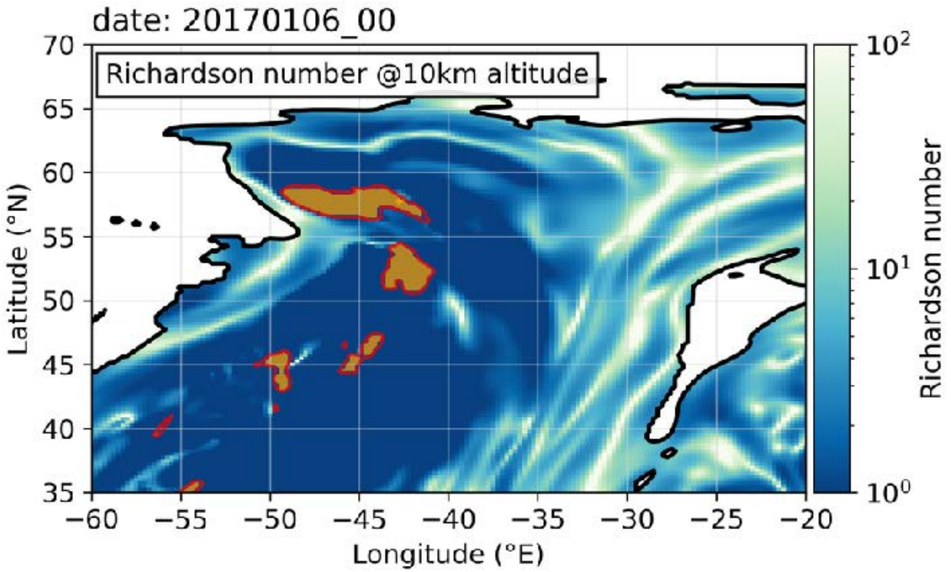


Random Forest result

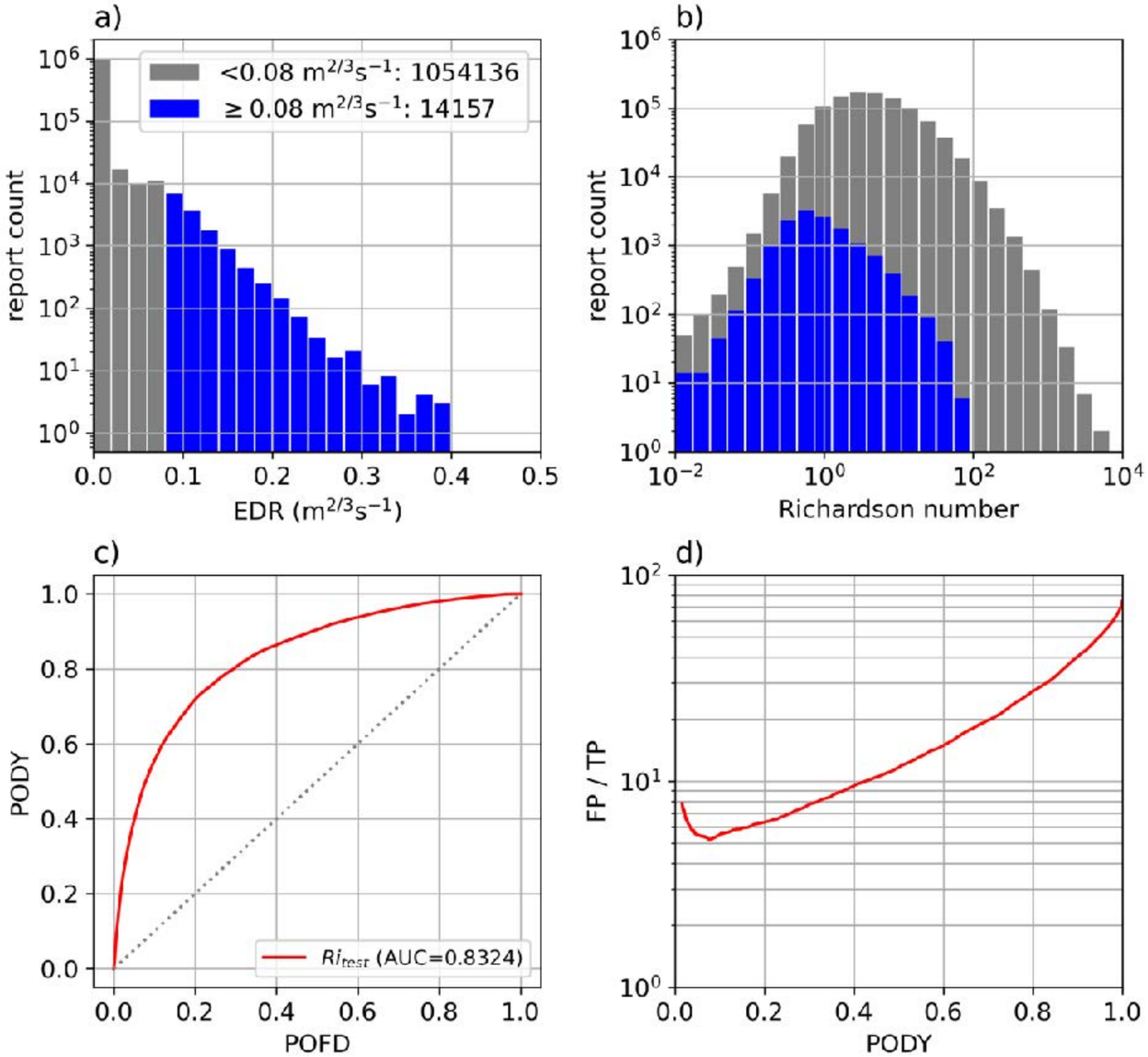
Inference / application

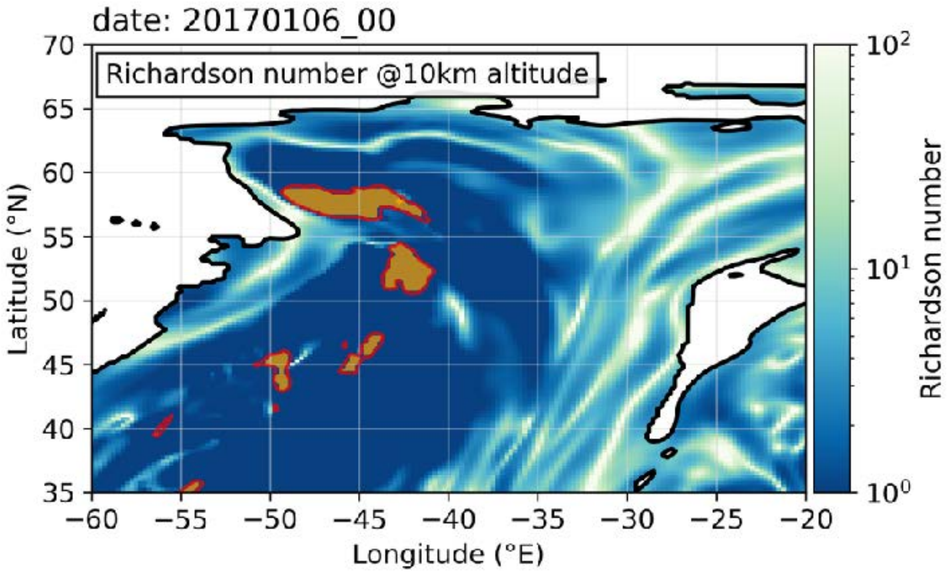


Inference / application

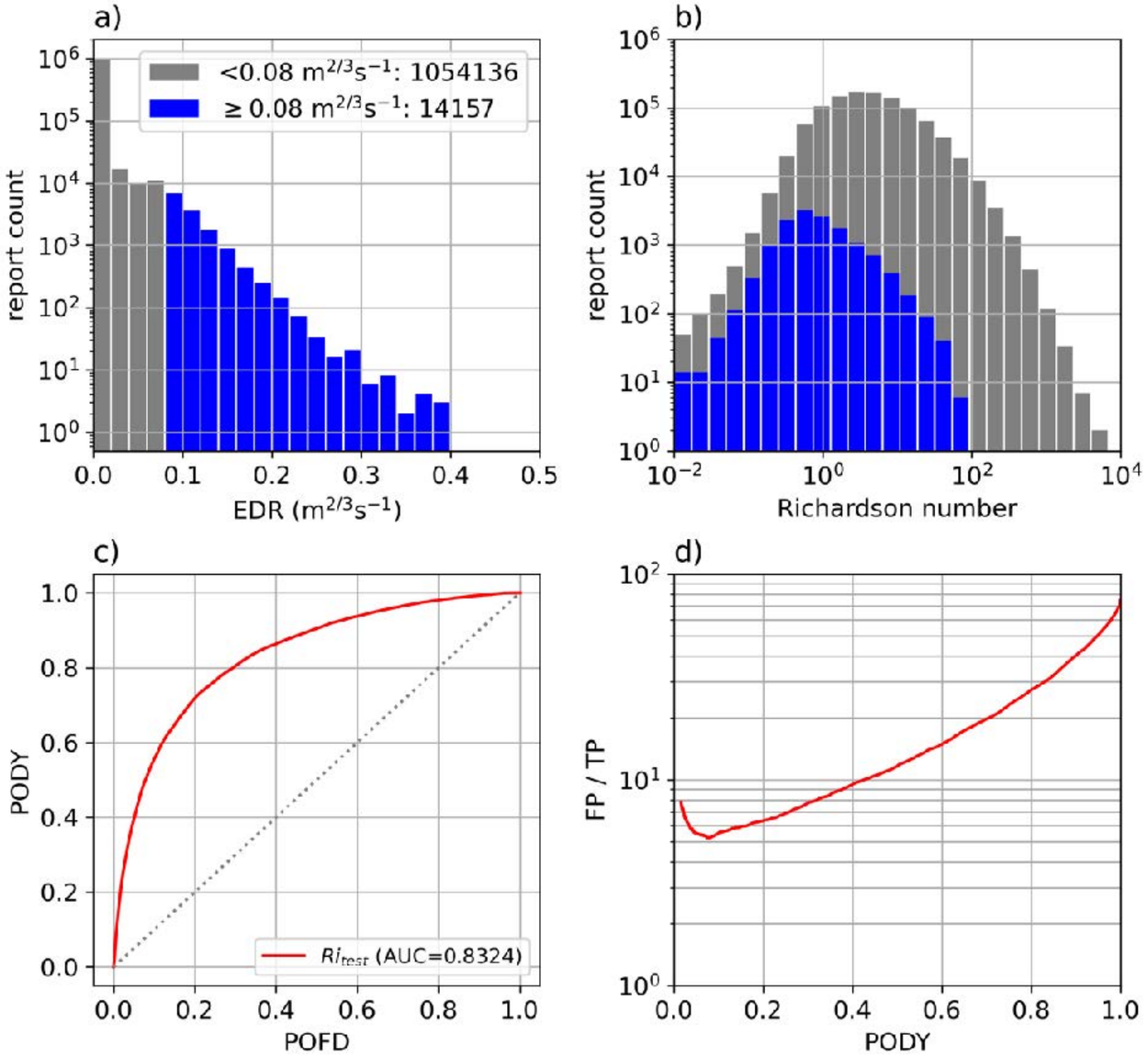


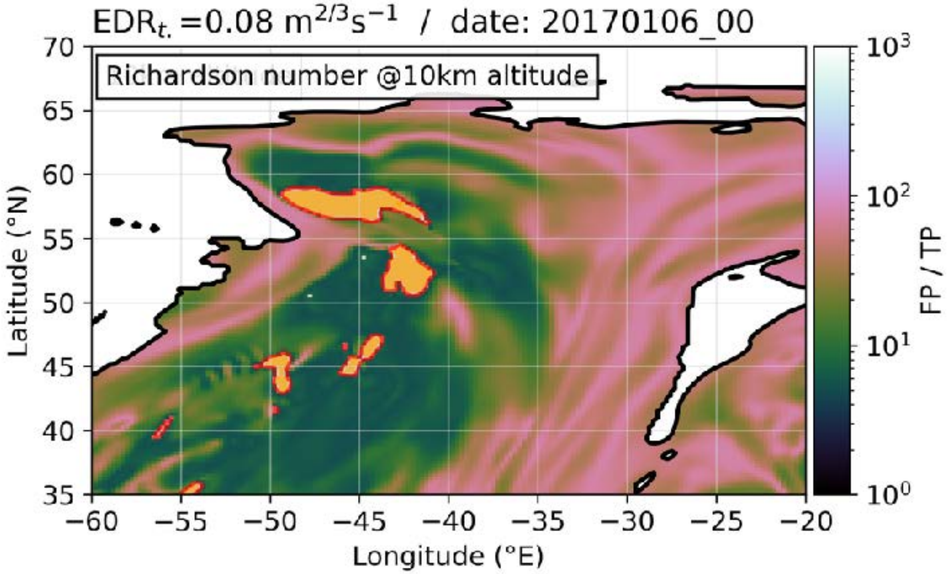
Inference / application



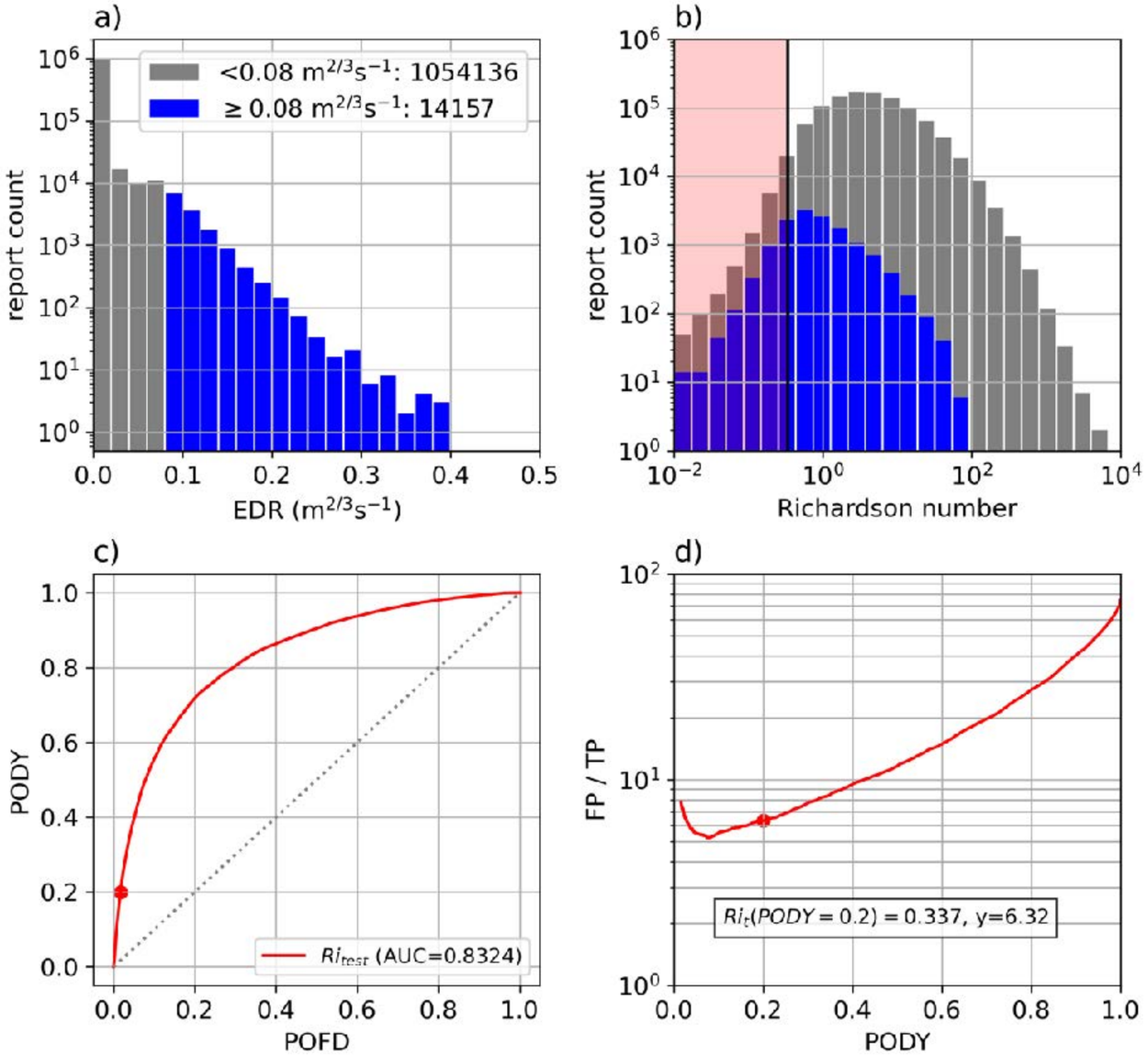


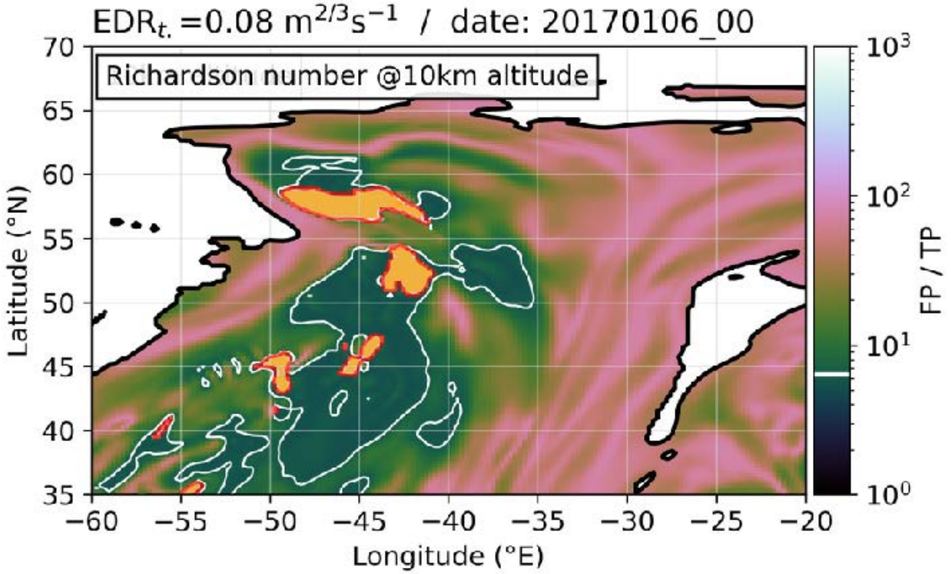
Inference / application



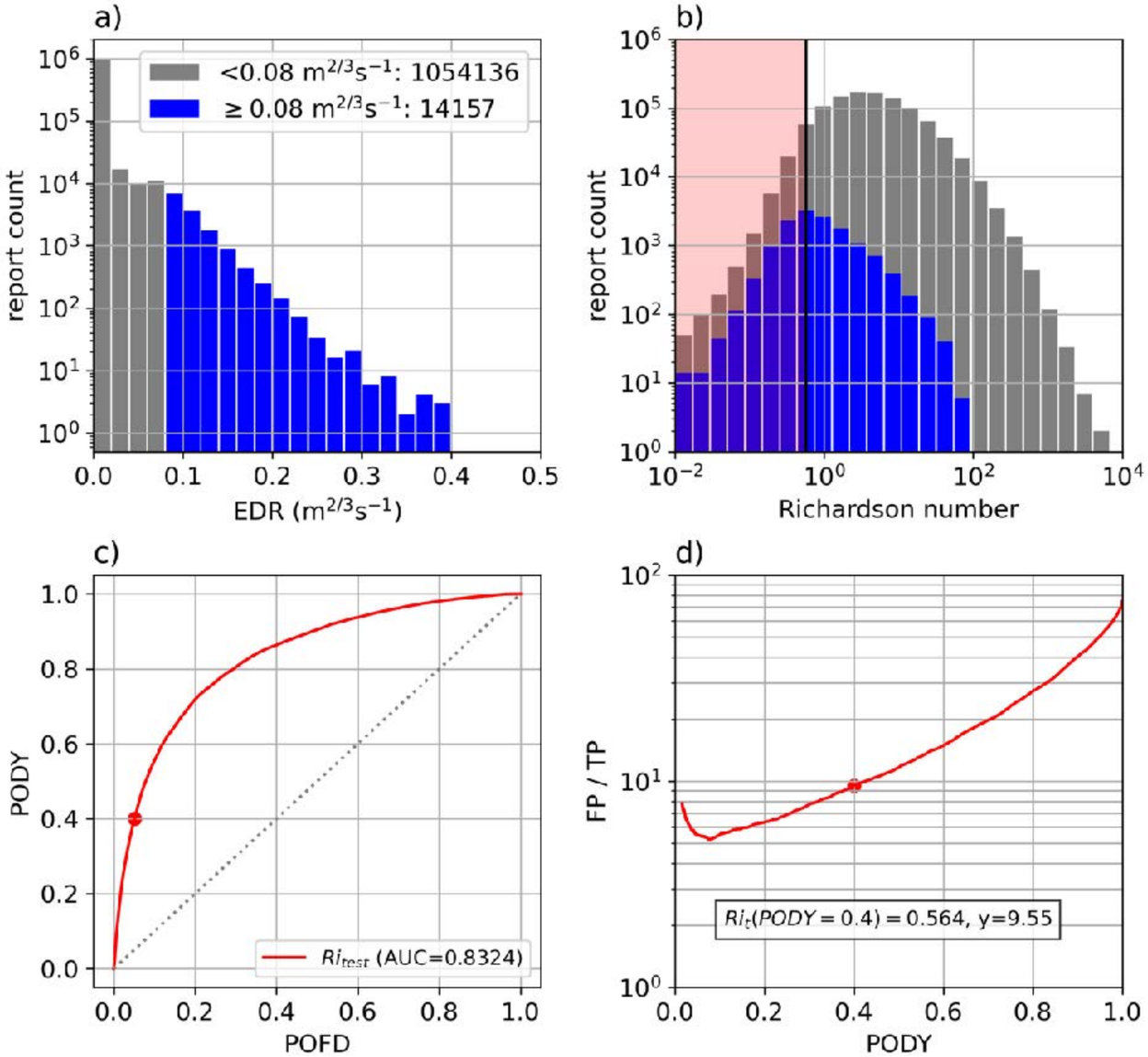


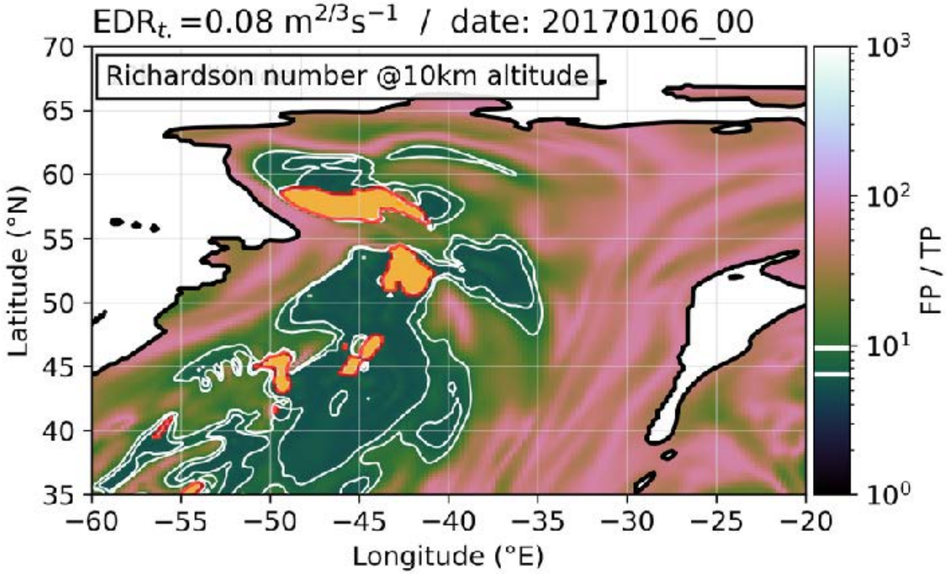
Inference / application



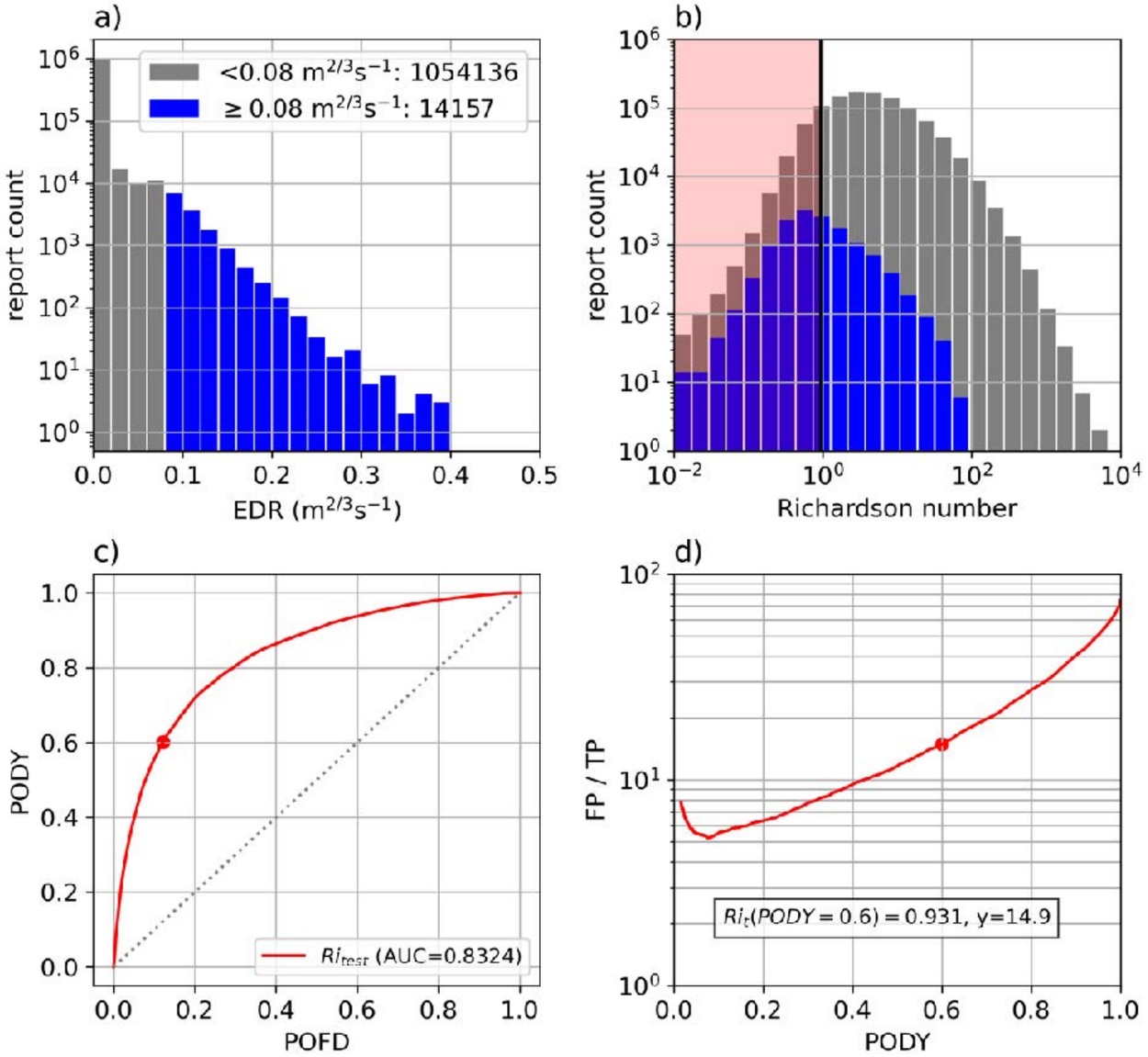


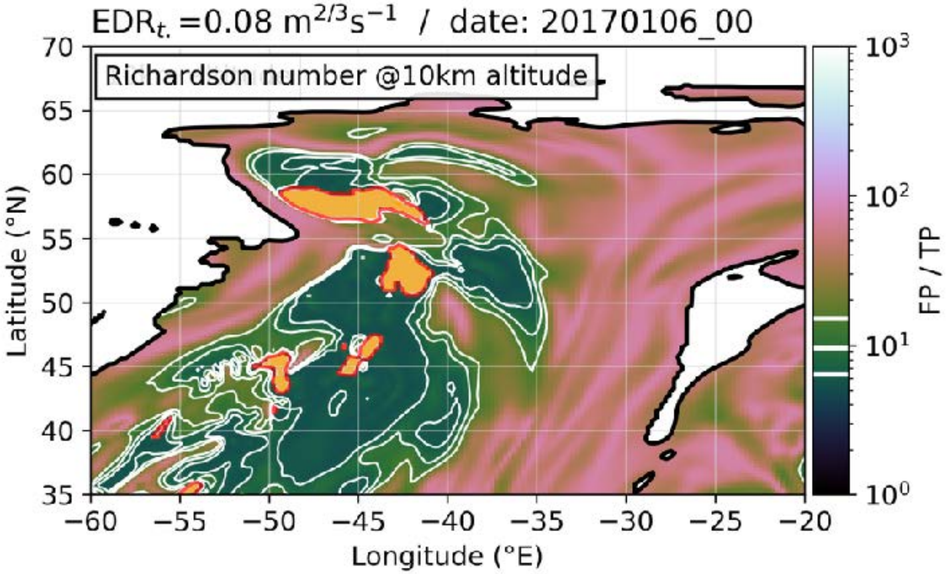
Inference / application



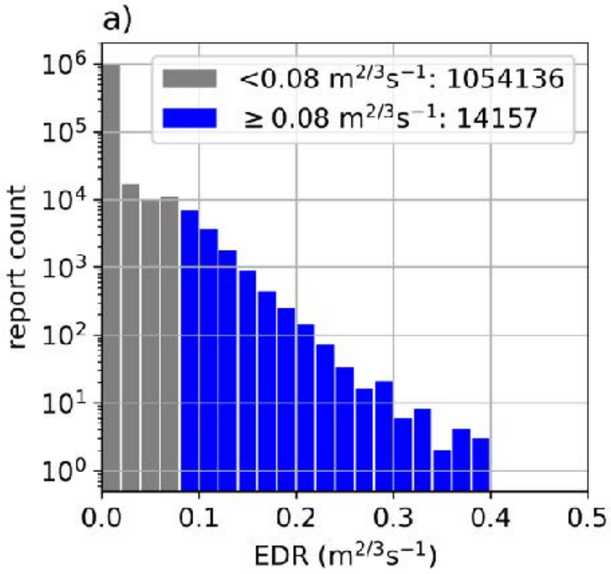


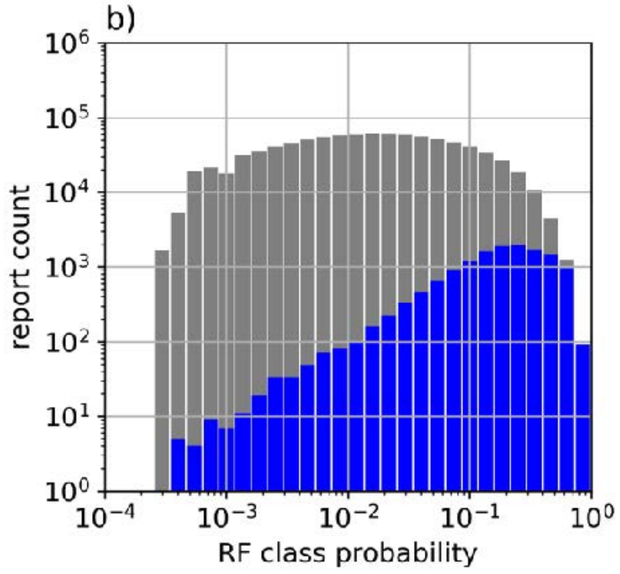
Inference / application

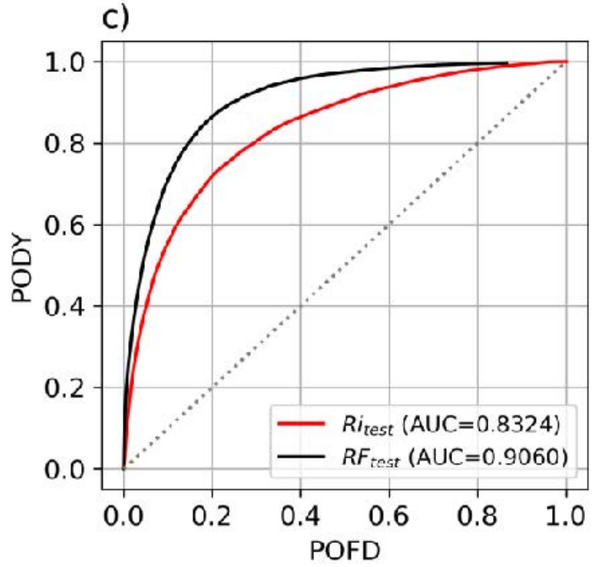


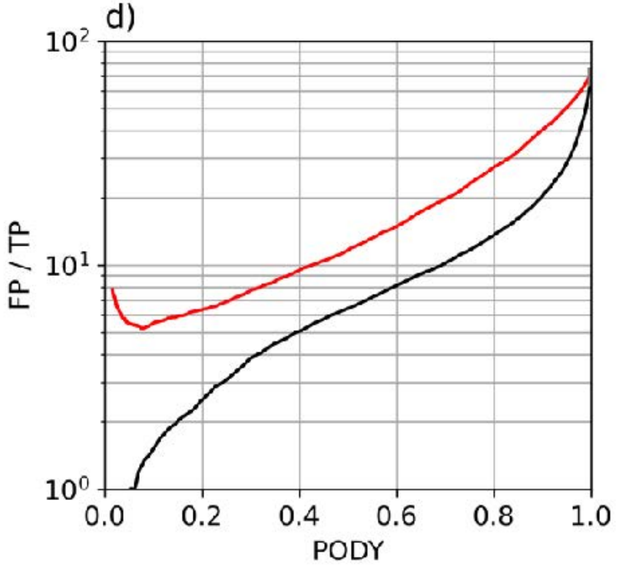


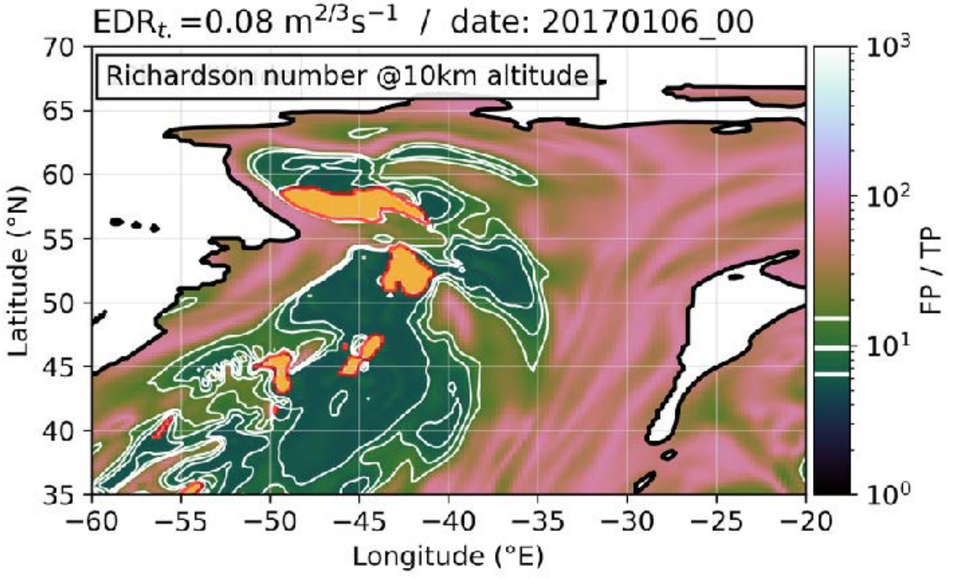
Inference / application

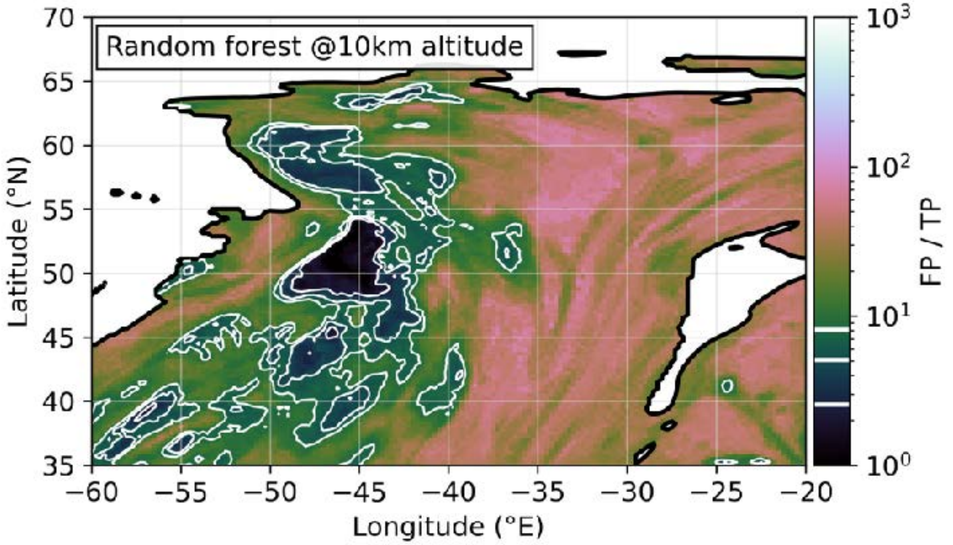




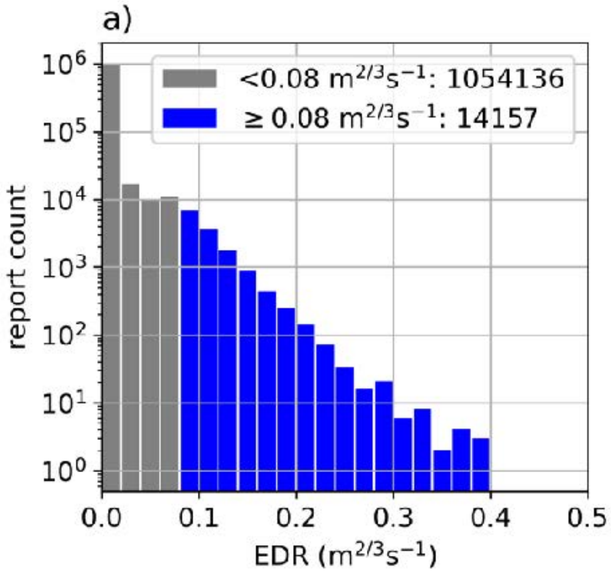


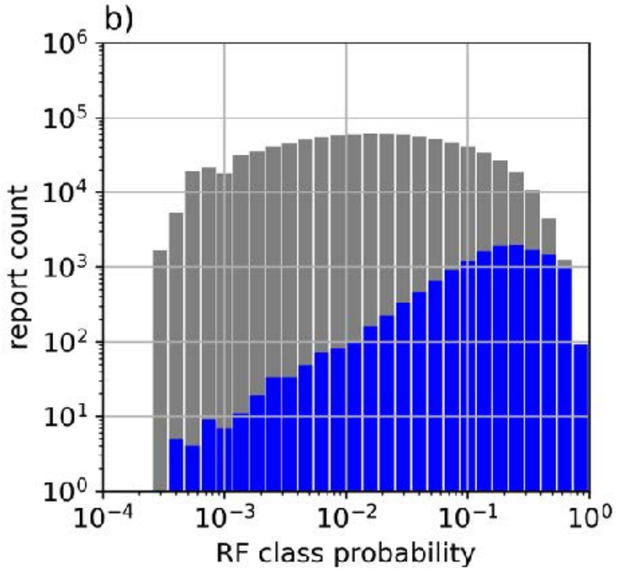


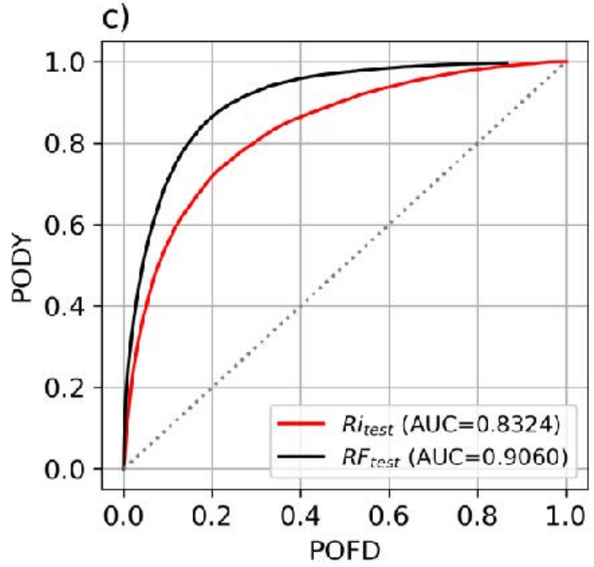


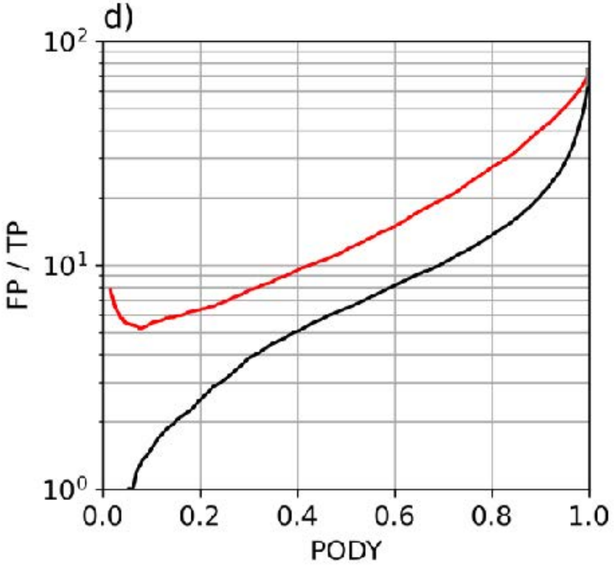


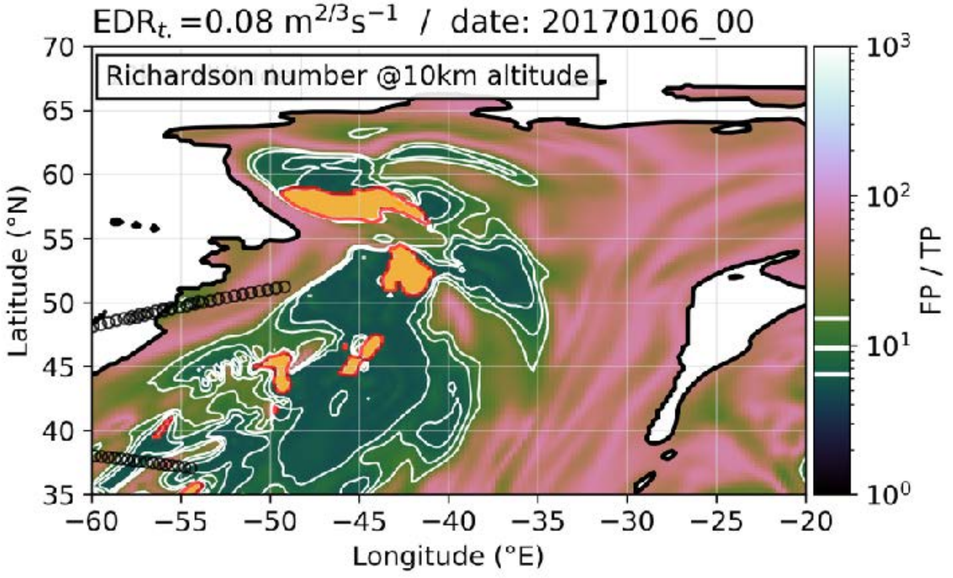
