Unlocking the potential of Time Series across different fields of application

Y. Maruccia[1], S. Cavuoti[1], G. Riccio[1], D. De Cicco[2], R. Crupi[3]

[1] INAF - Astronomical Observatory of Capodimonte
[2] University Federico II of Napoli
[3] Banca Intesa San Paolo

Activities carried out within the project *Centro Nazionale HPC, Big Data e Quantum Computing CN_00000013, Spoke 3 "Astrophysics and Cosmos Observations" - CUP C53C22000350006.*

Introduction and Relevance

Introduction

Time series analysis is a powerful tool used to extract insights from data collected over time, and its applications span across a **wide range of fields of application**.

By analysing trends, patterns, and seasonal variations, time series models allow for **accurate forecasting** and **anomaly detection**.

Beyond forecasting, classifying time series data allows us to group **similar patterns and behaviors**, adding even more depth to its utility.

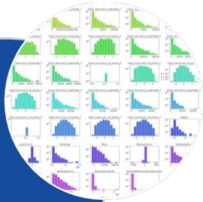
Moreover, the methodologies applied in one area can often be adapted to others, allowing **knowledge sharing and innovation** across different domains.

Aim of this work

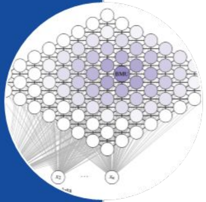
How the same methodologies used to analyse temporal data can be applied across seemingly unrelated fields.

Research Methodology

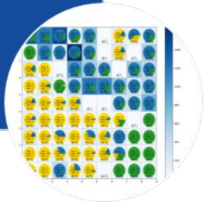
Method



Modeling time series using statistical features



Adopting Self-Organizing Maps



Identifying groups of objects with or without similar characteristics

Application: AGN

Dataset and features