Inference / application

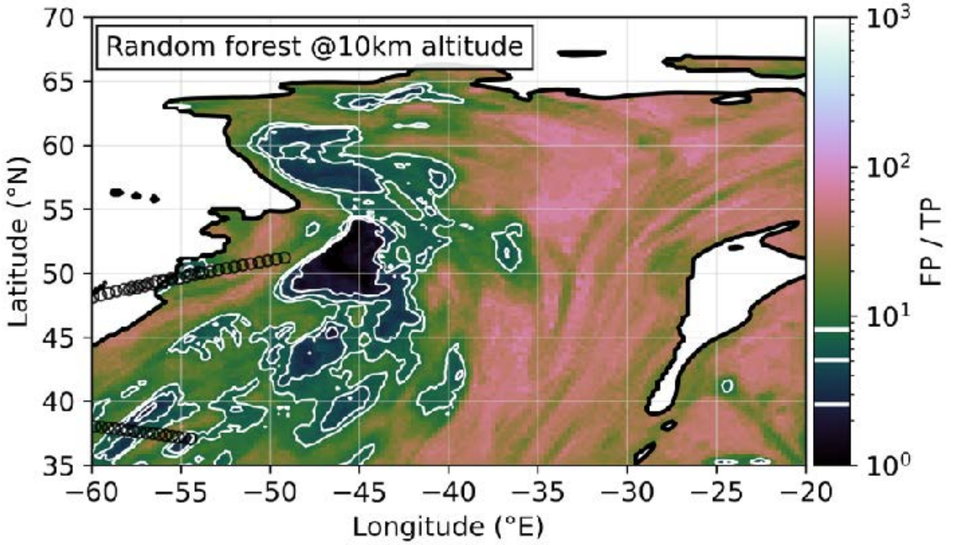




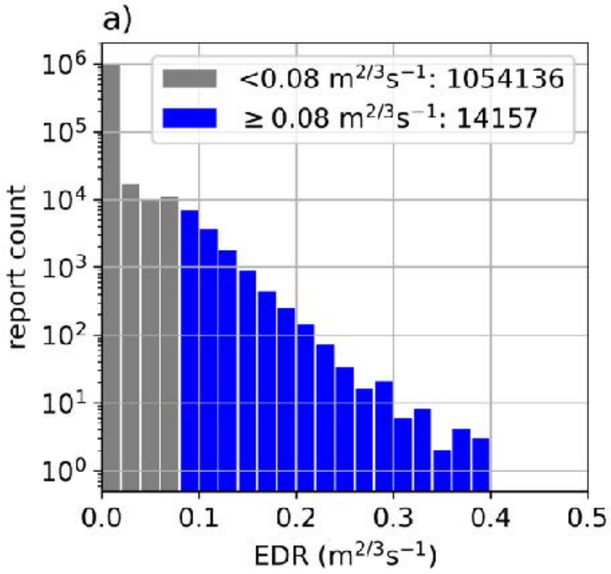


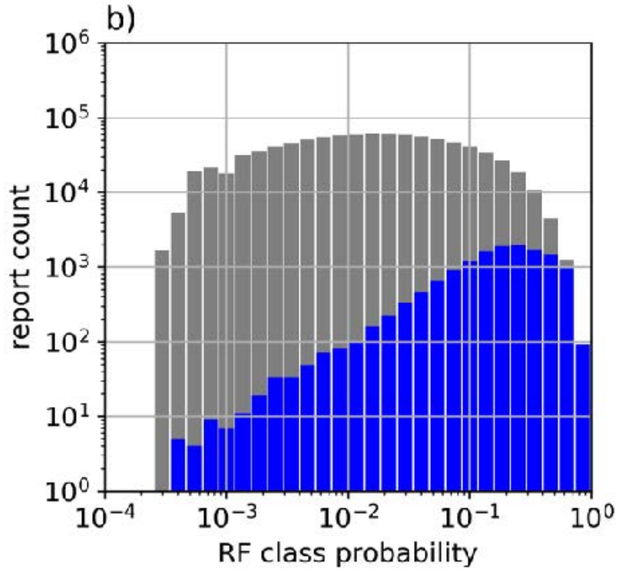


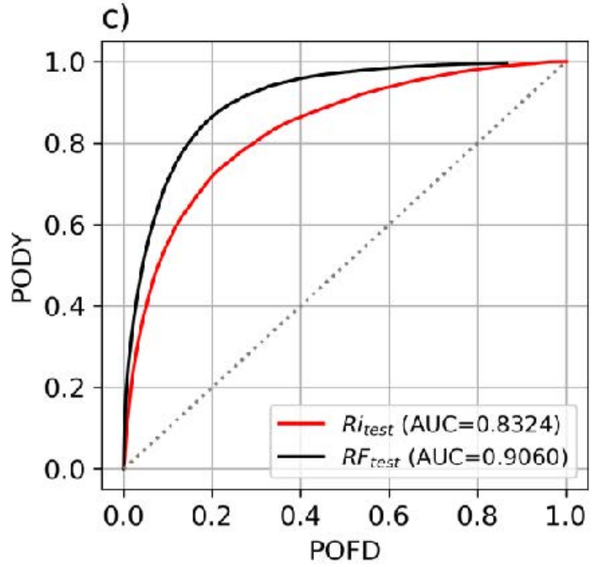


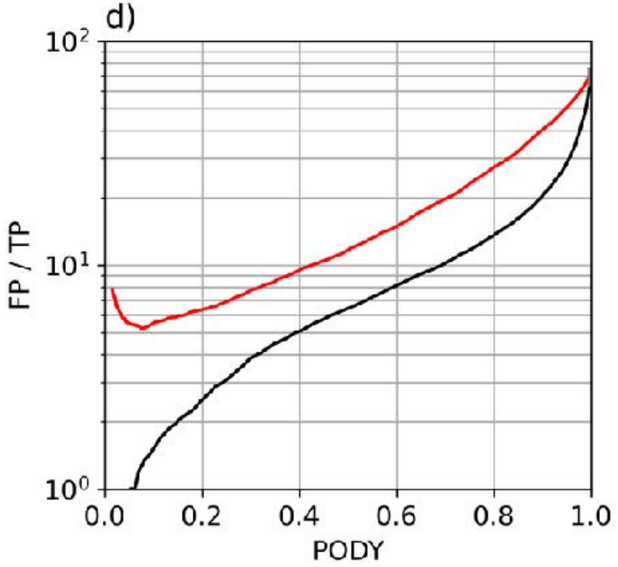


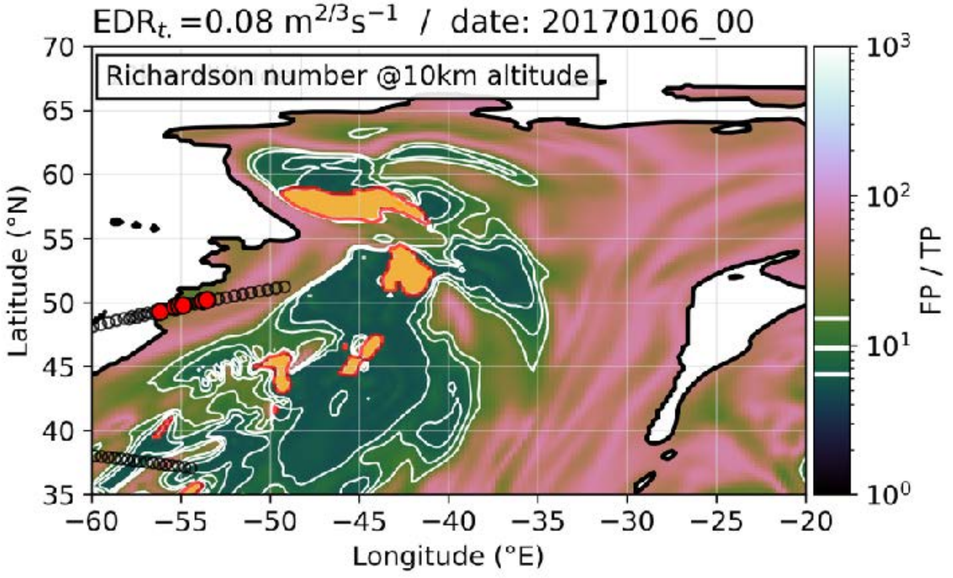
Inference / application

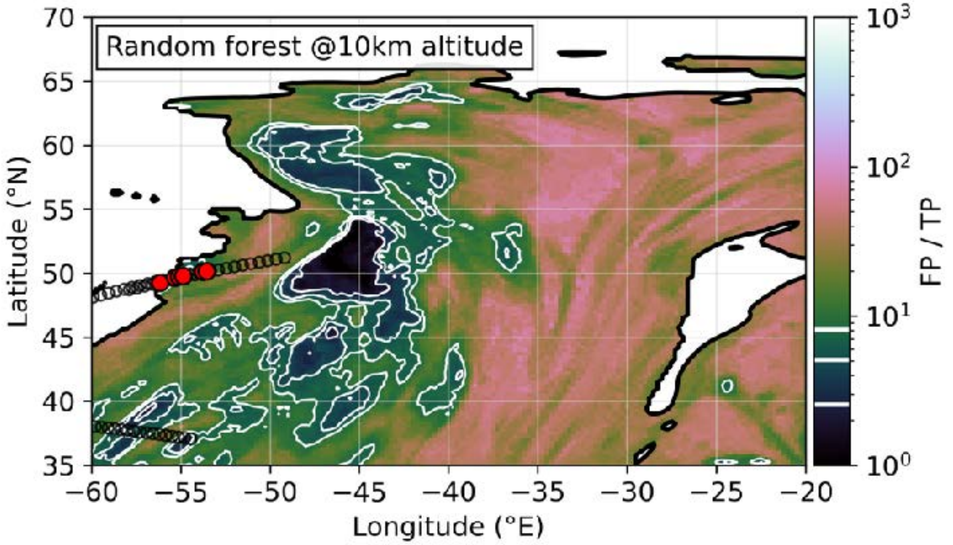




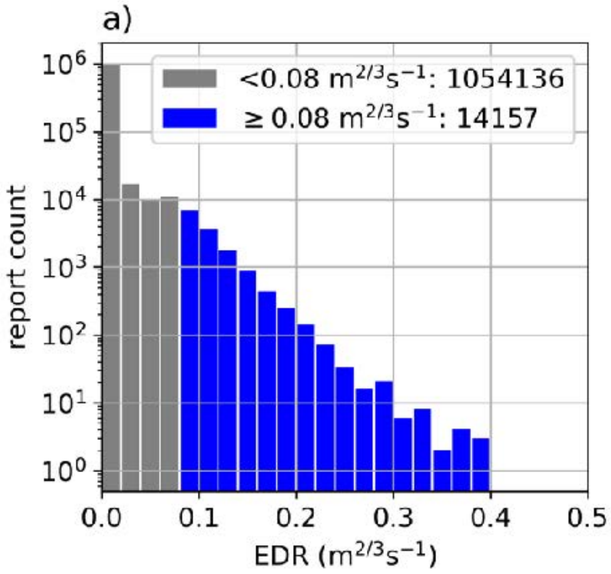


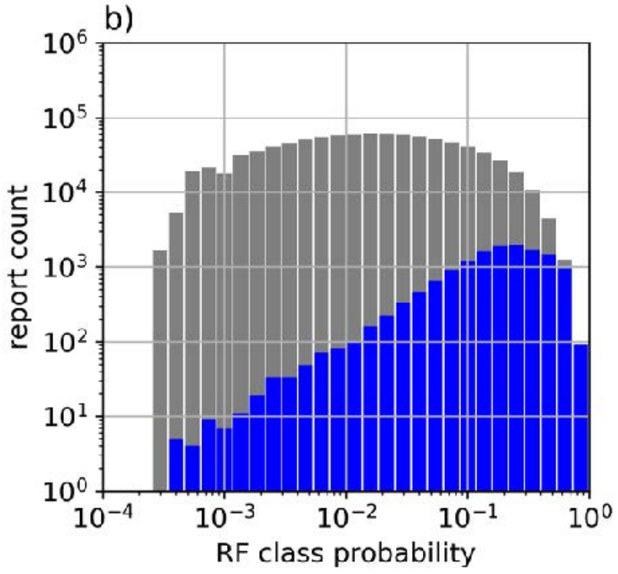


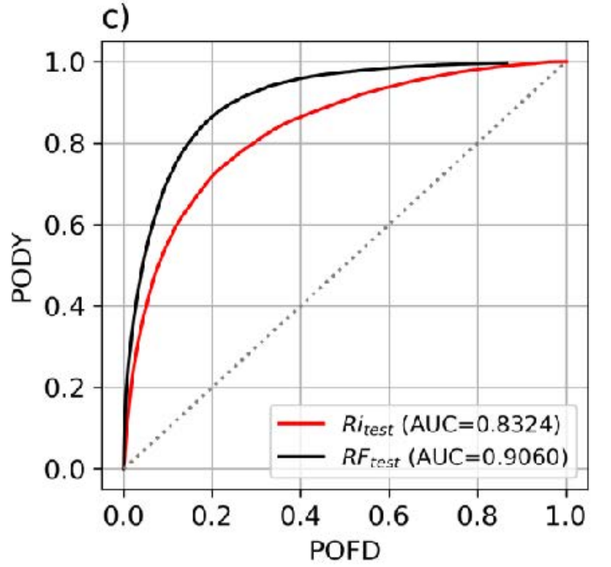


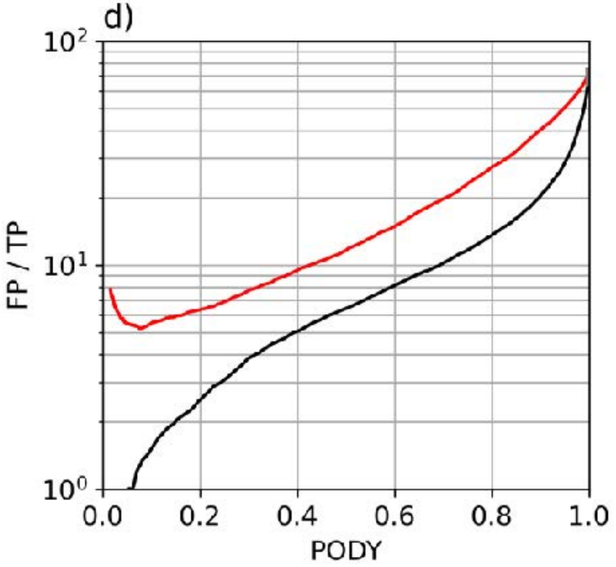


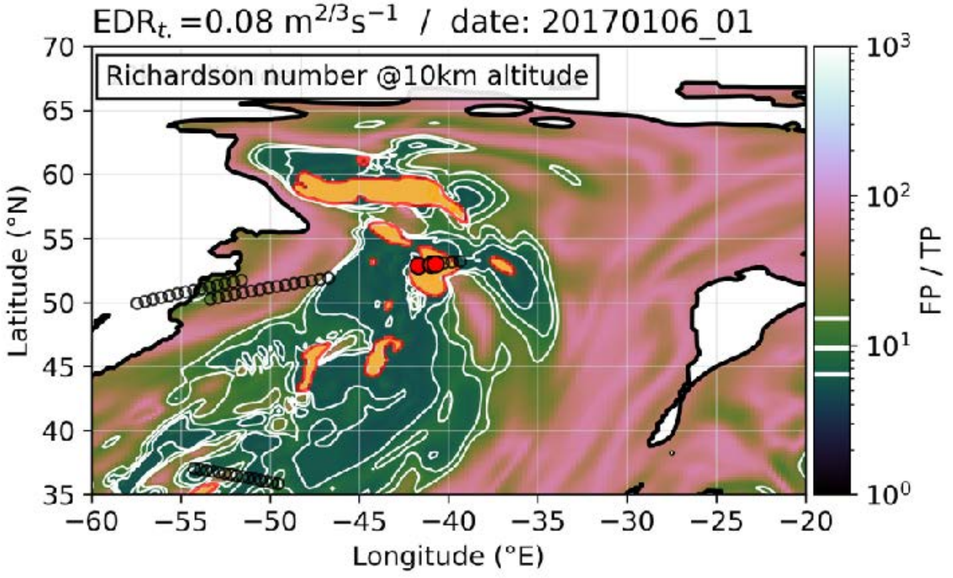
Inference / application

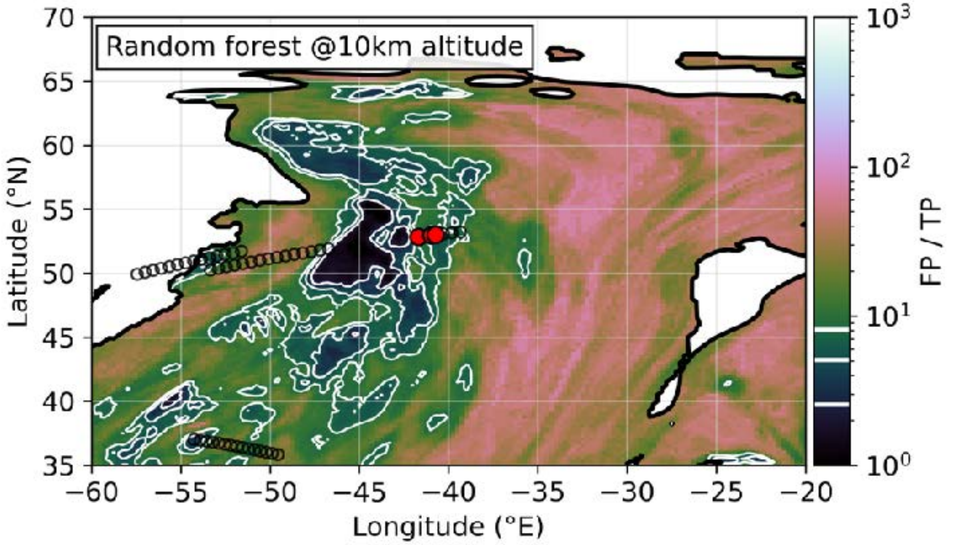




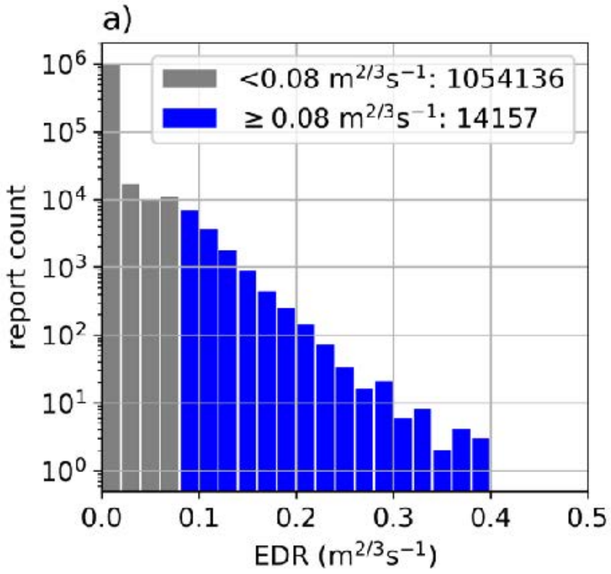


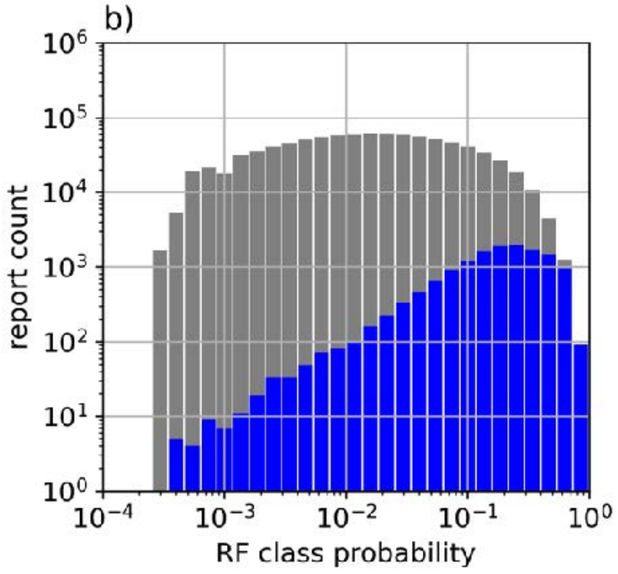


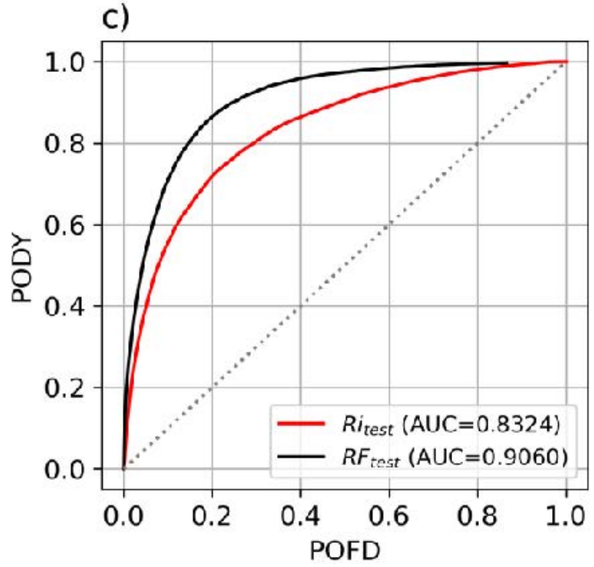


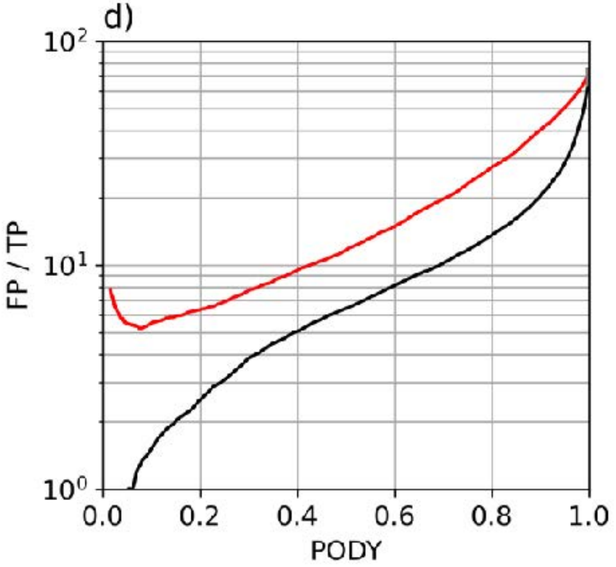


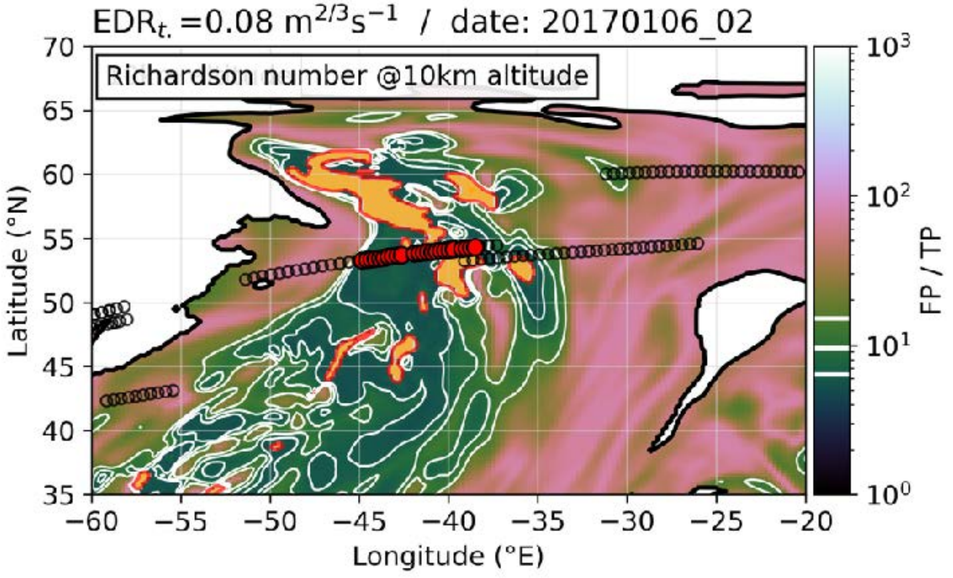
Inference / application

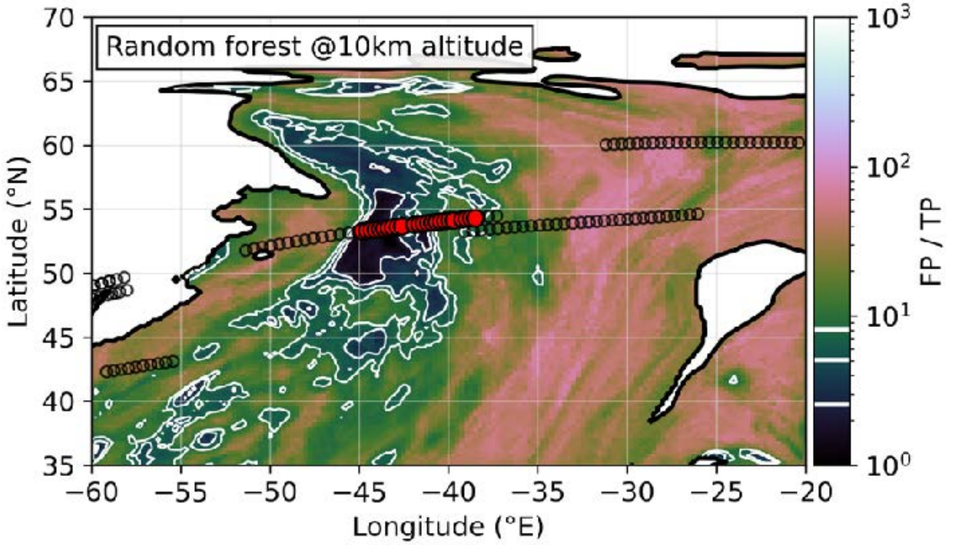




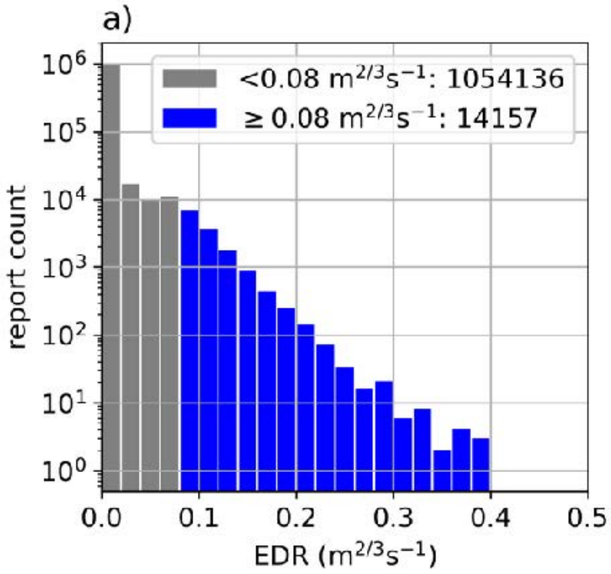


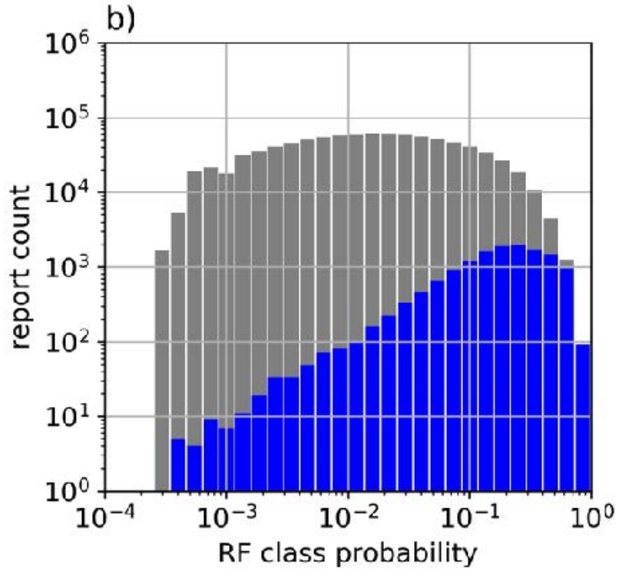


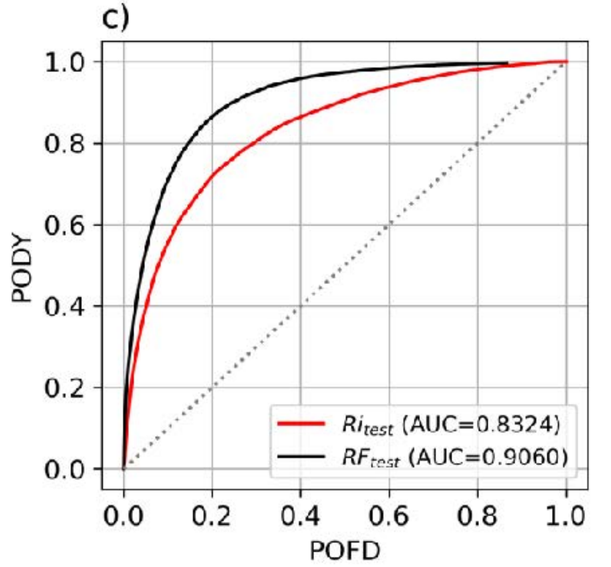


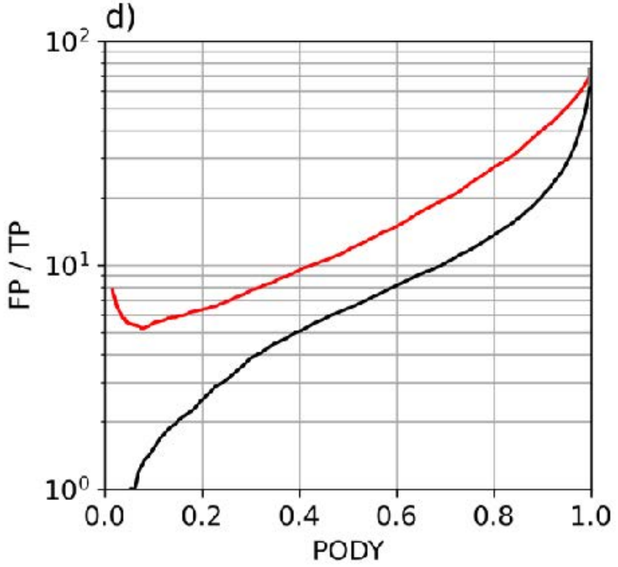


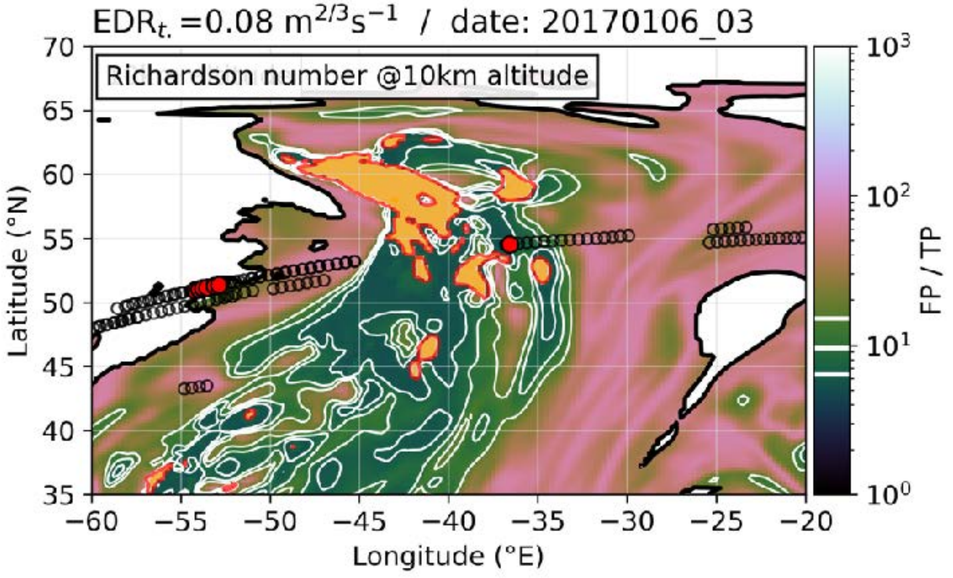
Inference / application

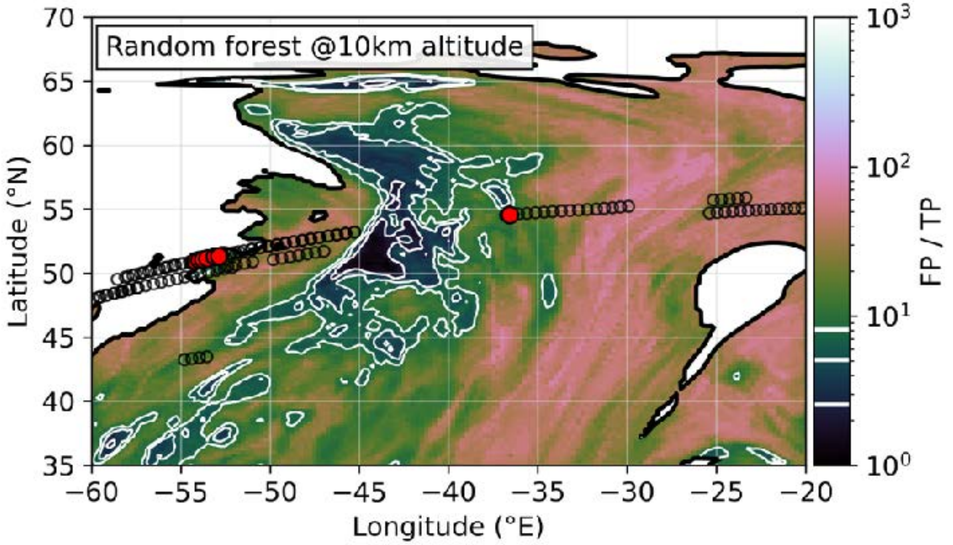




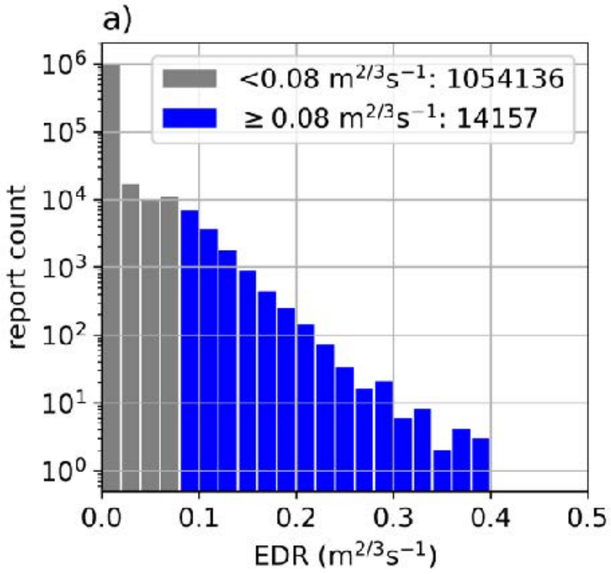


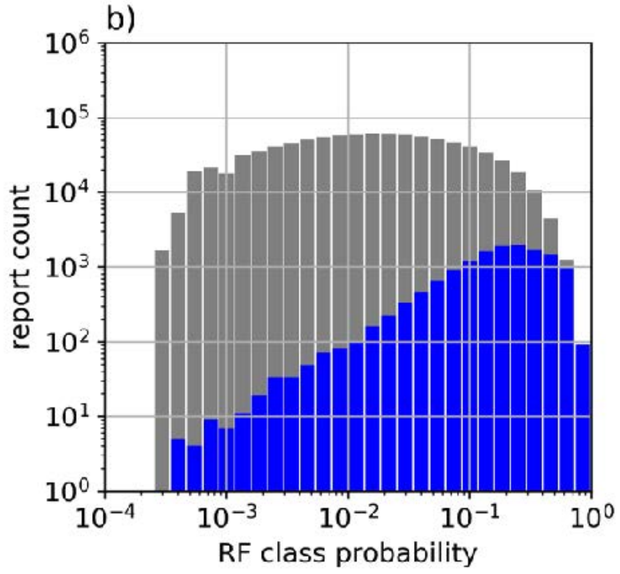


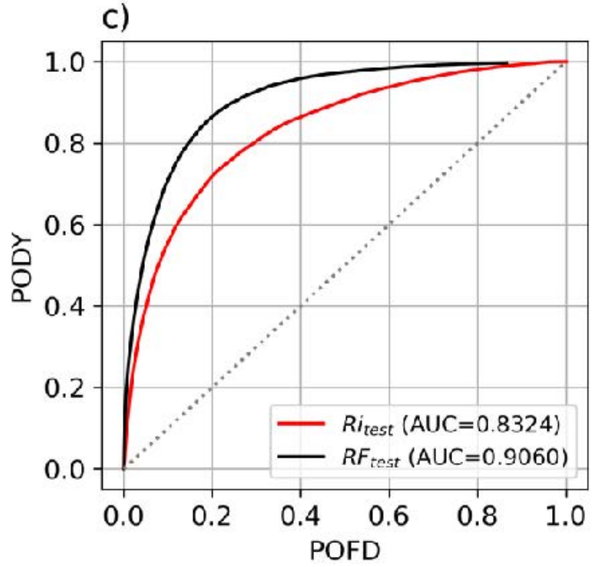


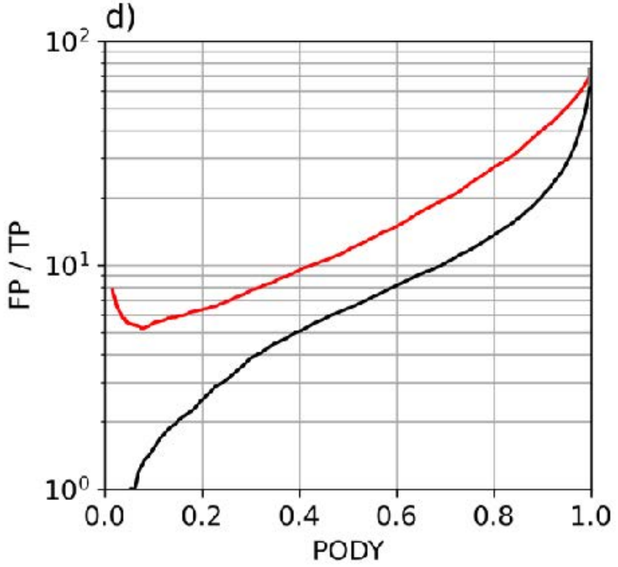


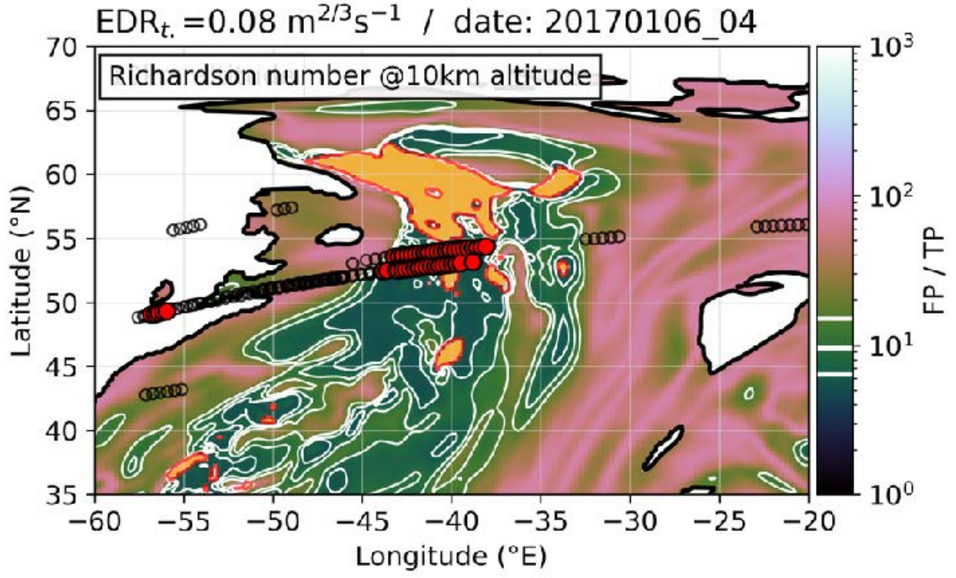
Inference / application

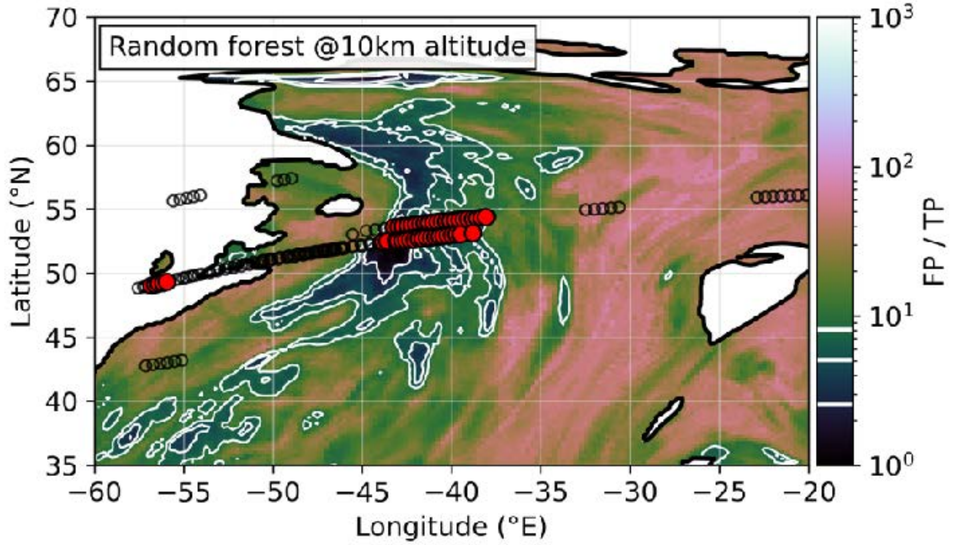




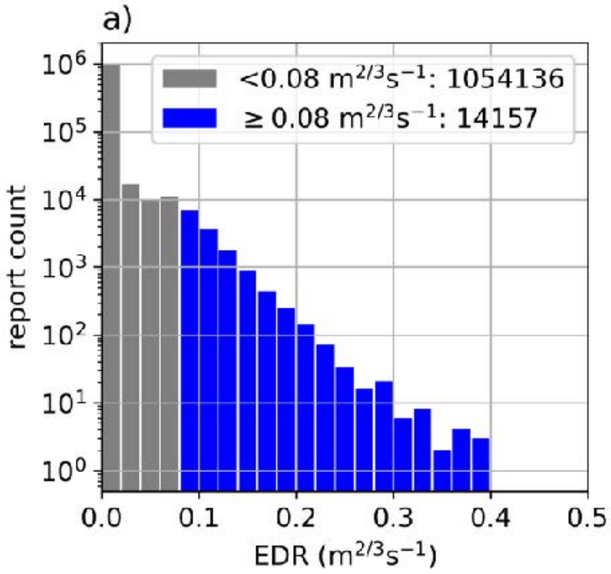


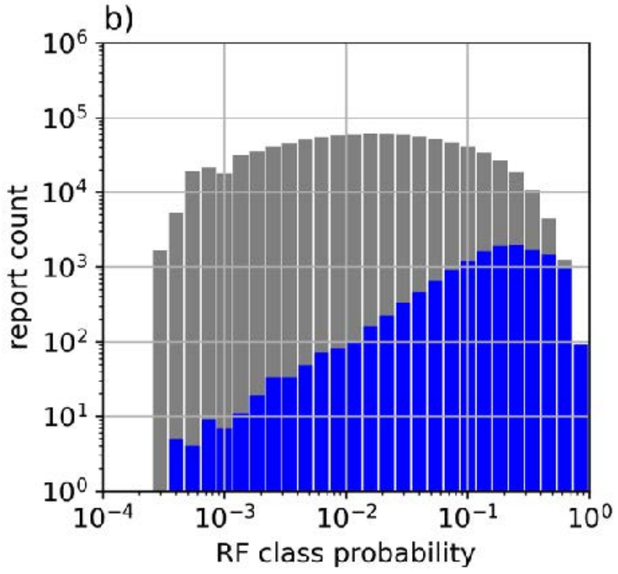


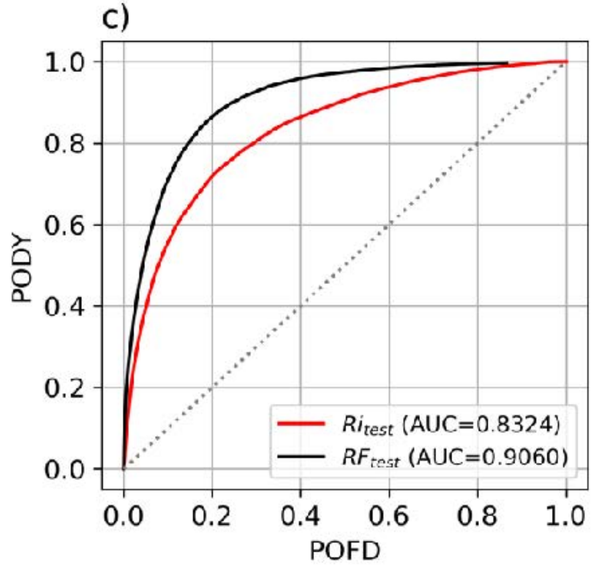


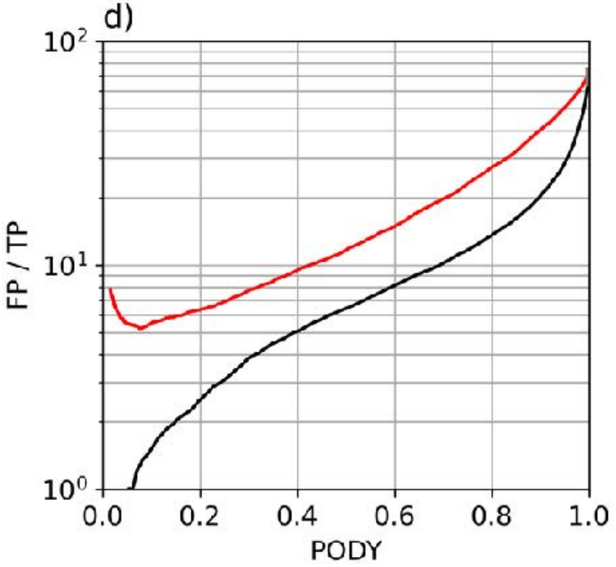


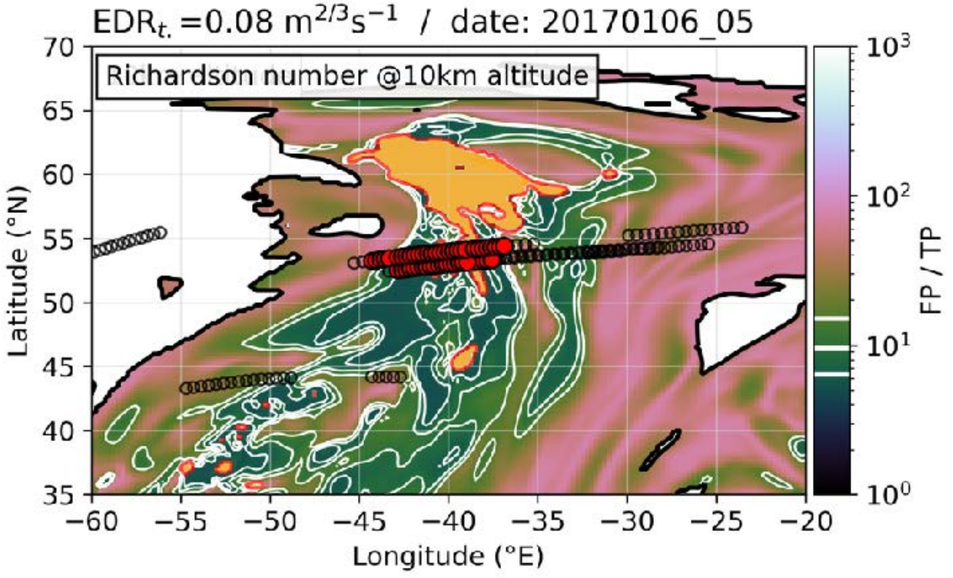
Inference / application

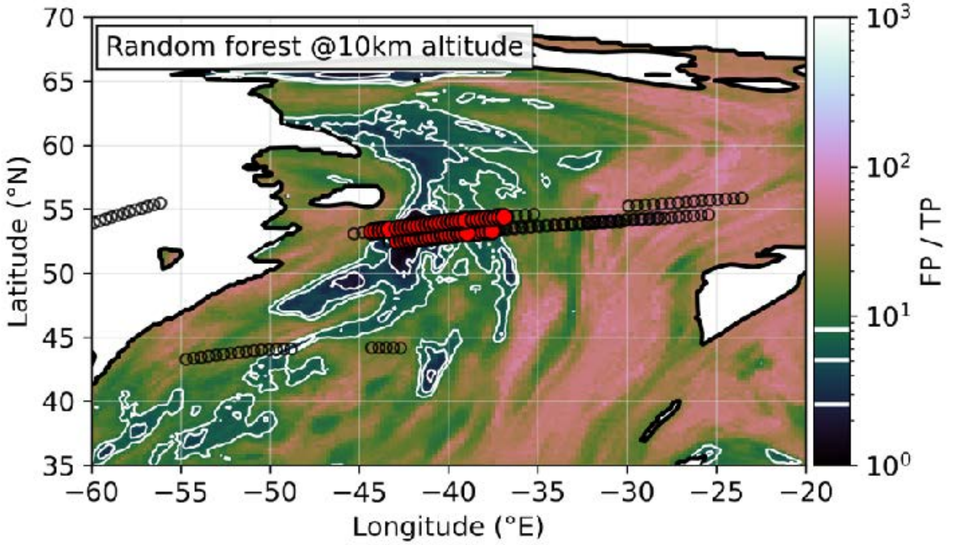




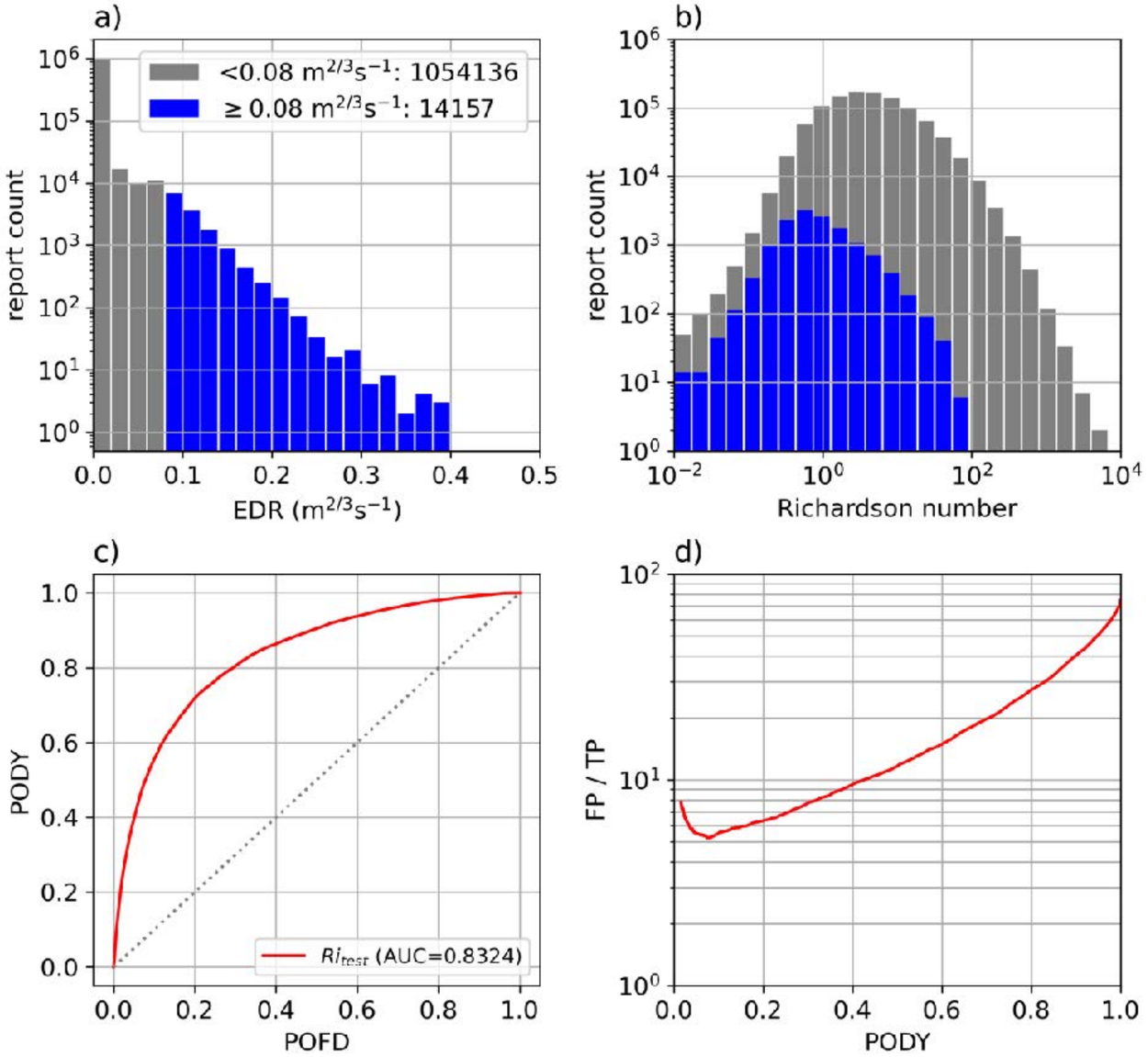


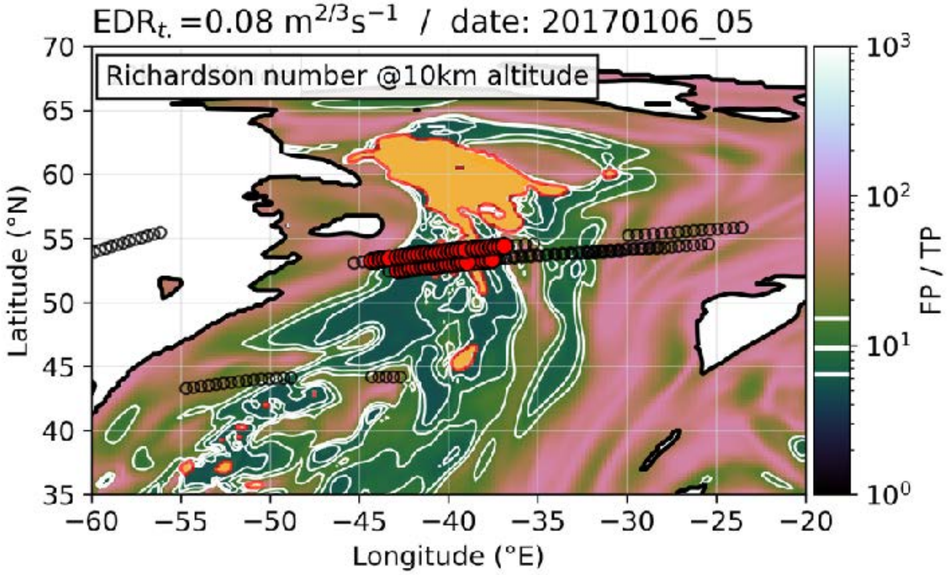




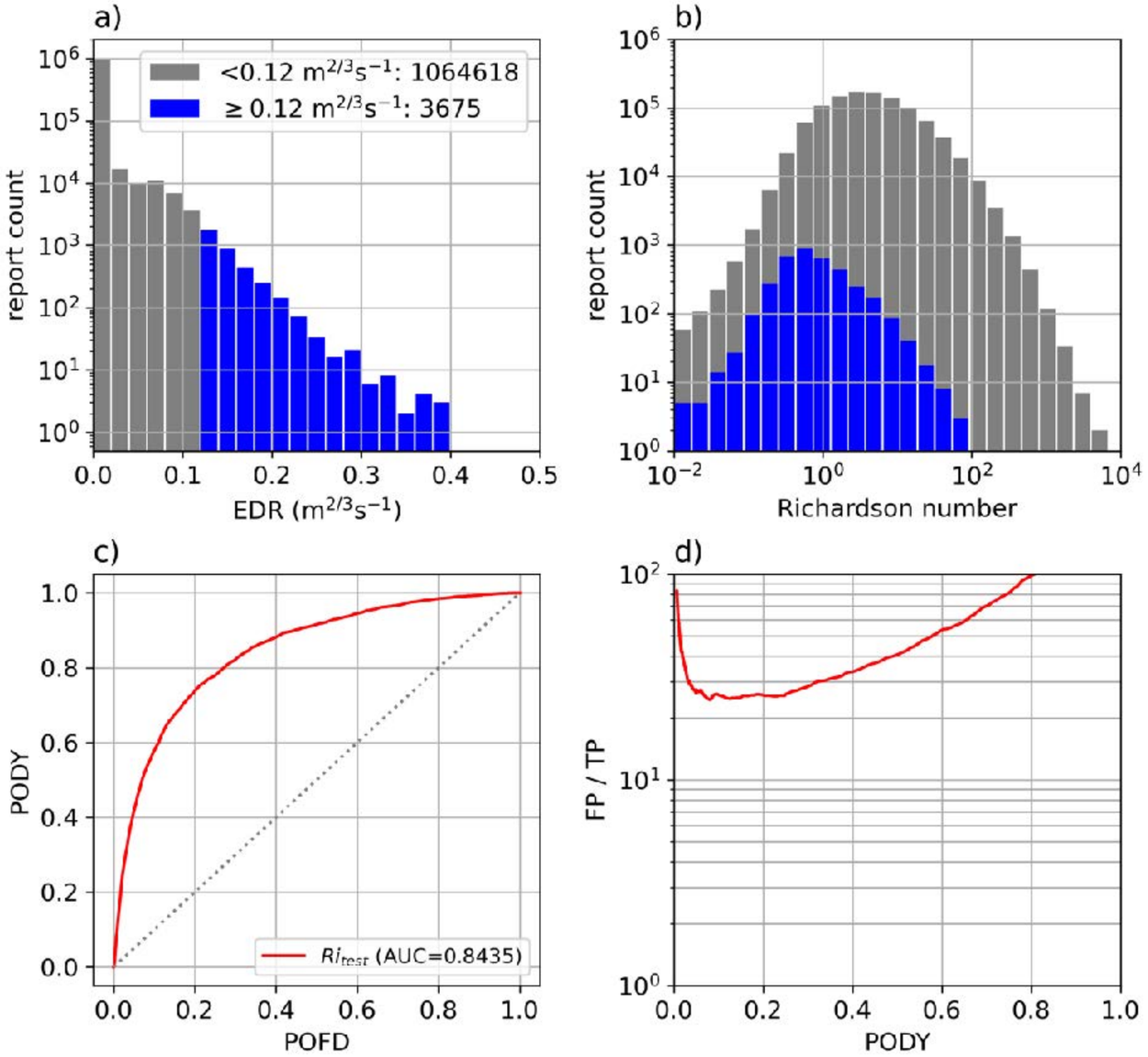


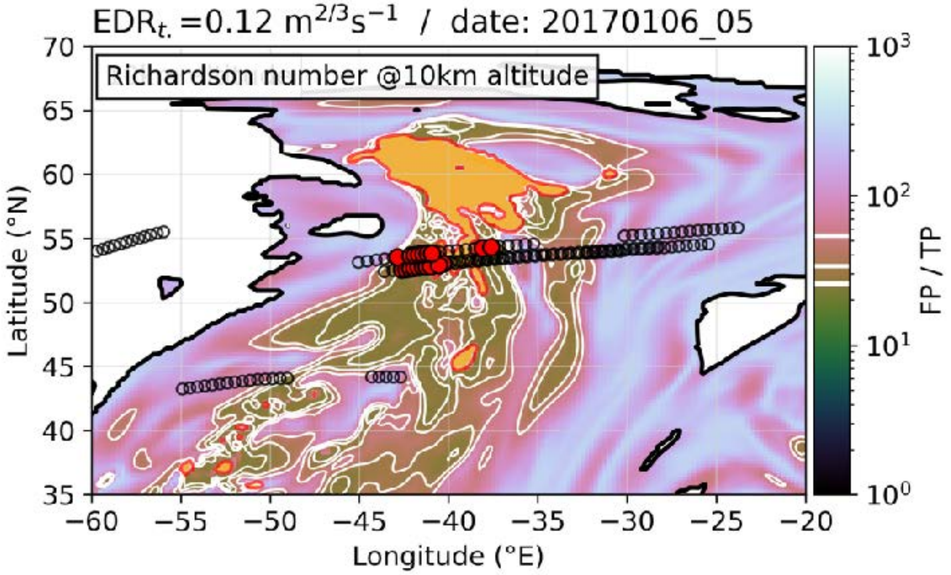
Inference / application



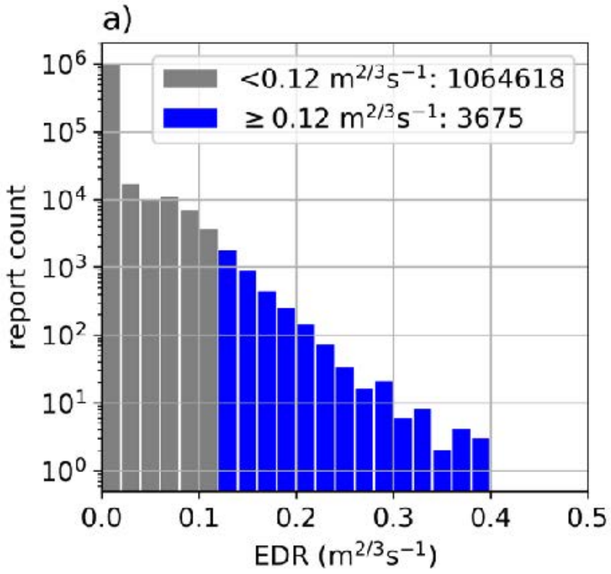


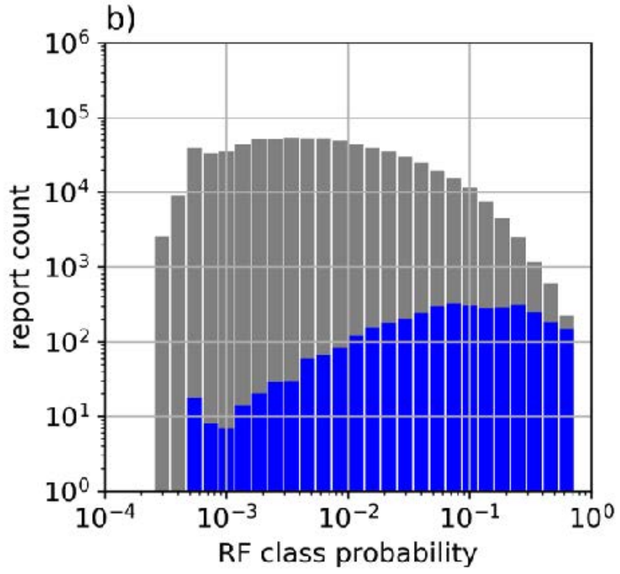
Inference / application

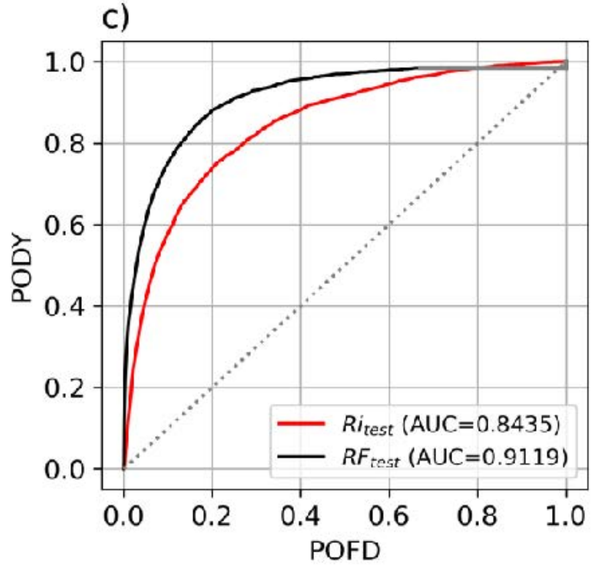


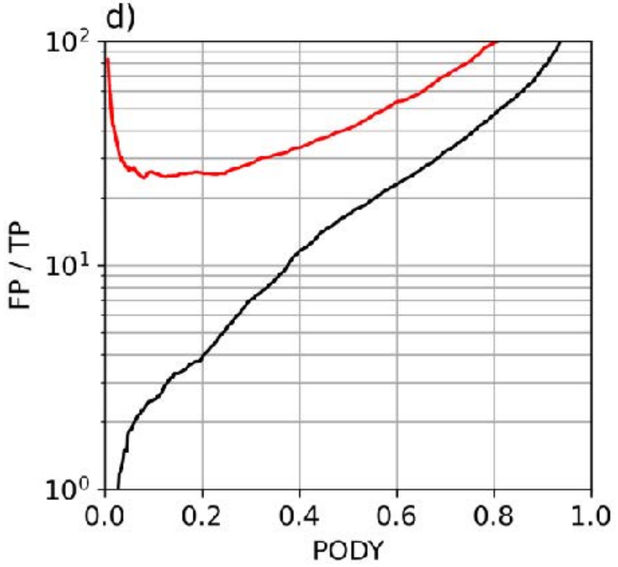


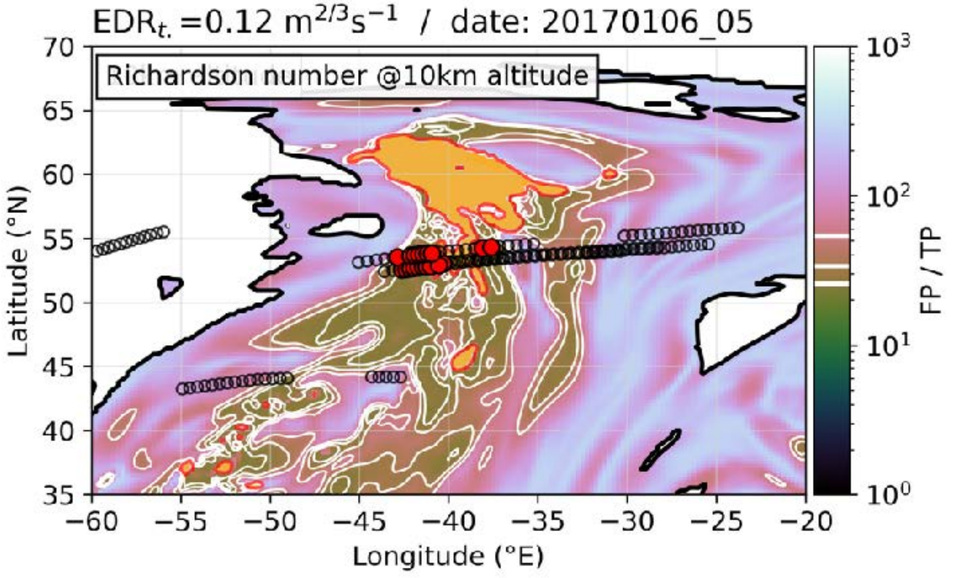
Inference / application

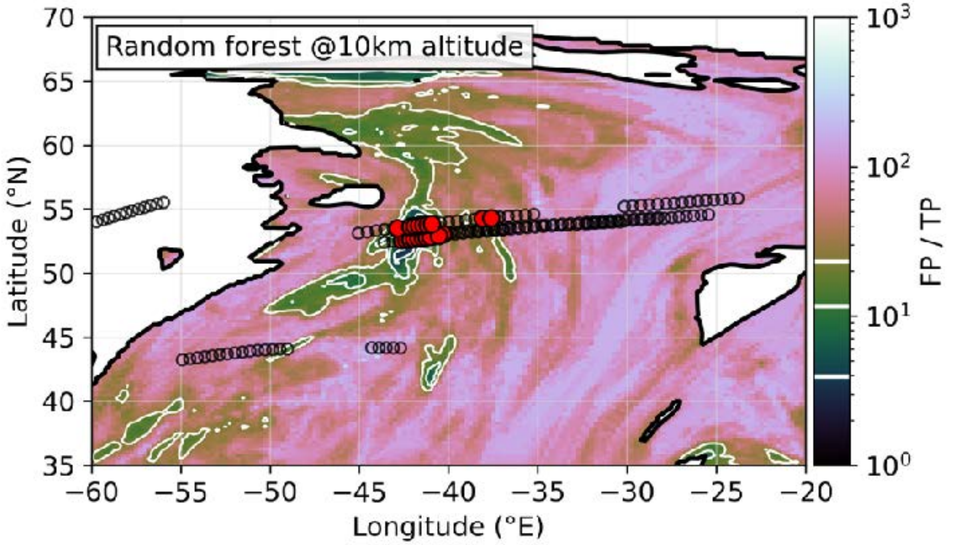




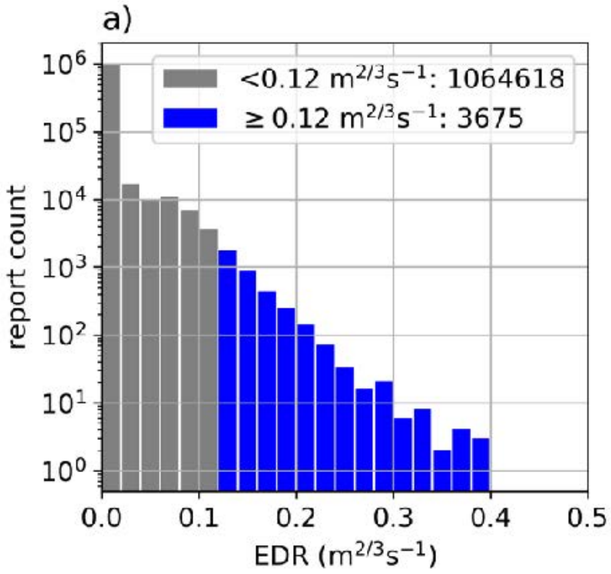


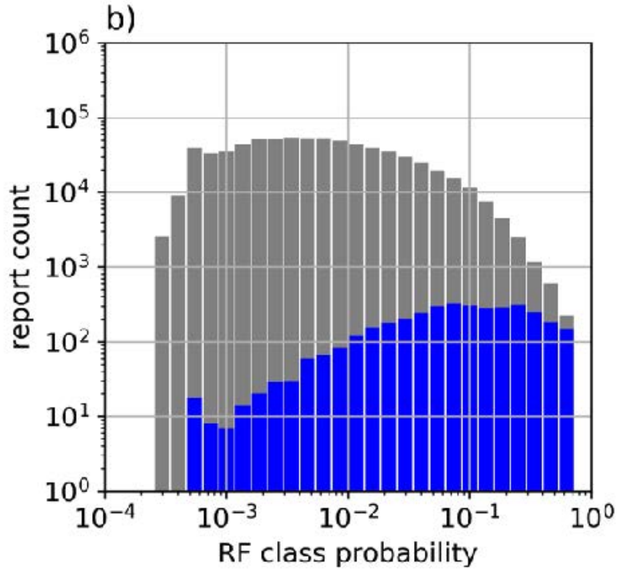


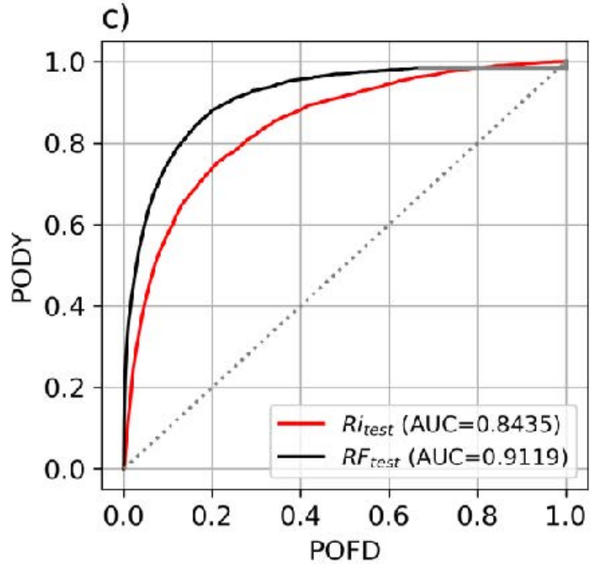


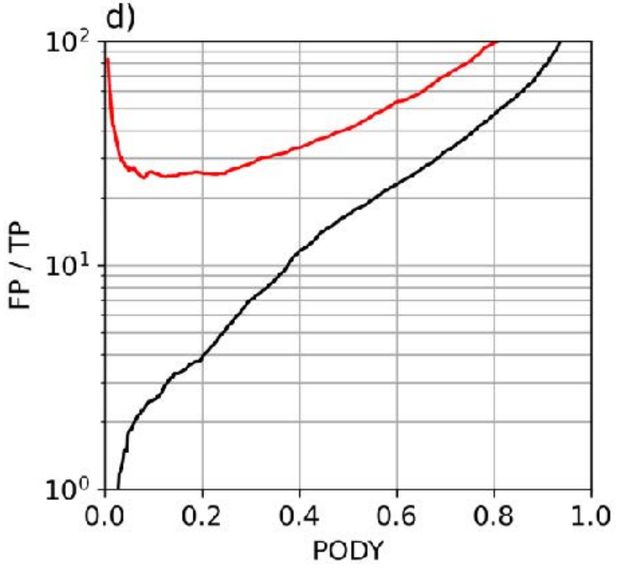


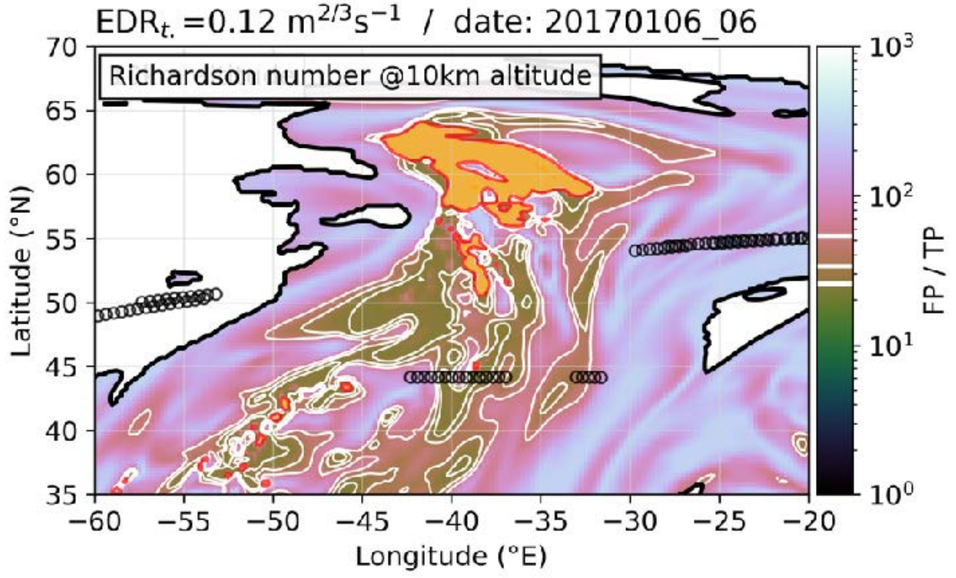
Inference / application

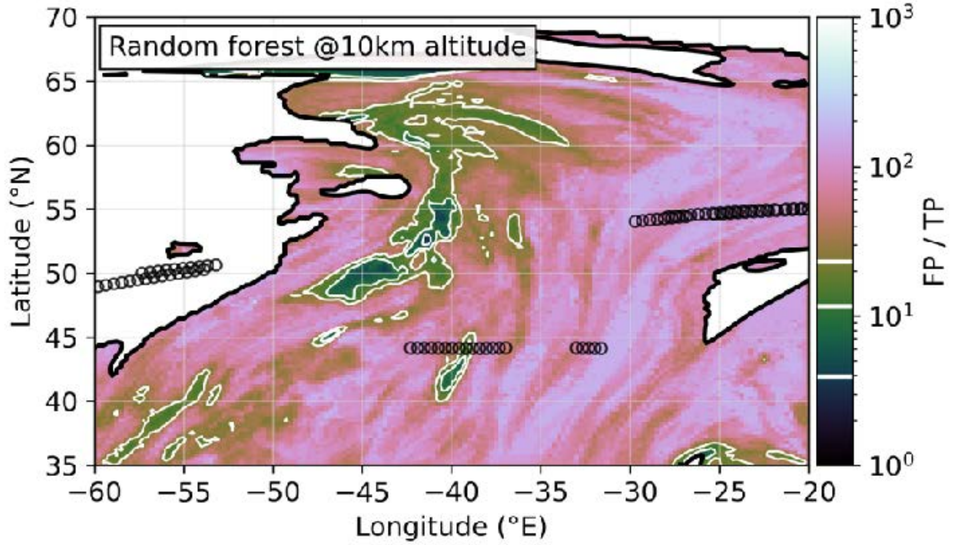




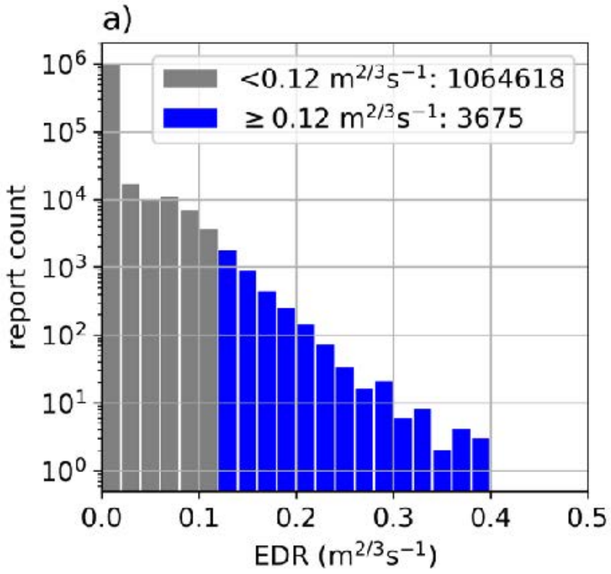


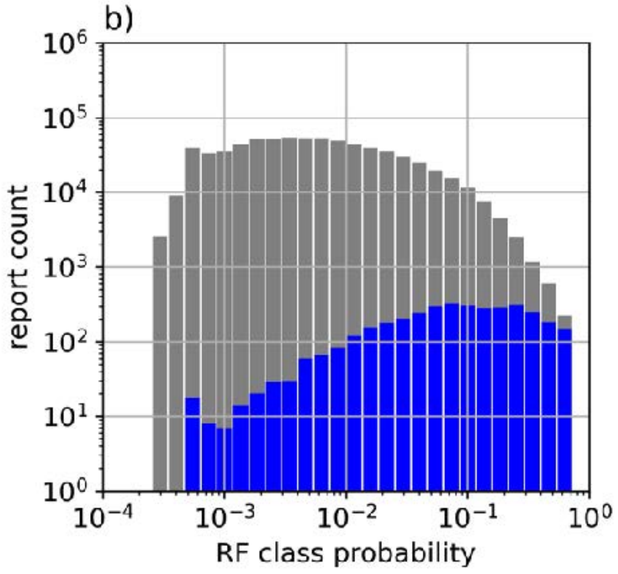


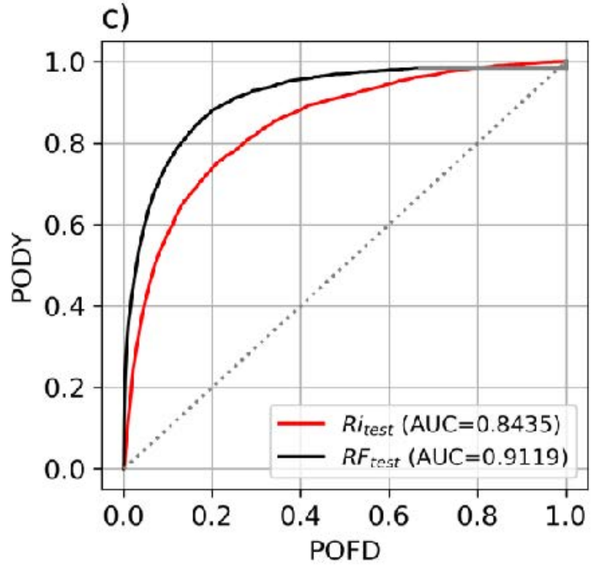


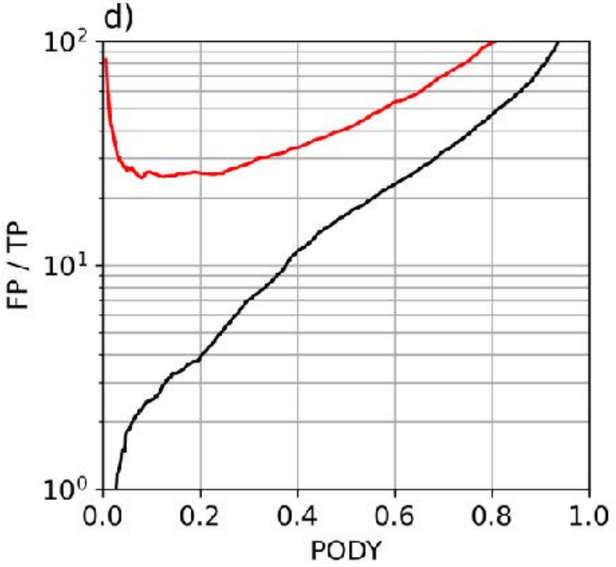


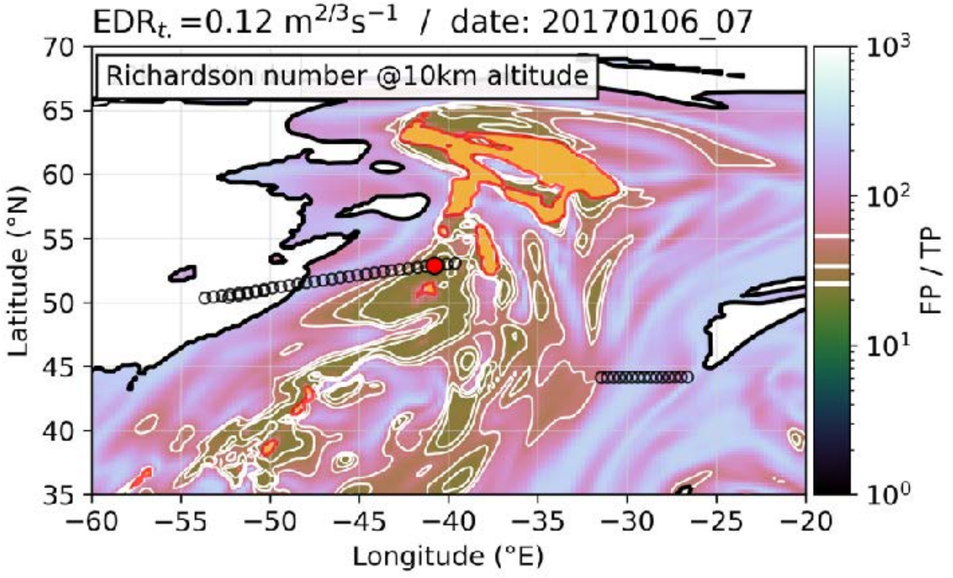
Inference / application

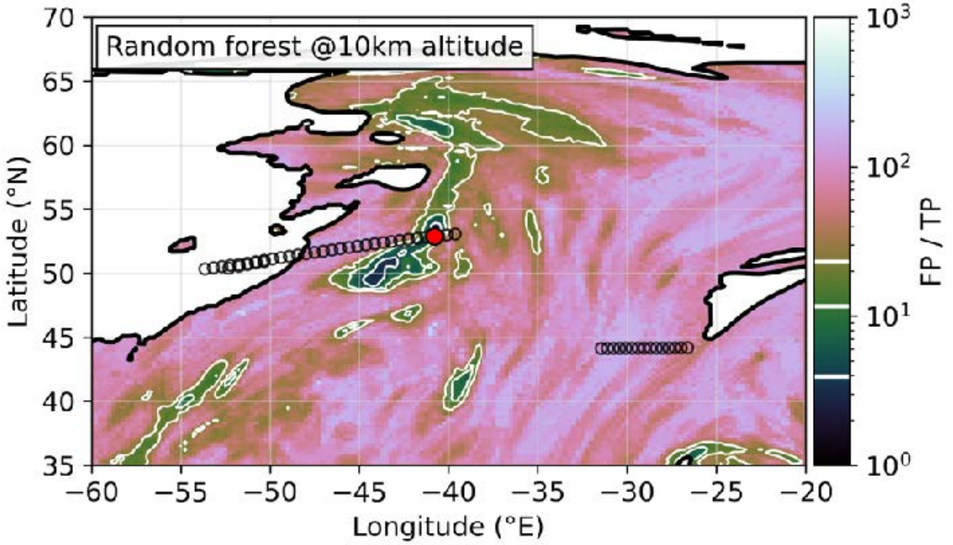




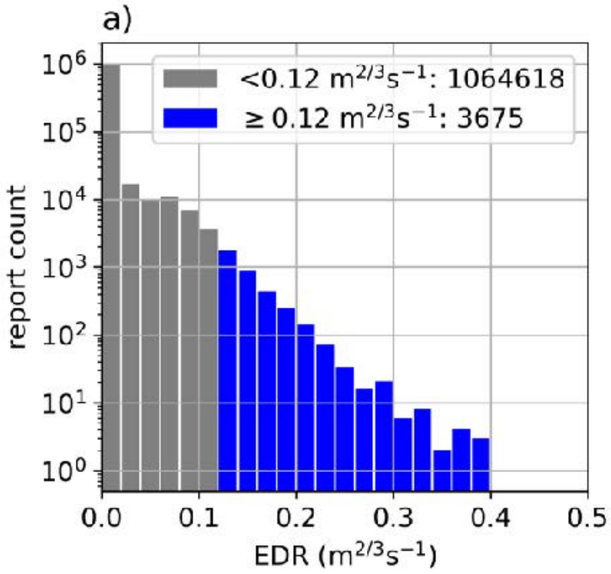


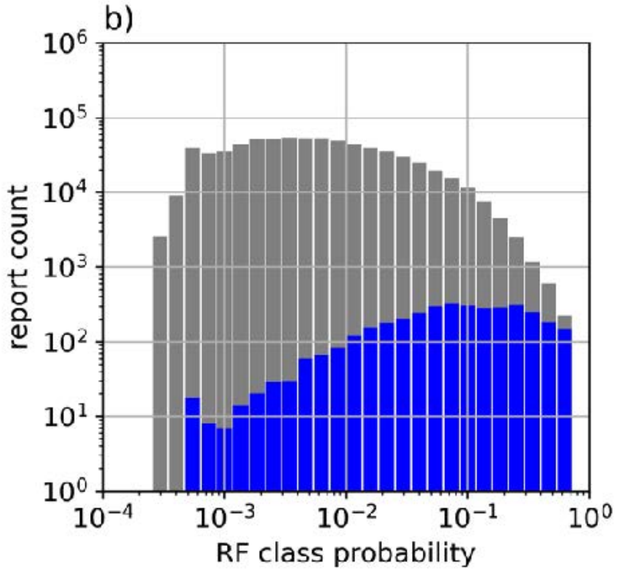


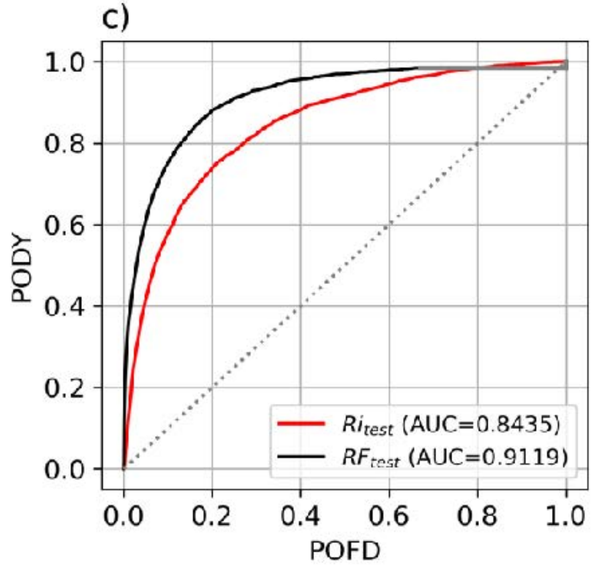


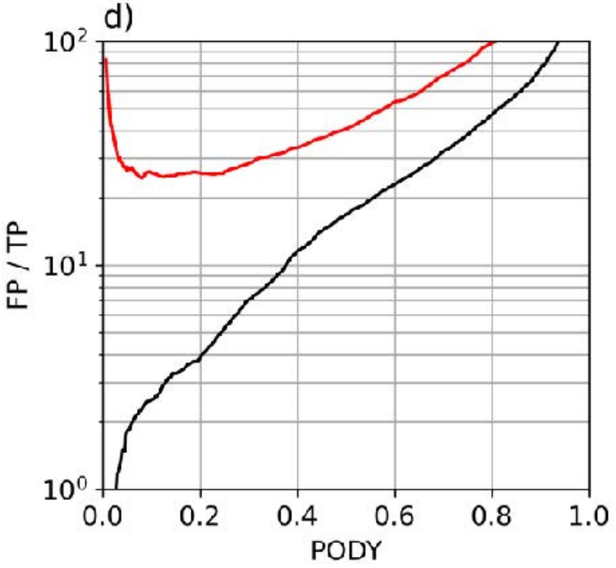


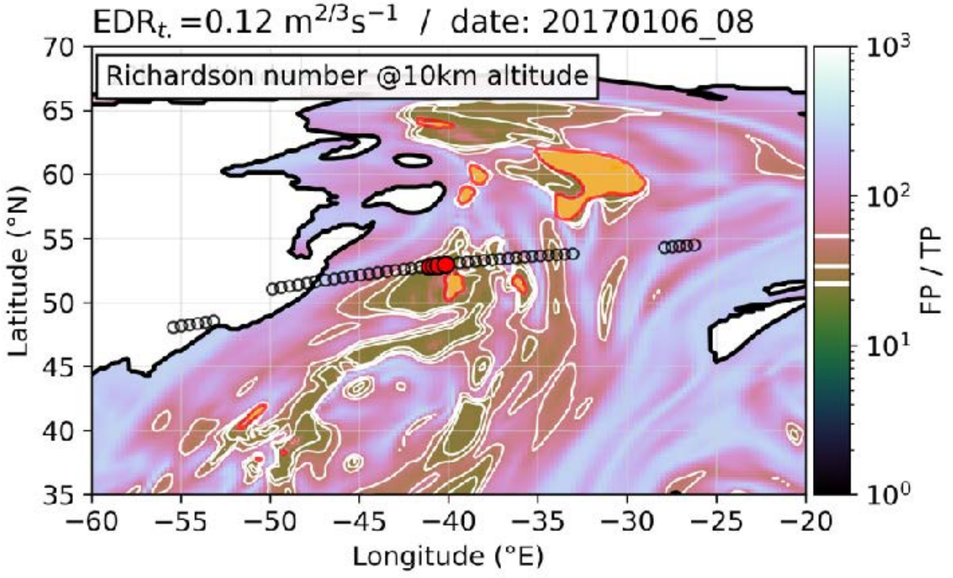
Inference / application

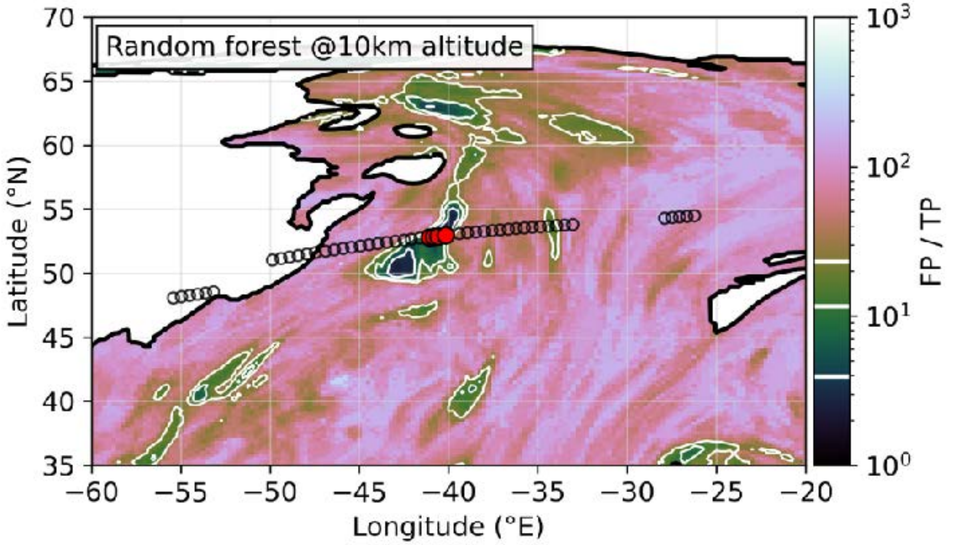




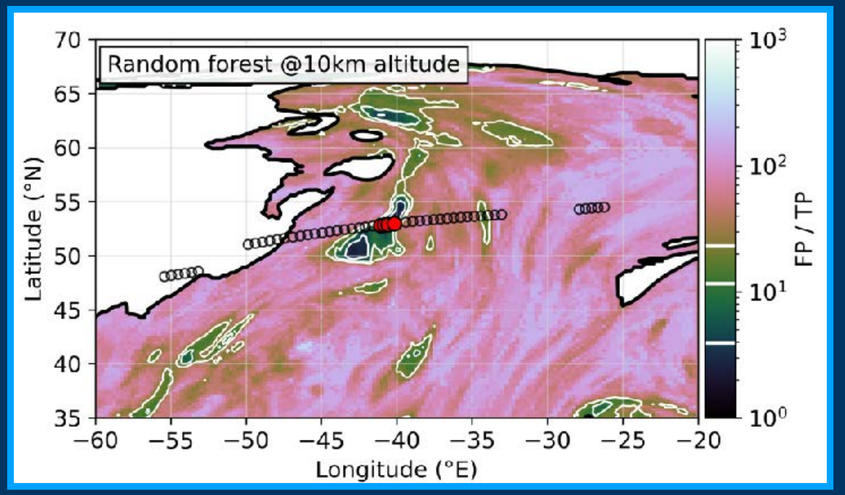








Applicability



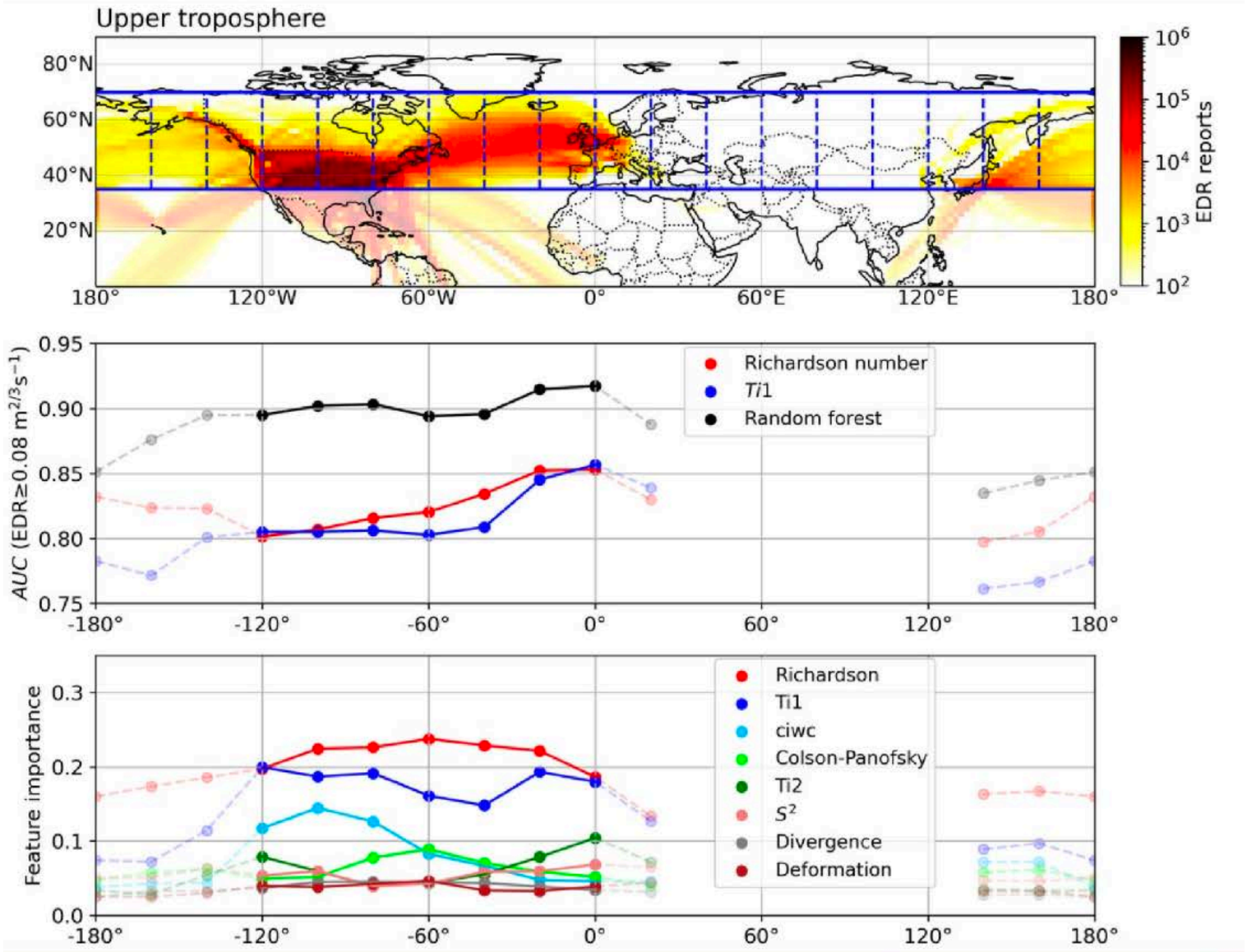
- Probability of occurrence of turbulence in ERA5:
    - Identify geographical hotspots
    - Hotspots in composites, relative to cyclone centre or jetstream
- Apply to IAGOS, analyze trace gas gradients relative to turbulence occurrence
- Compare with Thorpe analysis derived turbulence occurrence
- Turbulence forecast, for aviation or measurement campaigns
    - Comparison e.g. with GTG
- Analyze easiest to predict turbulence reports vs. hardest to predict
    - —> why do known diagnostics fail?

- Apply technique to any higher resolution NWP data (training + inference)

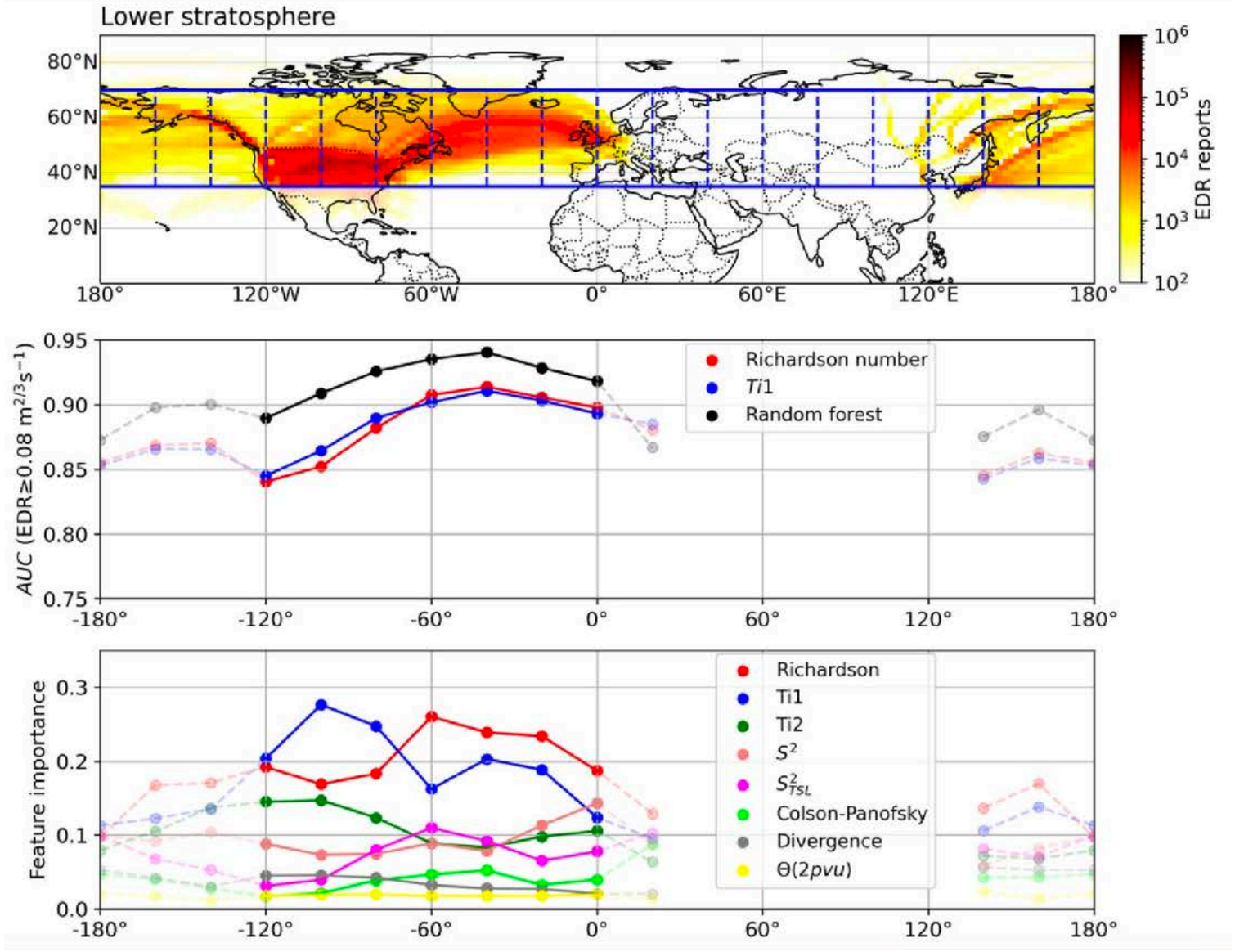
Thank you for your attention

- This study was made possible due to the data made available to the National Oceanic and Atmospheric Administration by the following commercial airlines: American, Delta, Federal Express, Northwest, United, and United Parcel Service.
- The ECMWF provided the ERA5 data which has been retrieved from the MARS system.
- In-service aircraft for a global observing system (IAGOS) data were created with support from the European Commission, national agencies in Germany (BMBF), France (MESR), and the UK (NERC), and the IAGOS member institutions (http://www.iagos.org/partners). The participating airlines (Lufthansa, Air France, Austrian, China Airlines, Iberia, Cathay Pacific, Air Namibia, Sabena) supported IAGOS by carrying the measurement equipment free of charge since 1994.

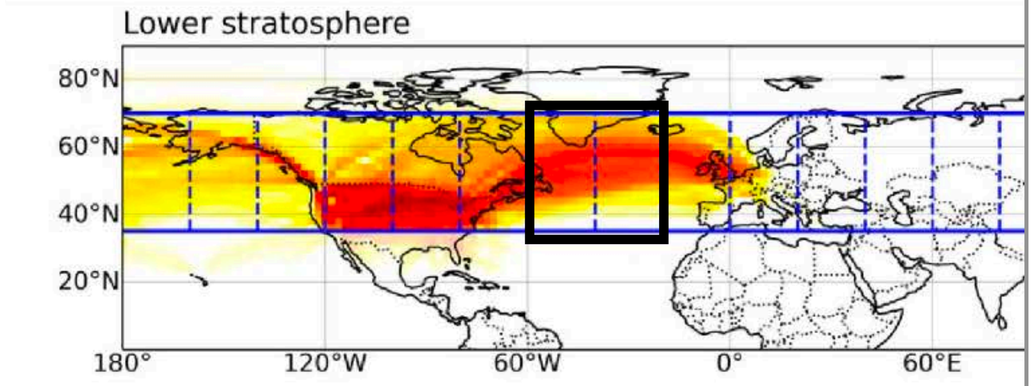
Outlook

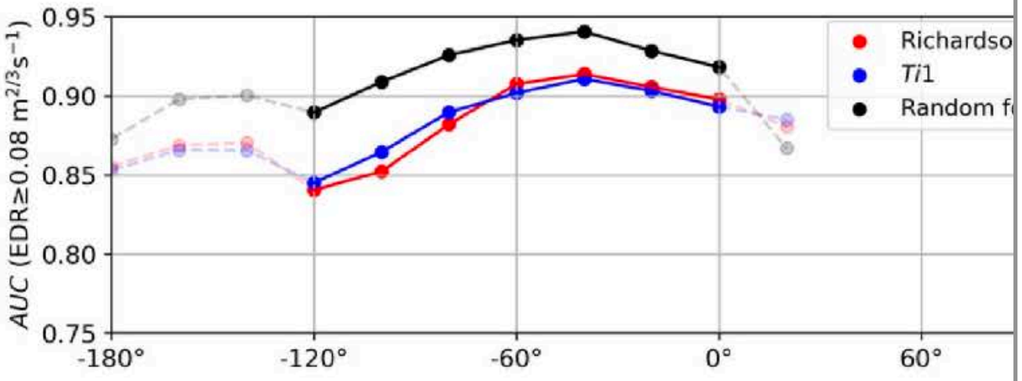

Upper troposphere
80°N
60°N
40°N
20°N
180°
120°W
60°W
0°
60°E
120°E
180°
10^6
10^5
10^4
10^3
10^2
EDR reports
AUC (EDR≥0.08 m$^{2/3}$s^{-1})
0.95
0.90
0.85
0.80
0.75
-180°
-120°
-60°
0°
60°
120°
180°
Richardson number
Ti1
Random forest
Feature importance
0.3
0.2
0.1
0.0
Richardson
Ti1
ciwc
Colson-Panofsky
Ti2
S^2
Divergence
Deformation

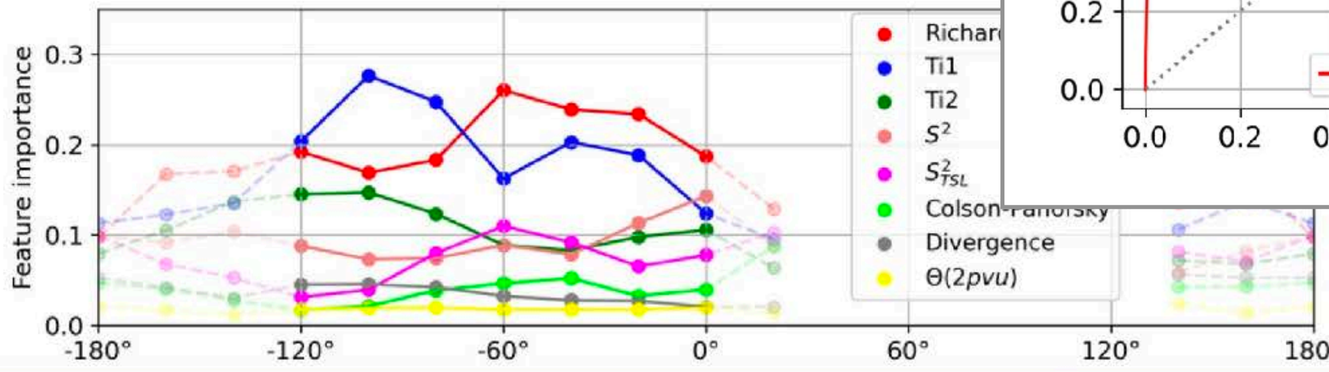
Outlook

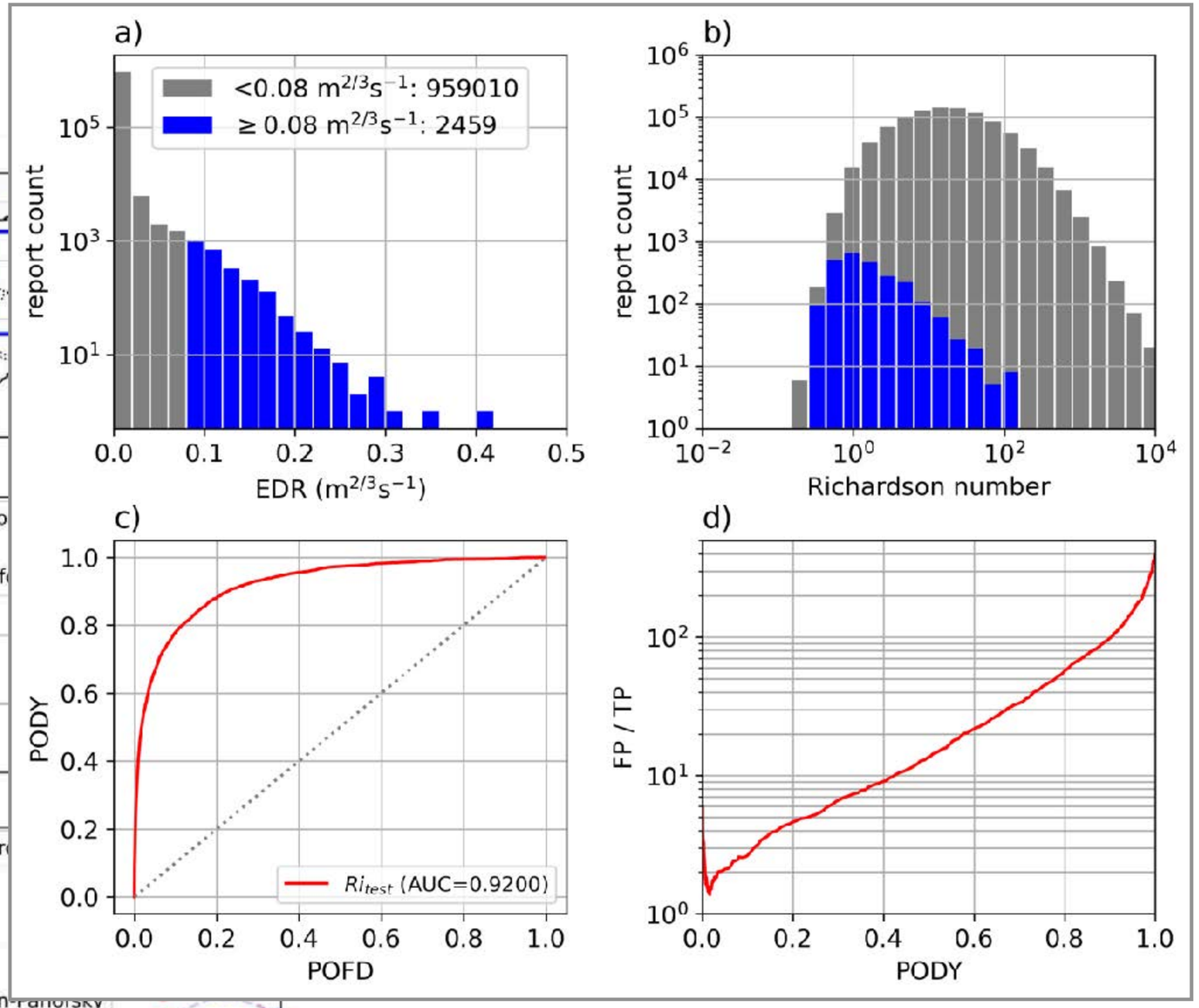

Lower stratosphere
80°N
60°N
40°N
20°N
180°
120°W
60°W
0°
60°E
120°E
180°
10^6
10^5
10^4
10^3
10^2
EDR reports
AUC (EDR≥0.08 m$^{2/3}$s^{-1})
0.95
0.90
0.85
0.80
0.75
-180°
-120°
-60°
0°
60°
120°
180°
Richardson number
Ti1
Random forest
Feature importance
0.3
0.2
0.1
0.0
Richardson
Ti1
Ti2
S^2
S^2_{TSL}
Colson-Panofsky
Divergence
Θ(2pvu)

Outlook

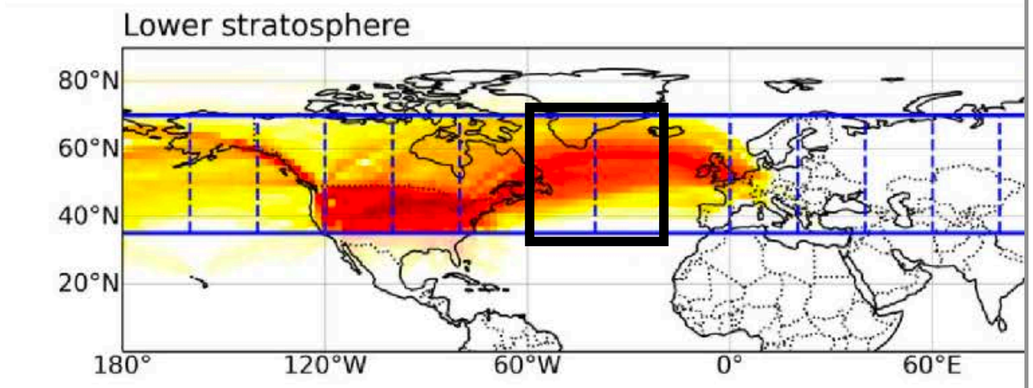

Lower stratosphere
80°N
60°N
40°N
20°N
180°
120°W
60°W
0°
60°E

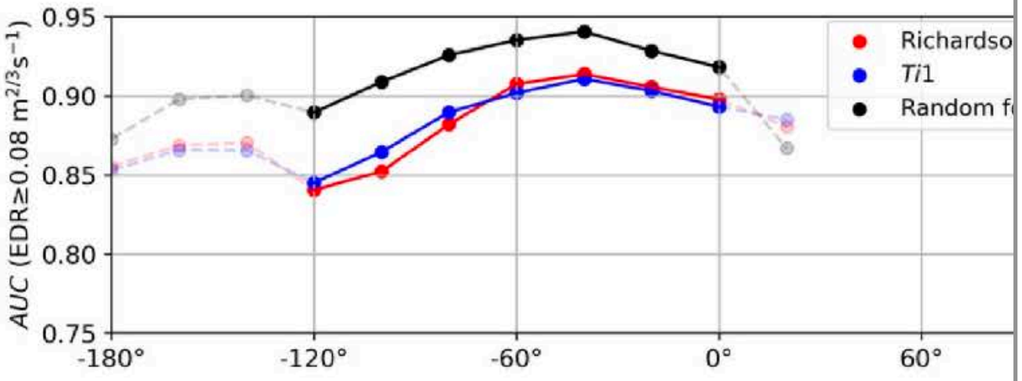

AUC (EDR≥0.08 m^2/3s^-1)
0.95
0.90
0.85
0.80
0.75
-180°
-120°
-60°
0°
60°
Richardso
Ti1
Random f

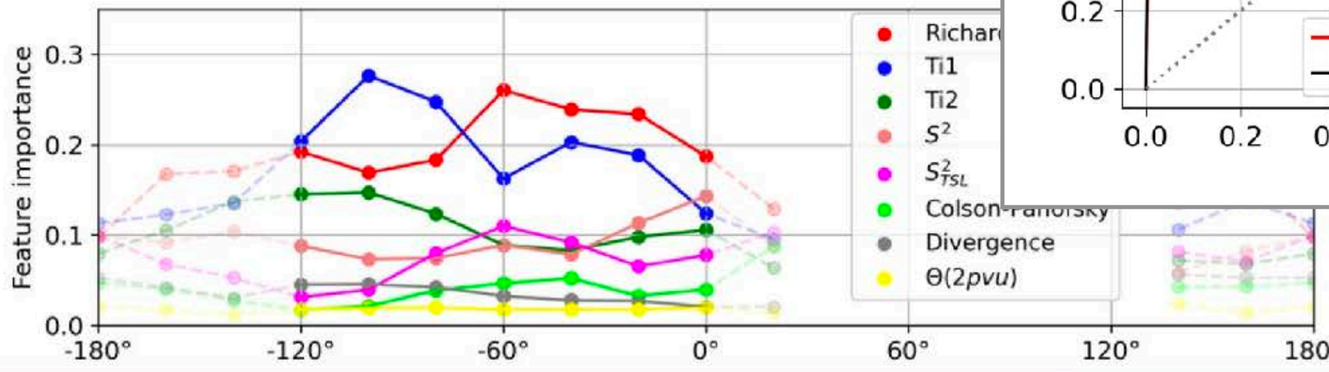

Feature importance
0.3
0.2
0.1
0.0
-180°
-120°
-60°
0°
60°
120°
180°
Richar
Ti1
Ti2
S^2
S^2_fsL
Colson-Panofsky
Divergence
Θ(2pvu)

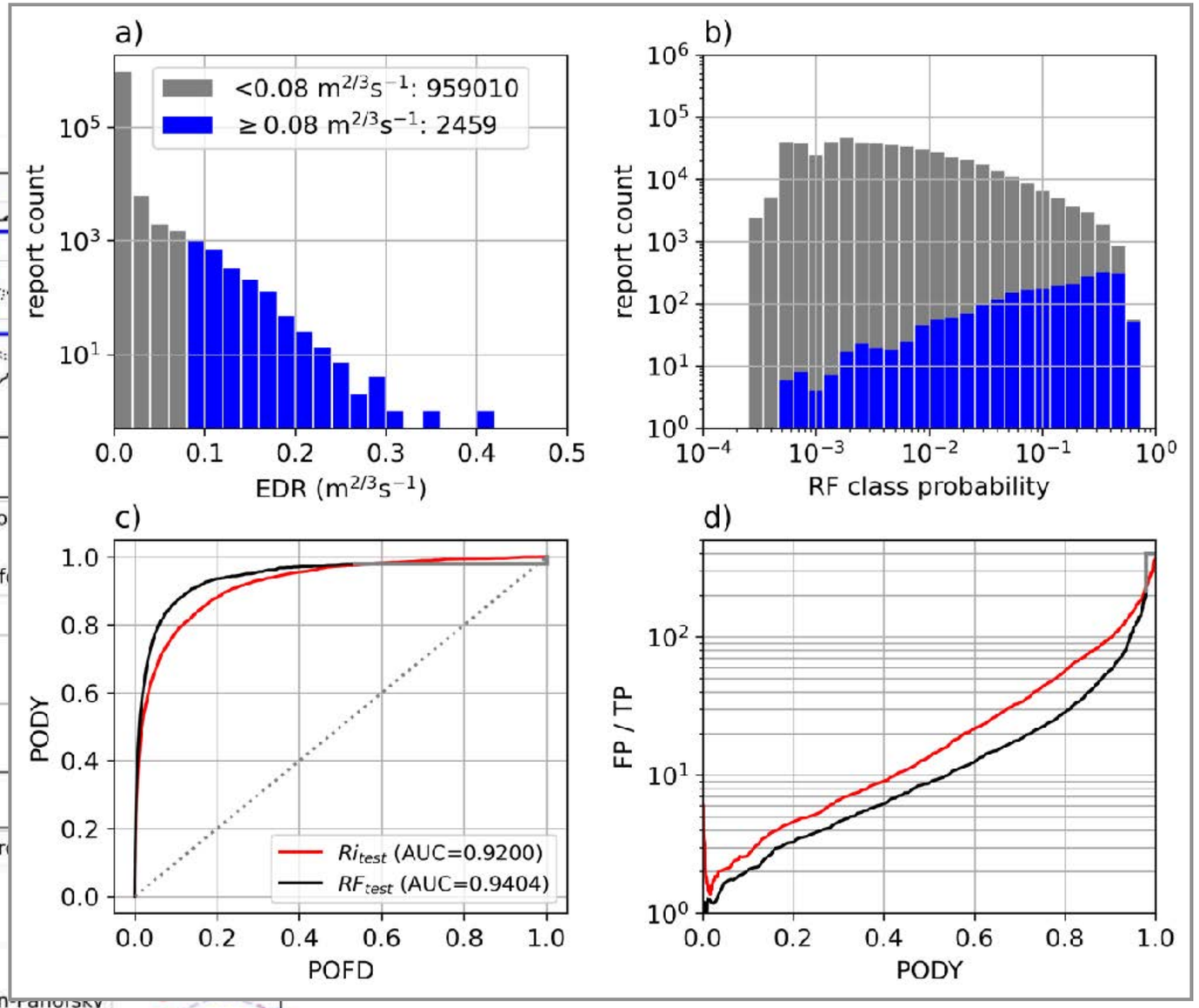

a)
<0.08 m^2/3s^-1: 959010
≥ 0.08 m^2/3s^-1: 2459
report count
EDR (m^2/3s^-1)
b)
report count
Richardson number
c)
PODY
POFD
Ri_test (AUC=0.9200)
d)
FP / TP
PODY

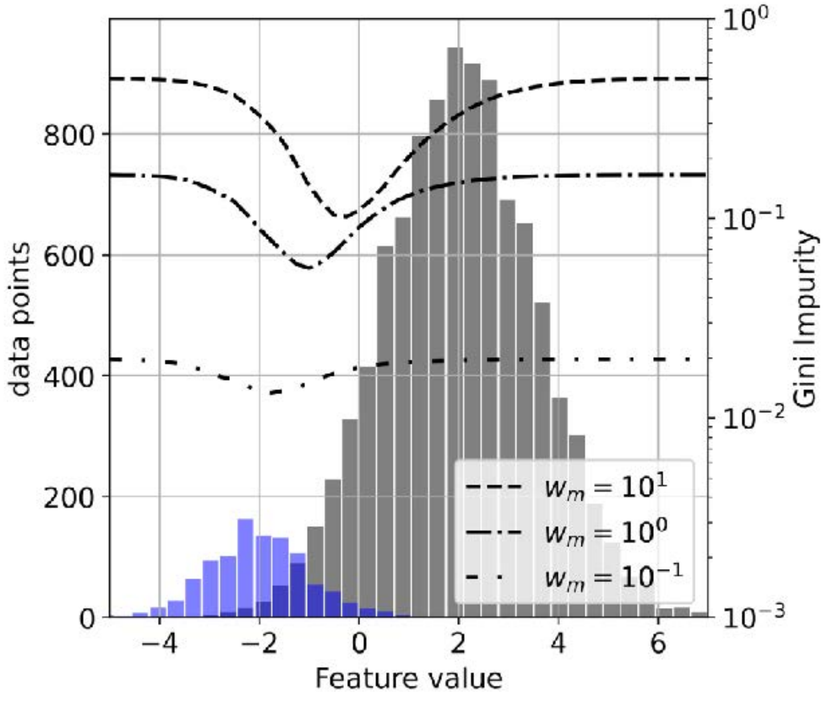
Outlook

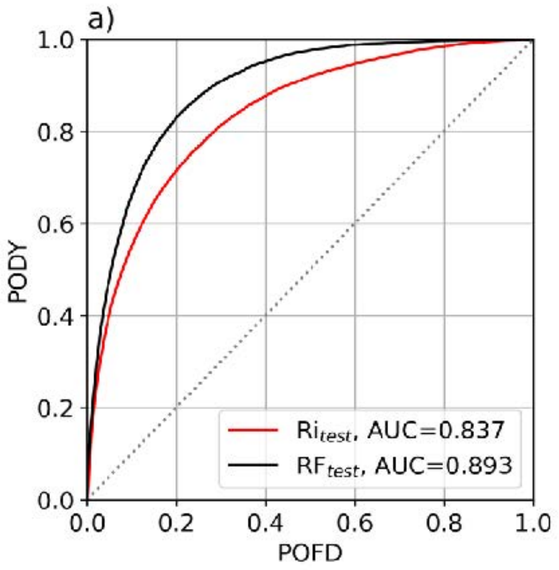

Lower stratosphere
80°N
60°N
40°N
20°N
180°
120°W
60°W
0°
60°E

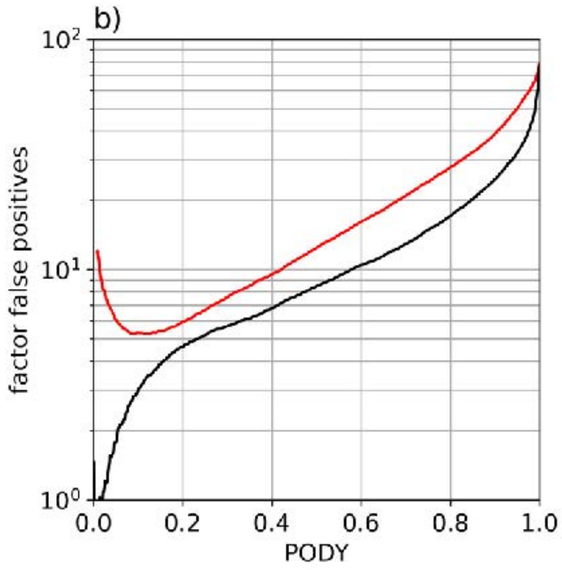

AUC (EDR≥0.08 m^2/3s^-1)
0.95
0.90
0.85
0.80
0.75
-180°
-120°
-60°
0°
60°
Richardso
Ti1
Random f

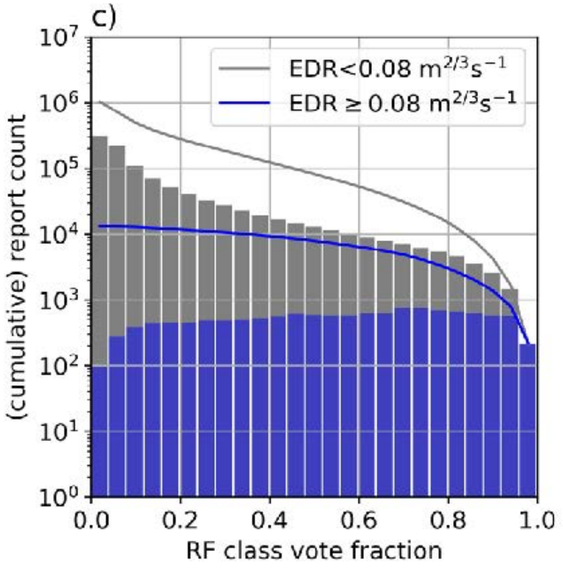

Feature importance
0.3
0.2
0.1
0.0
-180°
-120°
-60°
0°
60°
120°
180°
Richar
Ti1
Ti2
S^2
S^2_fsL
Colson-Panofsky
Divergence
Θ(2pvu)


a)
<0.08 m^2/3s^-1: 959010
≥ 0.08 m^2/3s^-1: 2459
report count
EDR (m^2/3s^-1)
b)
report count
RF class probability
c)
POD Y
POFD
Ri_test (AUC=0.9200)
RF_test (AUC=0.9404)
d)
FP / TP
PODY





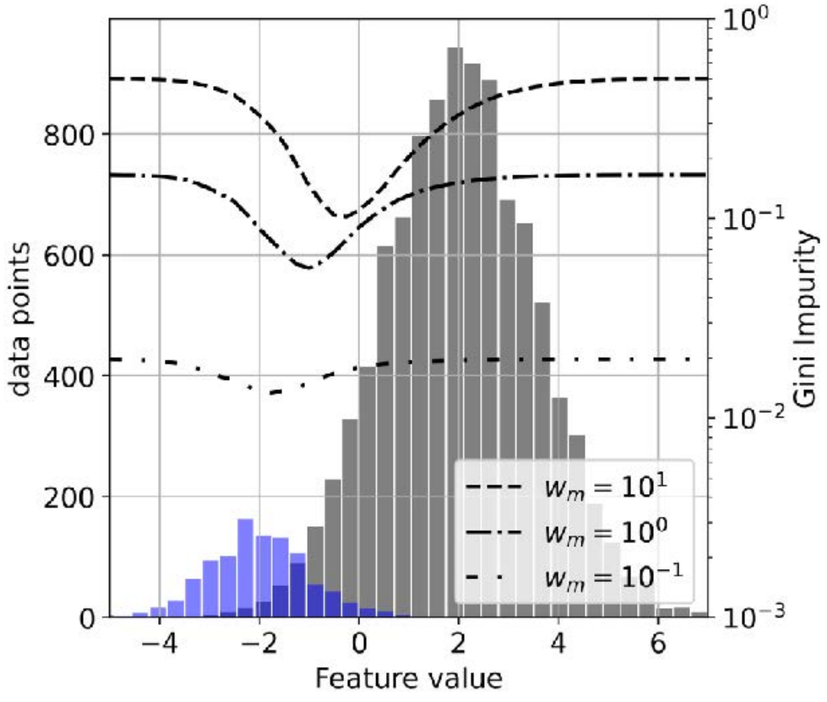


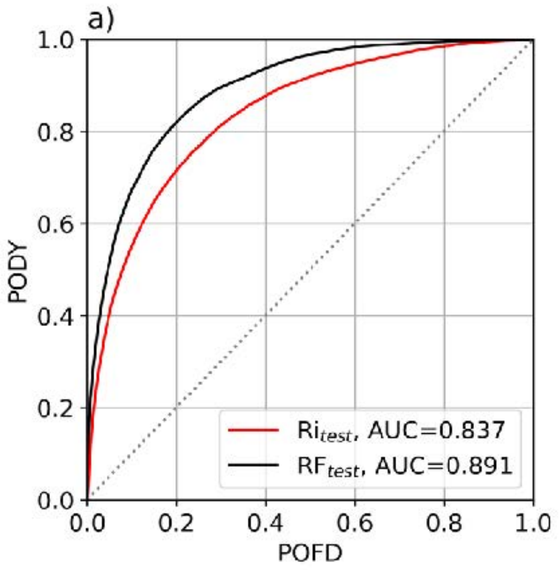


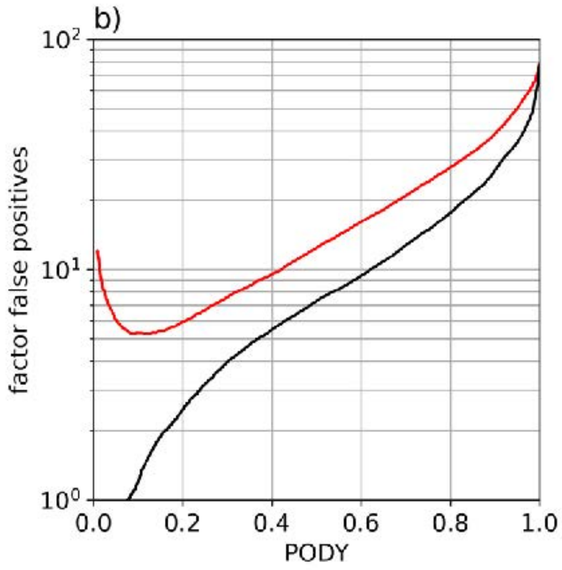
Region:
35°N-70°N, 40°W ± 20°
Upper troposphere

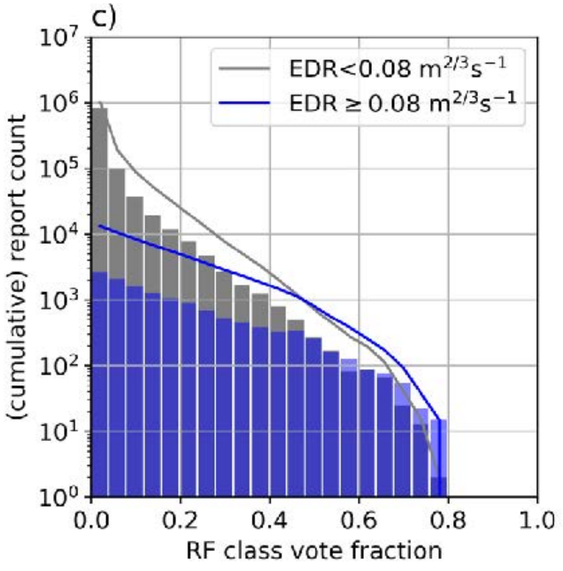
Reports:
EDR<0.08 m$^{2/3}$s^{-1}: 2670893
EDR ≥ 0.08 m$^{2/3}$s^{-1}: 34204

RF settings:
trees=100
max_features=13
max_depth=9
$w_m = 100$





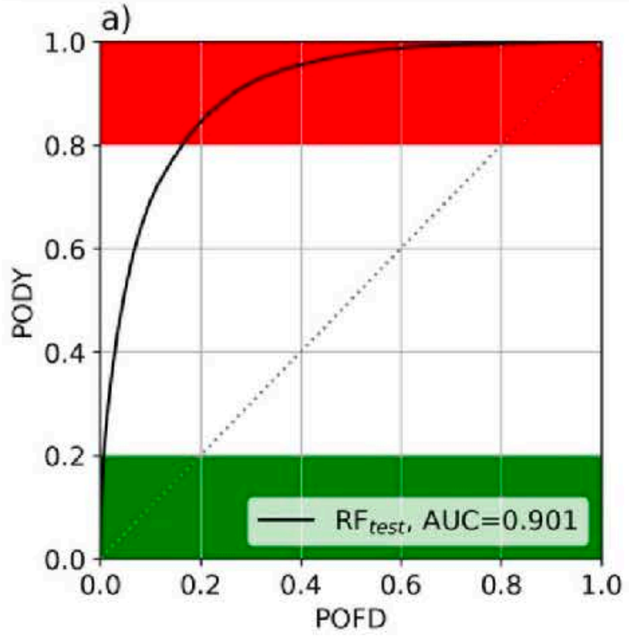


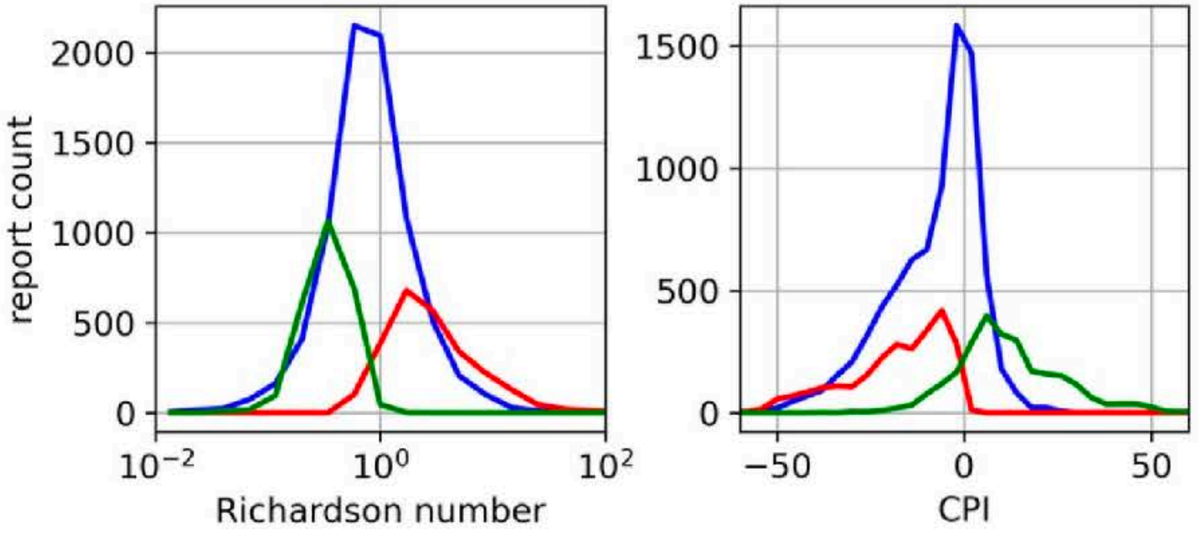


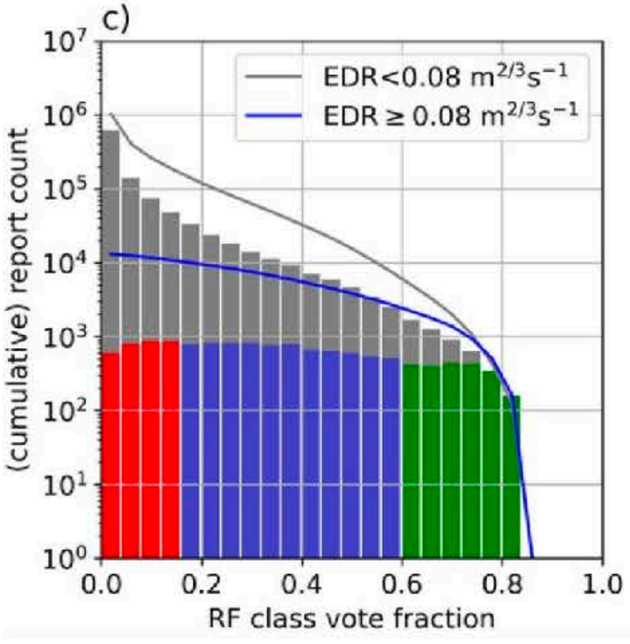
Region:
35°N-70°N, 40°W ± 20°
Upper troposphere

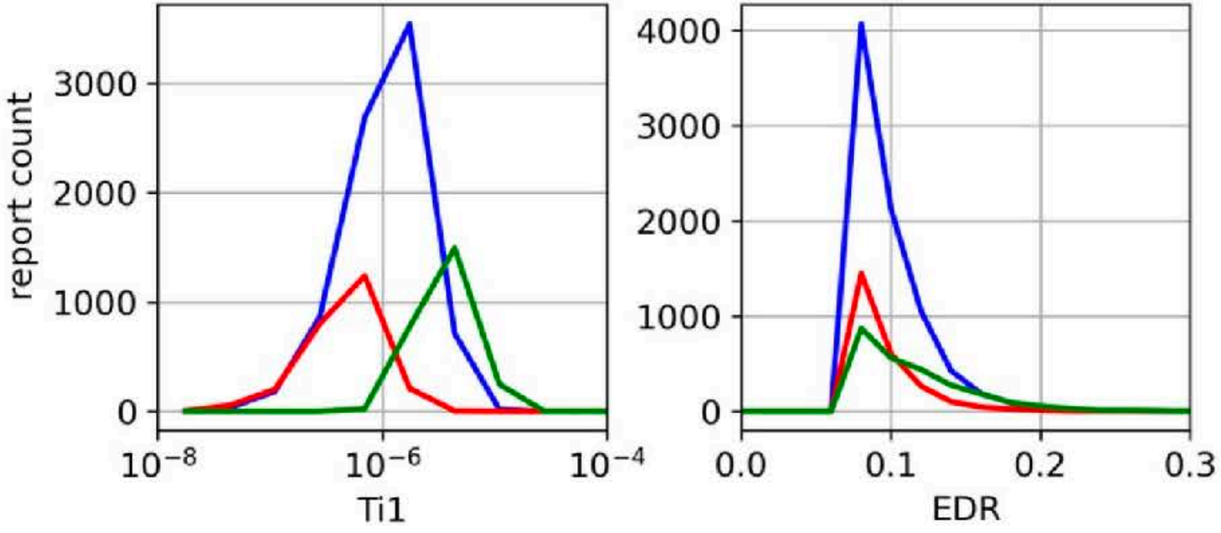
Reports:
EDR<0.08 m$^{2/3}$s^{-1}: 2670893
EDR ≥ 0.08 m$^{2/3}$s^{-1}: 34204

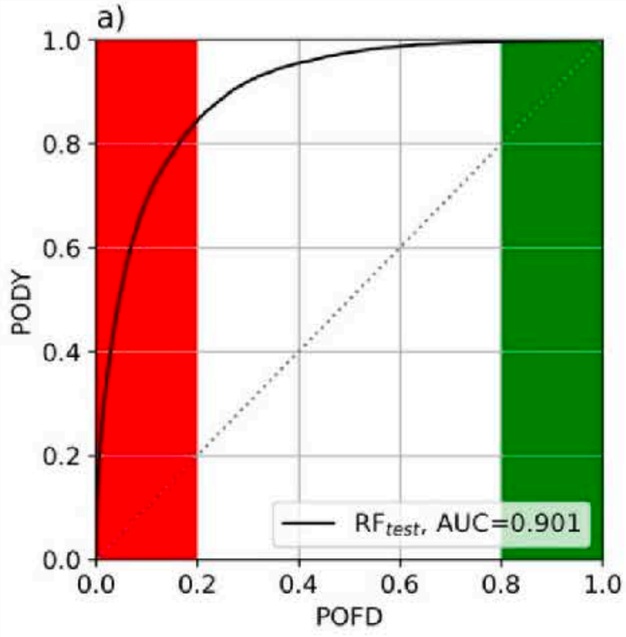
RF settings:
trees=100
max_features=13
max_depth=7
$w_m = 3$

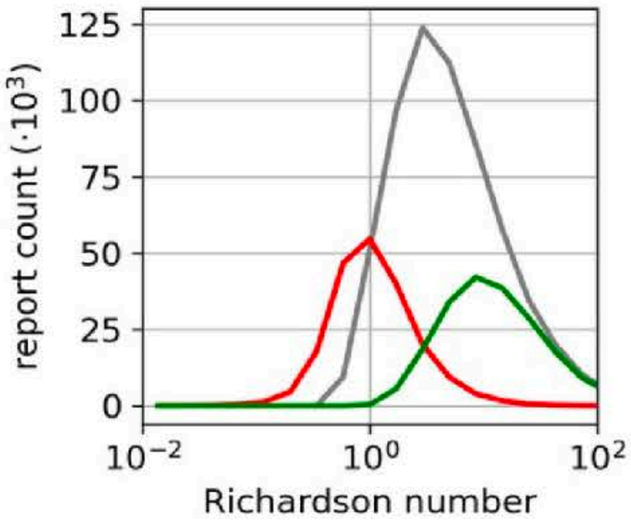

a)
1.0
0.8
0.6
0.4
0.2
0.0
POD Y
0.0
0.2
0.4
0.6
0.8
1.0
POFD
RF$_{test}$, AUC=0.901

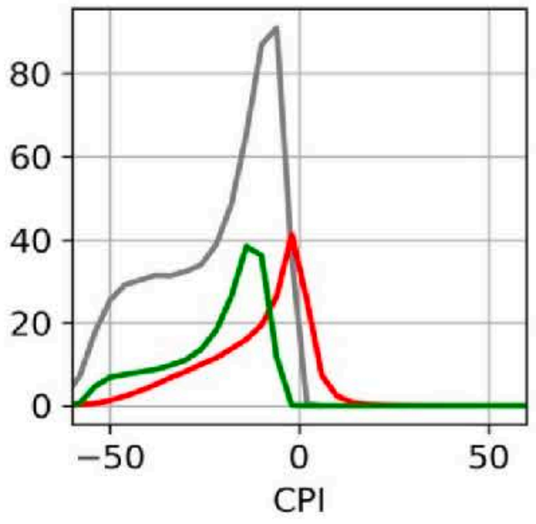

report count
2000
1500
1000
500
0
10^{-2}
10^{0}
10^{2}
Richardson number
1500
1000
500
0
−50
0
50
CPI

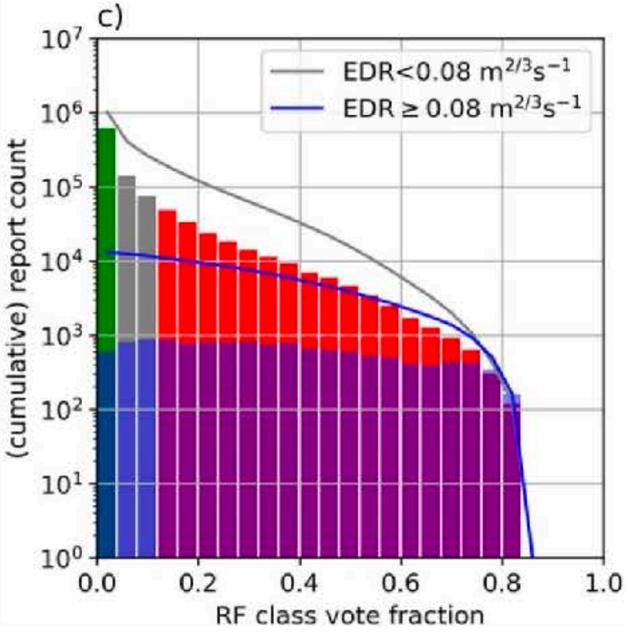

c)
10^{7}
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
(cumulative) report count
EDR<0.08 m$^{2/3}$s^{-1}
EDR ≥ 0.08 m$^{2/3}$s^{-1}
0.0
0.2
0.4
0.6
0.8
1.0
RF class vote fraction

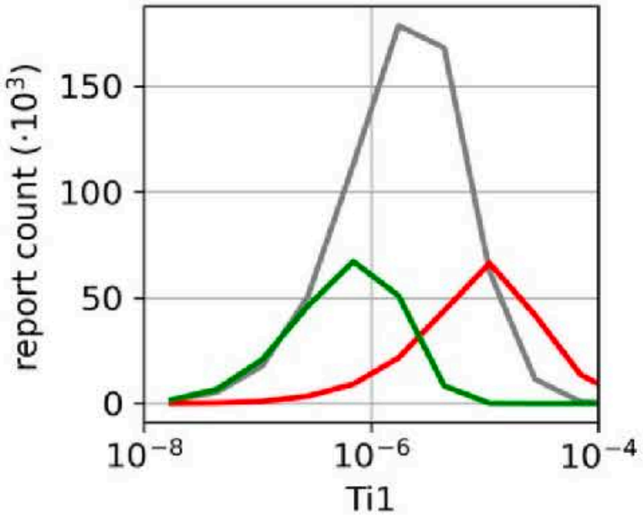

report count
3000
2000
1000
0
10^{-8}
10^{-6}
10^{-4}
Ti1
4000
3000
2000
1000
0
0.0
0.1
0.2
0.3
EDR

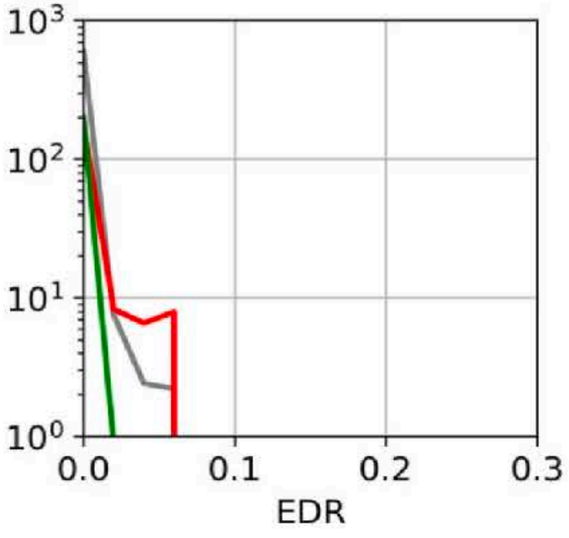

a)
1.0
0.8
0.6
0.4
0.2
0.0
PODY
0.0
0.2
0.4
0.6
0.8
1.0
POFD
RF$_{test}$, AUC=0.901


125
100
75
50
25
0
report count (·10^3)
10^{-2}
10^0
10^2
Richardson number

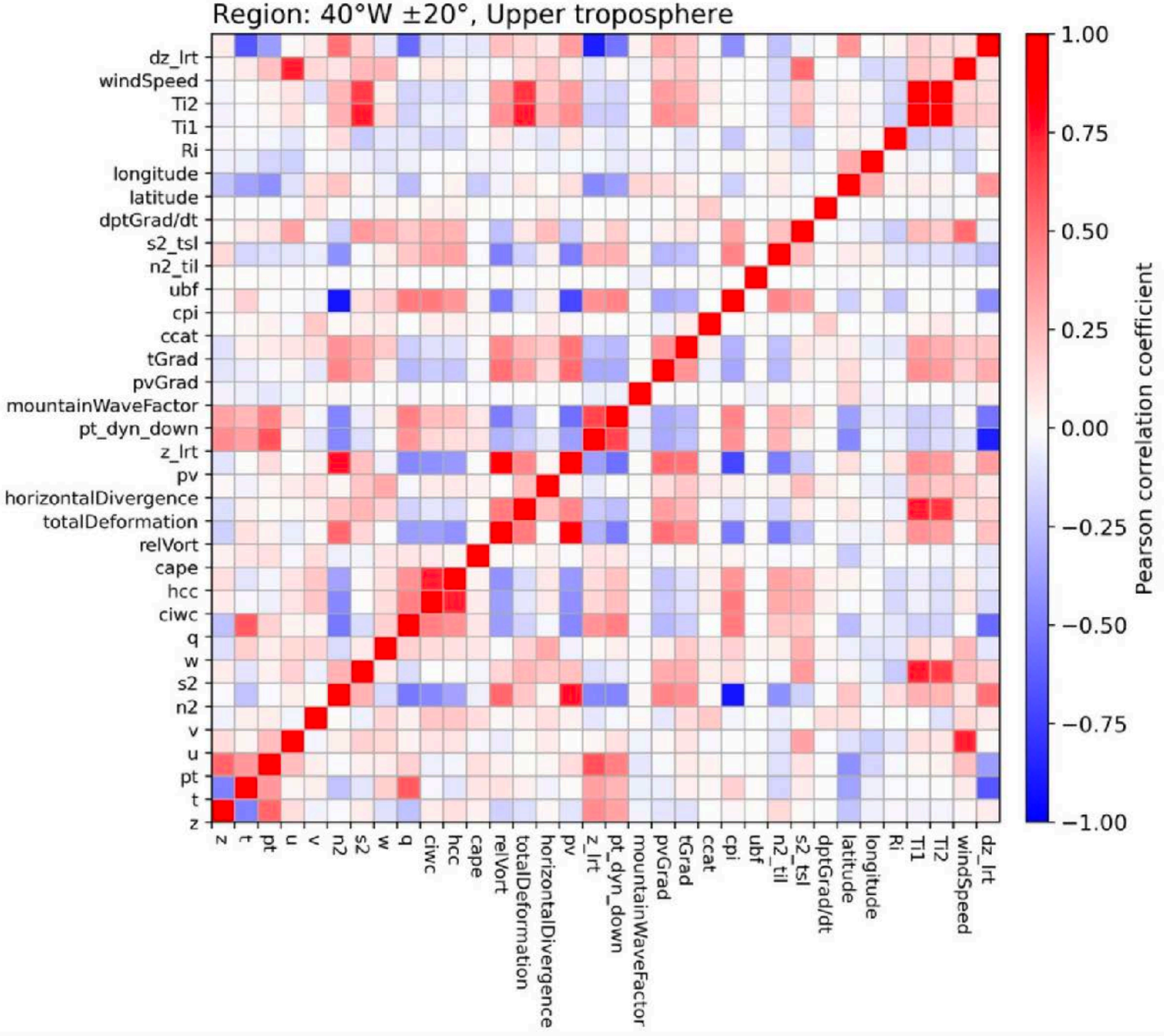

80
60
40
20
0
−50
0
50
CPI

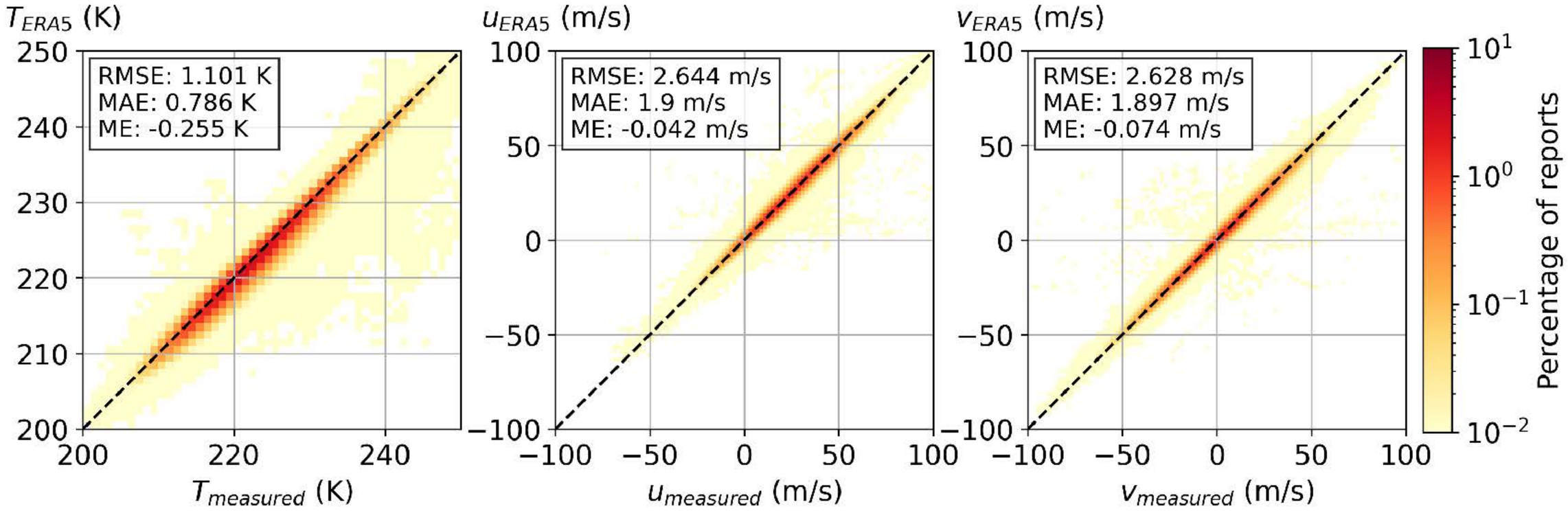

c)
EDR<0.08 m$^{2/3}$s^{-1}
EDR ≥ 0.08 m$^{2/3}$s^{-1}
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
(cumulative) report count
0.0
0.2
0.4
0.6
0.8
1.0
RF class vote fraction

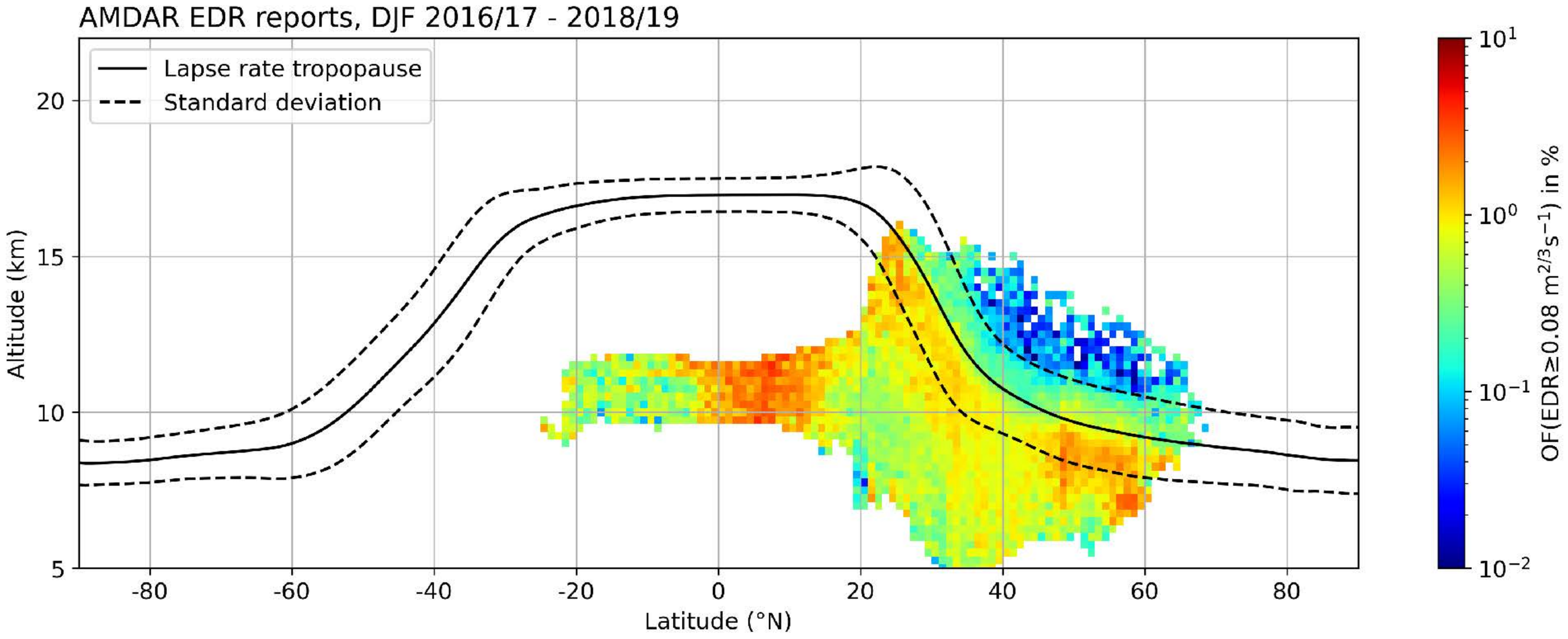

150
100
50
0
report count (·10^3)
10^{-8}
10^{-6}
10^{-4}
Ti1

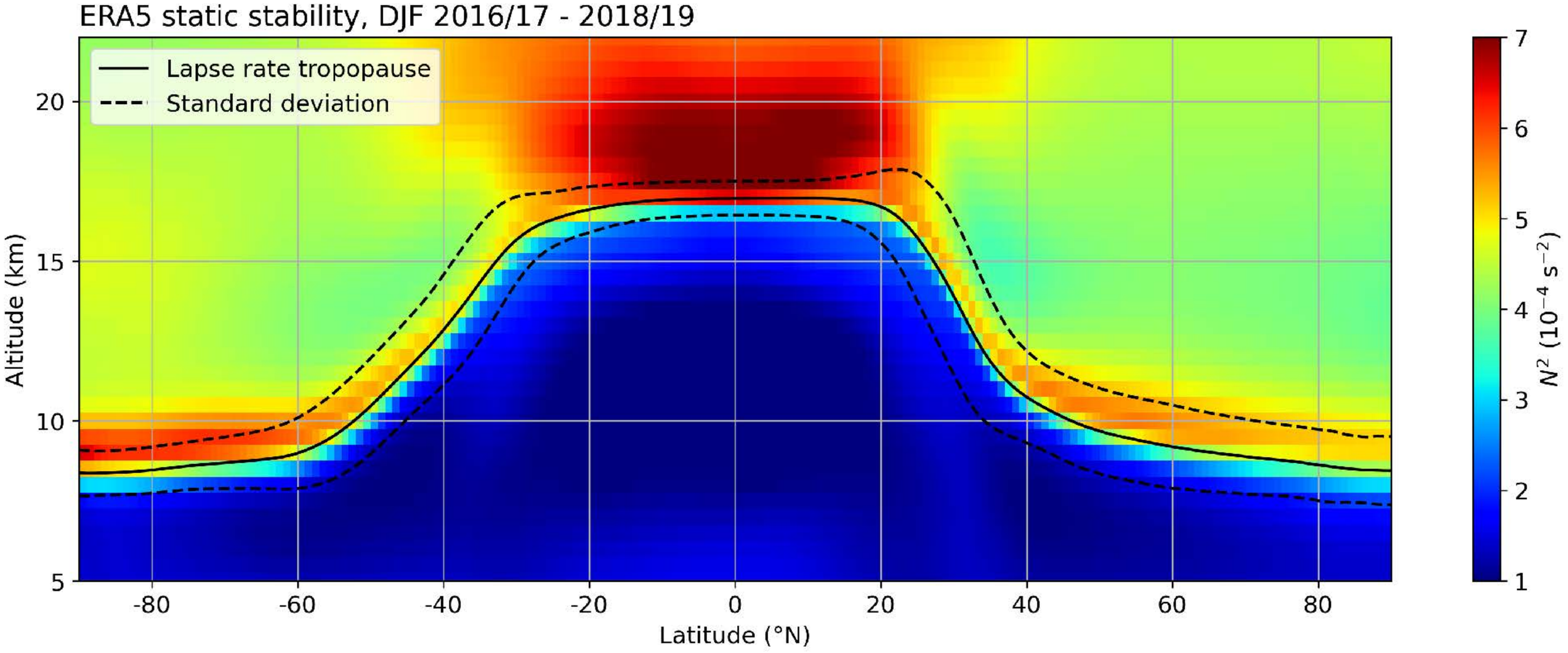

10^3
10^2
10^1
10^0
0.0
0.1
0.2
0.3
EDR

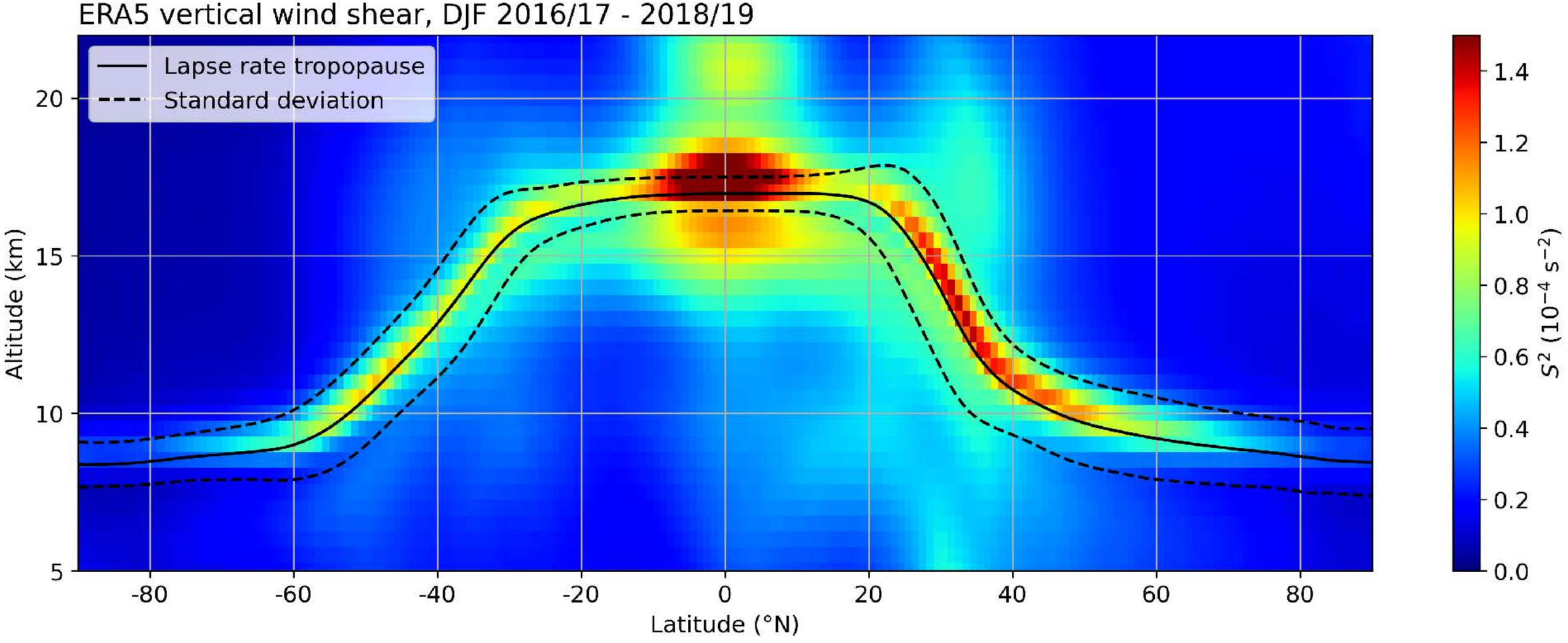
Trainings run ID
1
2
3
4
5
Dec
Jan 2017
Feb
Mar
Apr
May
Sept
Oct
Nov
Dec
Jan 2018
Feb
Mar
Apr
May
Sept
Oct
Nov
Dec
Jan 2019
Feb
Mar
Apr
May
Sept
Oct
Nov
Dec

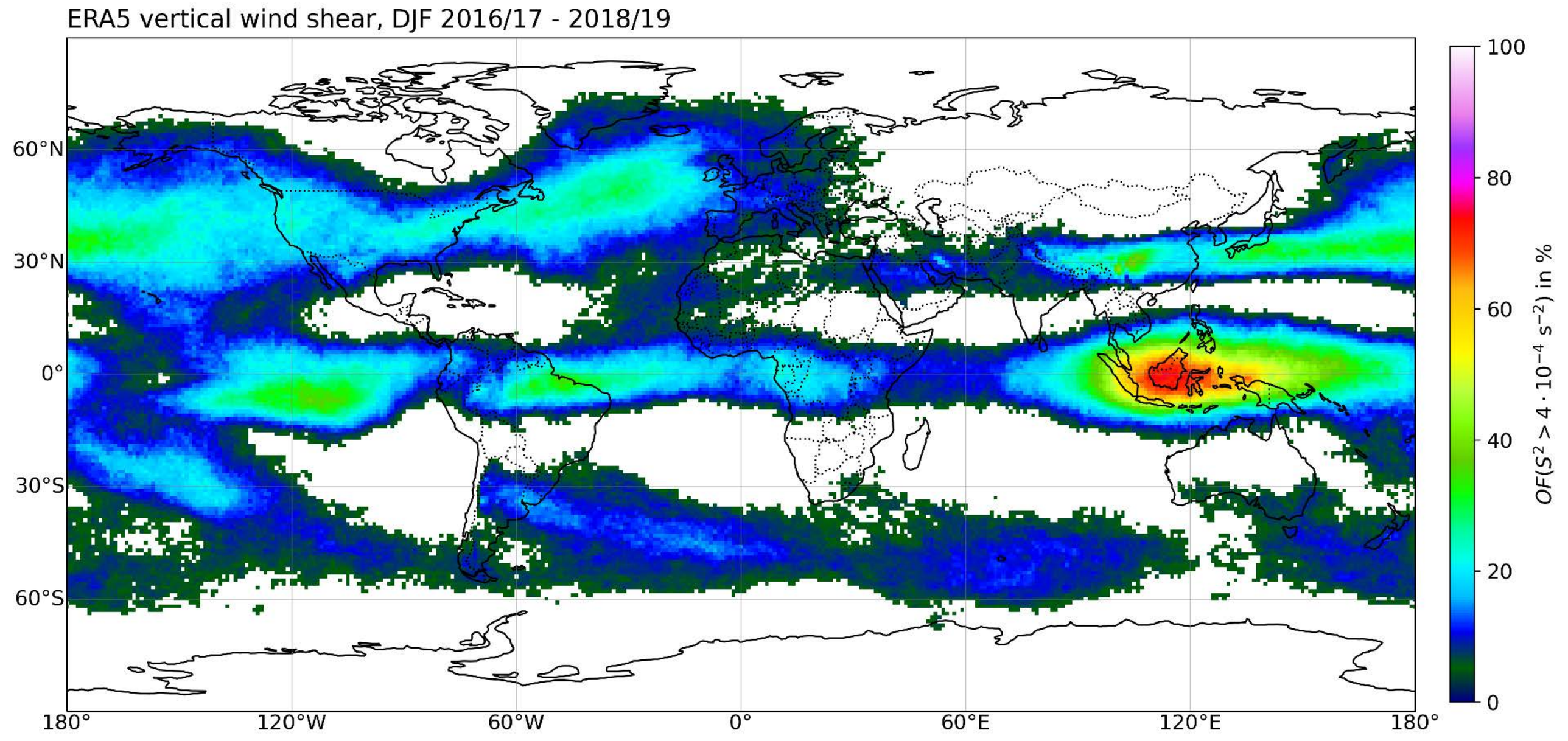

Region: 40°W ±20°, Upper troposphere
dz_lrt
windSpeed
Ti2
Ti1
Ri
longitude
latitude
dptGrad/dt
s2_tsl
n2_til
ubf
cpi
ccat
tGrad
pvGrad
mountainWaveFactor
pt_dyn_down
z_lrt
pv
horizontalDivergence
totalDeformation
relVort
cape
hcc
ciwc
q
w
s2
n2
v
u
pt
t
z
Pearson correlation coefficient
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

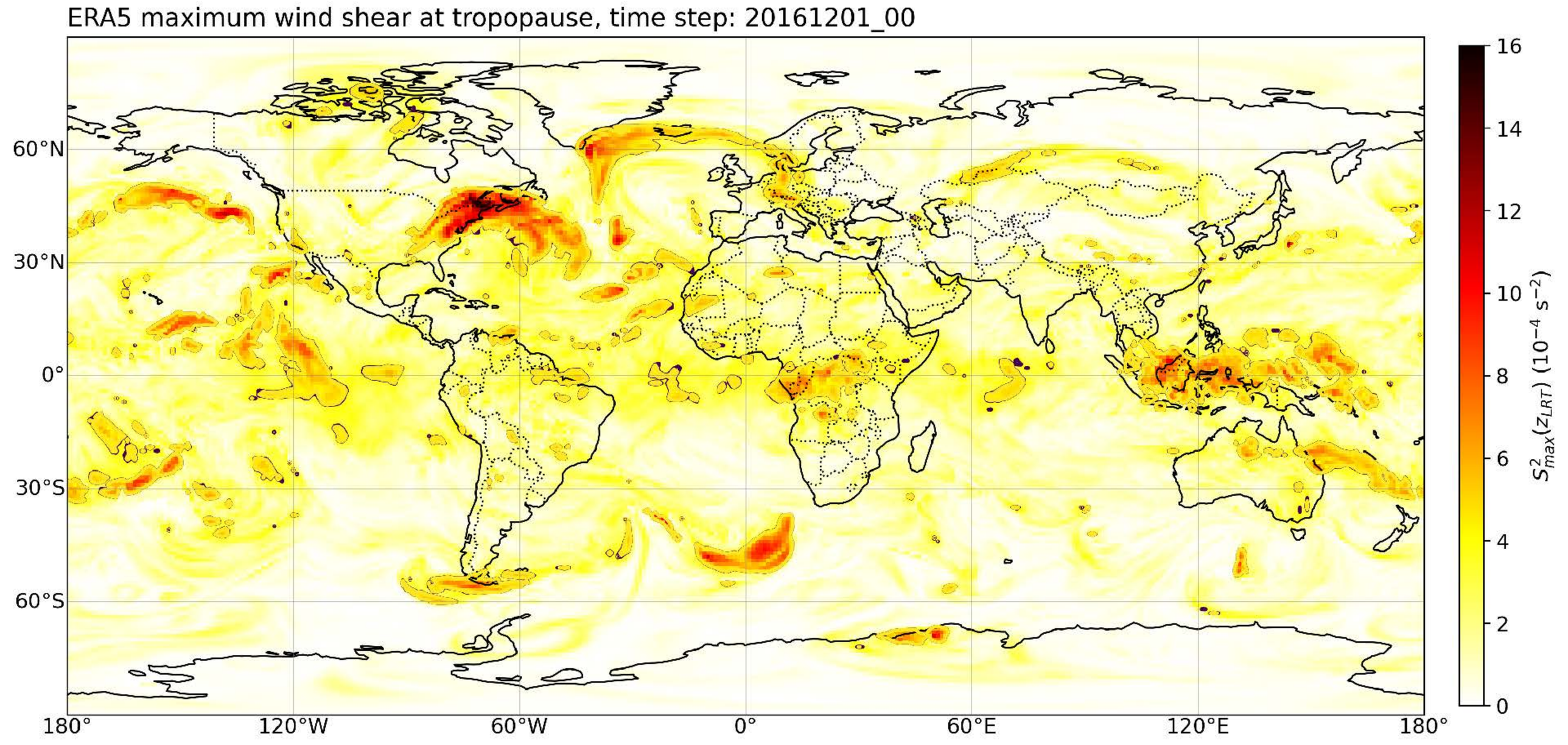

T_{ERA5} (K)
RMSE: 1.101 K
MAE: 0.786 K
ME: -0.255 K
250
240
230
220
210
200
200
220
240
$T_{measured}$ (K)
u_{ERA5} (m/s)
RMSE: 2.644 m/s
MAE: 1.9 m/s
ME: -0.042 m/s
100
50
0
−50
−100
−100
−50
0
50
100
$u_{measured}$ (m/s)
v_{ERA5} (m/s)
RMSE: 2.628 m/s
MAE: 1.897 m/s
ME: -0.074 m/s
100
50
0
−50
−100
−100
−50
0
50
100
$v_{measured}$ (m/s)
10^{1}
10^{0}
10^{-1}
10^{-2}
Percentage of reports

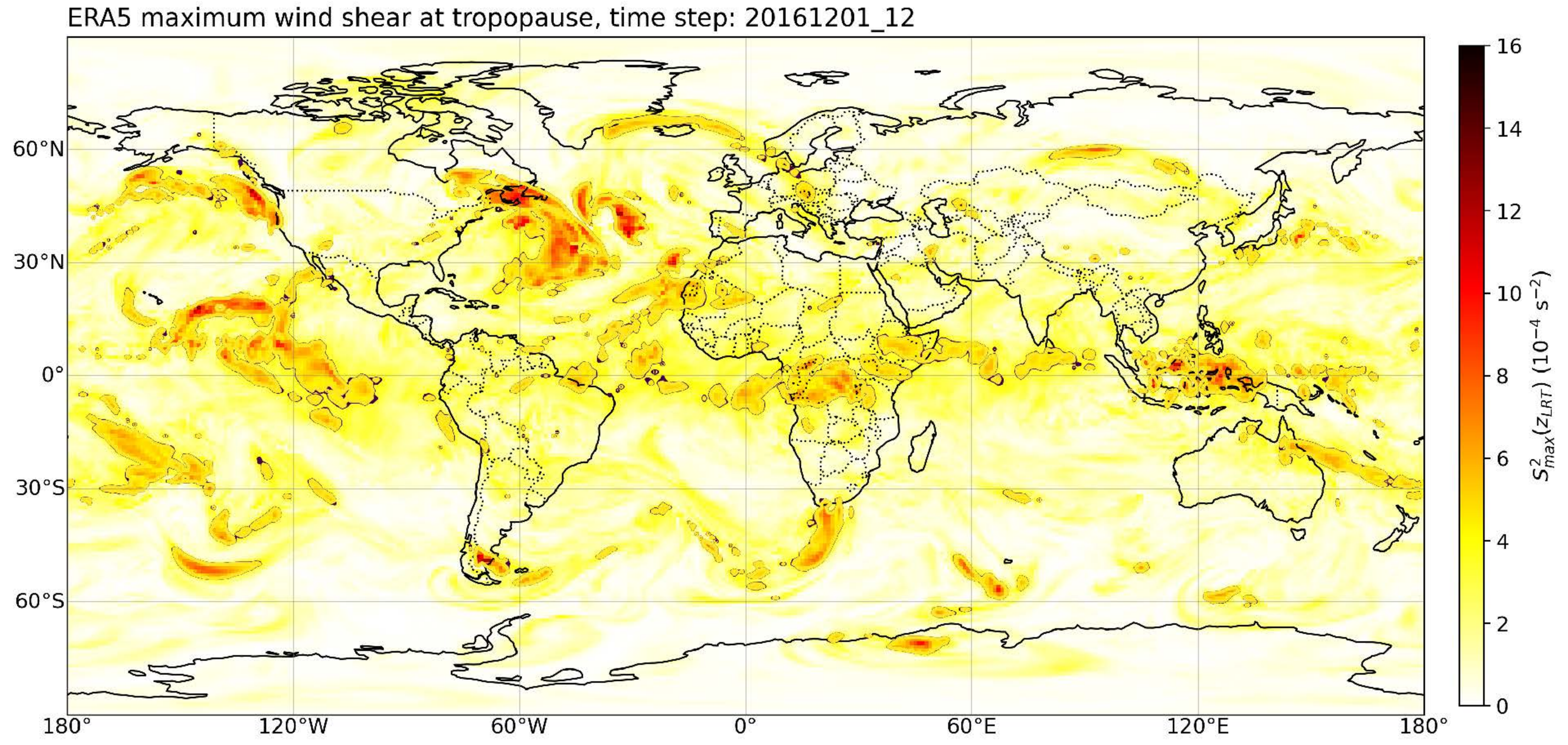

AMDAR EDR reports, DJF 2016/17 - 2018/19
Lapse rate tropopause
Standard deviation
Altitude (km)
20
15
10
5
-80
-60
-40
-20
0
20
40
60
80
Latitude (°N)
10^{1}
10^{0}
10^{-1}
10^{-2}
OF(EDR≥0.08 $m^{2/3}s^{-1}$) in %

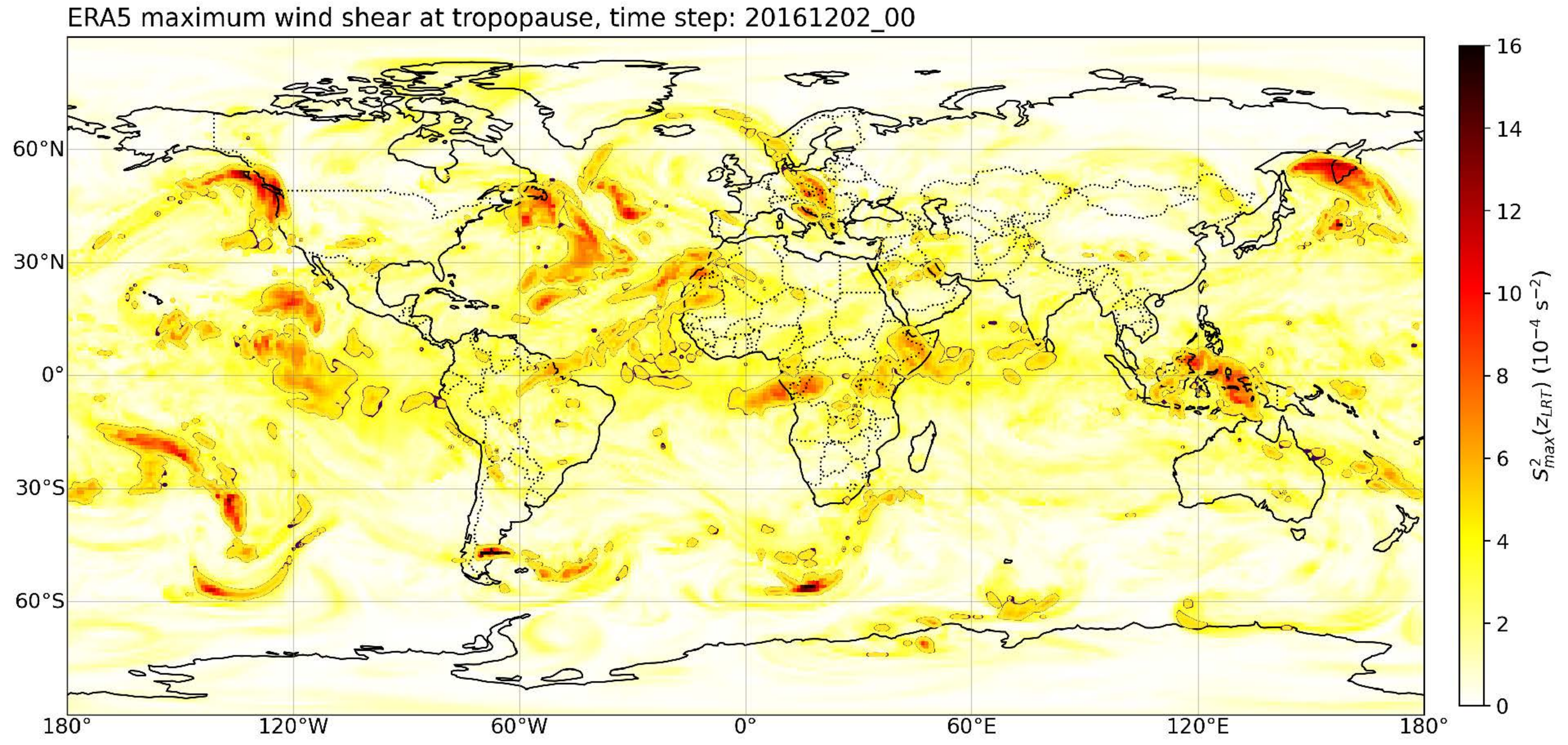

ERA5 static stability, DJF 2016/17 - 2018/19
Lapse rate tropopause
Standard deviation
Altitude (km)
20
15
10
5
-80
-60
-40
-20
0
20
40
60
80
Latitude (°N)
7
6
5
4
3
2
1
N^2 (10^{-4} s^{-2})

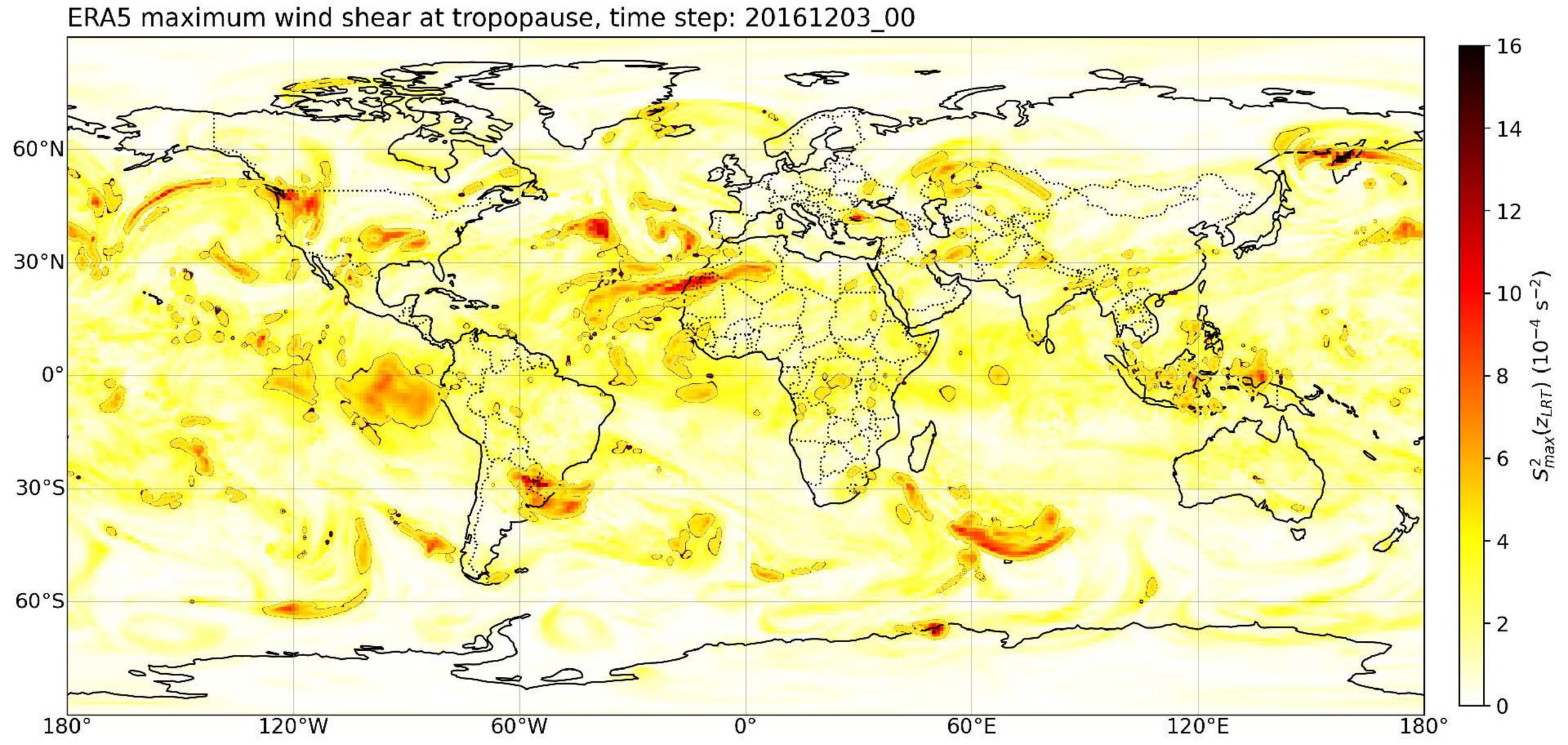

ERA5 vertical wind shear, DJF 2016/17 - 2018/19
Lapse rate tropopause
Standard deviation
Altitude (km)
20
15
10
5
-80
-60
-40
-20
0
20
40
60
80
Latitude (°N)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
S^2 (10^{-4} s^{-2})

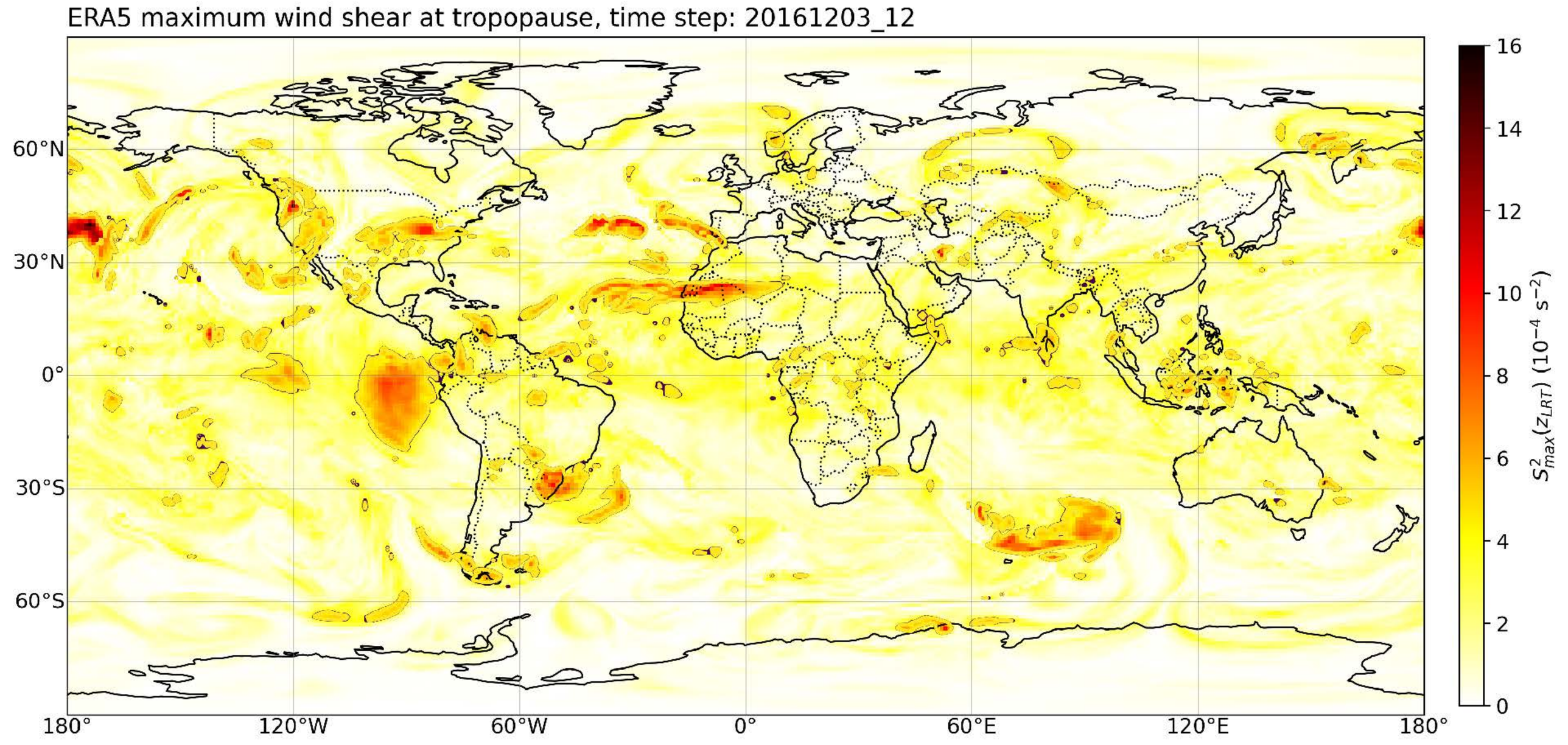

ERA5 vertical wind shear, DJF 2016/17 - 2018/19
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
100
80
60
40
20
0
$OF(S^2 > 4 \cdot 10^{-4}\ s^{-2})$ in %

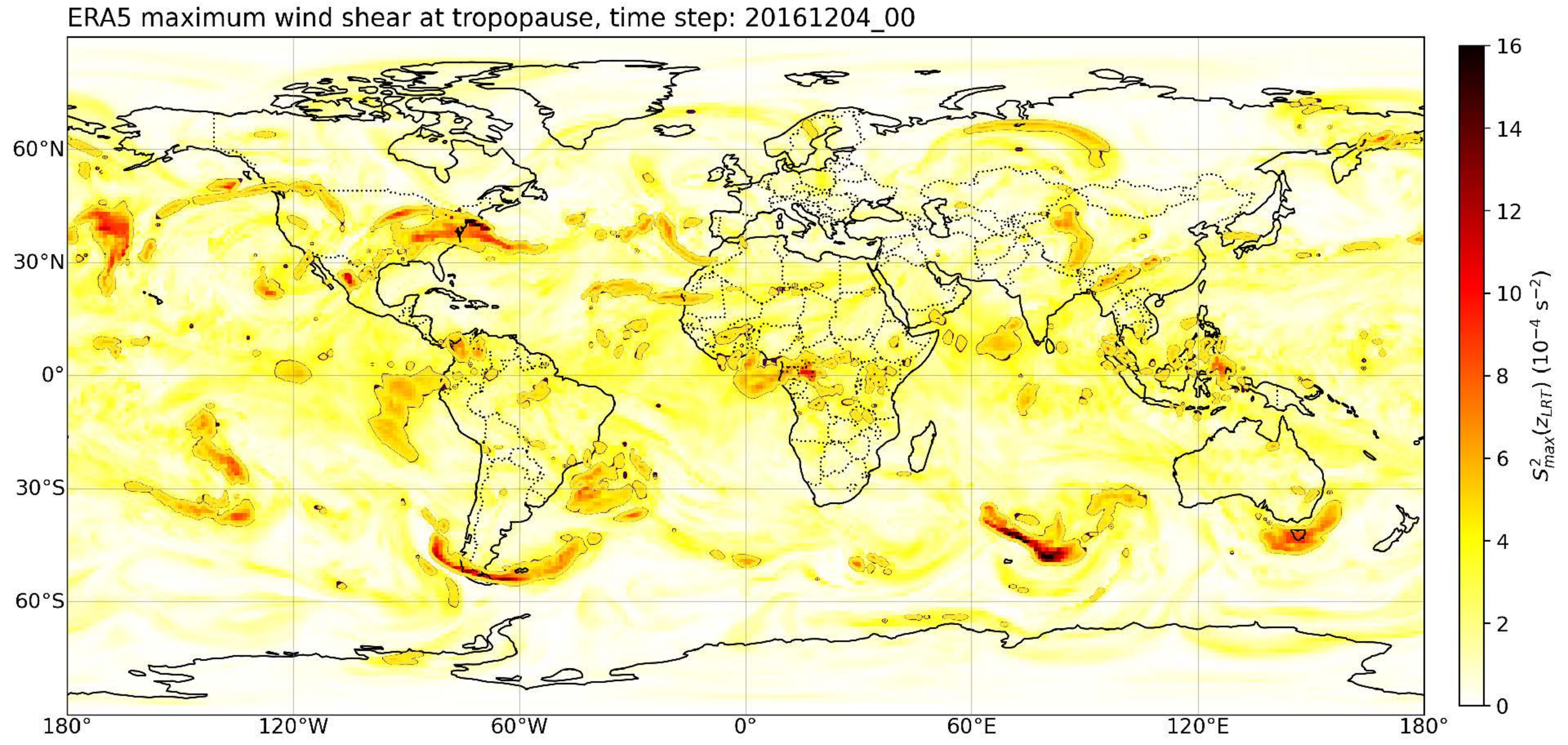

ERA5 maximum wind shear at tropopause, time step: 20161201_00
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
16
14
12
10
8
6
4
2
0
$S^2_{max}(z_{LRT})$ $(10^{-4}\ s^{-2})$

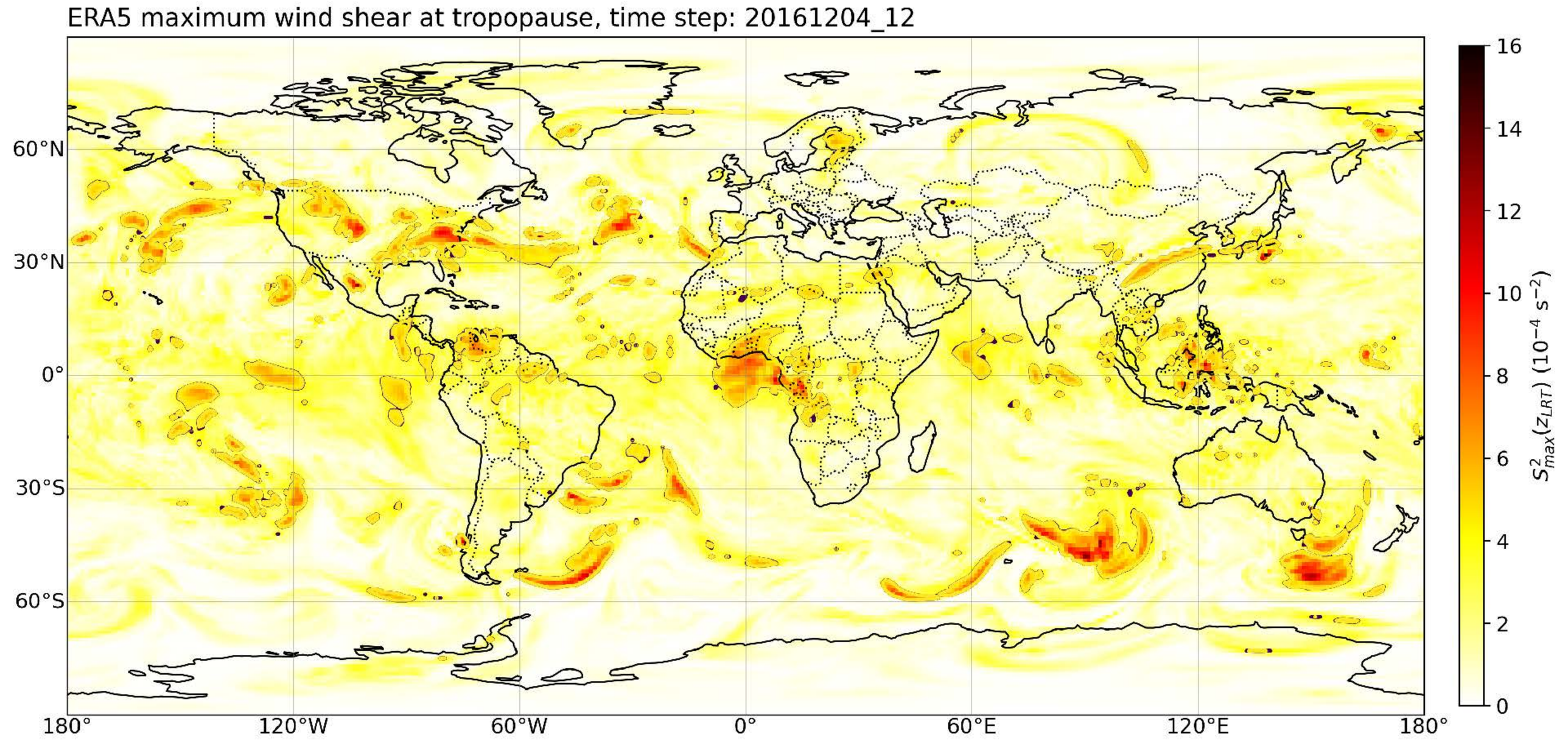

ERA5 maximum wind shear at tropopause, time step: 20161201_12
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
16
14
12
10
8
6
4
2
0
$S^2_{max}(z_{LRT})$ (10^{-4} s^{-2})


ERA5 maximum wind shear at tropopause, time step: 20161202_00
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
16
14
12
10
8
6
4
2
0
$S^2_{max}(z_{LRT})$ $(10^{-4}\ s^{-2})$


ERA5 maximum wind shear at tropopause, time step: 20161203_00
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
16
14
12
10
8
6
4
2
0
$S^2_{max}(z_{LRT})$ $(10^{-4}\ s^{-2})$


ERA5 maximum wind shear at tropopause, time step: 20161203_12
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
16
14
12
10
8
6
4
2
0
$S^2_{max}(z_{LRT})$ (10^{-4} s^{-2})


ERA5 maximum wind shear at tropopause, time step: 20161204_00
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
16
14
12
10
8
6
4
2
0
$S^2_{max}(z_{LRT})$ (10^{-4} s^{-2})


ERA5 maximum wind shear at tropopause, time step: 20161204_12
60°N
30°N
0°
30°S
60°S
180°
120°W
60°W
0°
60°E
120°E
180°
16
14
12
10
8
6
4
2
0
$S^2_{max}(z_{LRT})$ $(10^{-4}\ s^{-2})$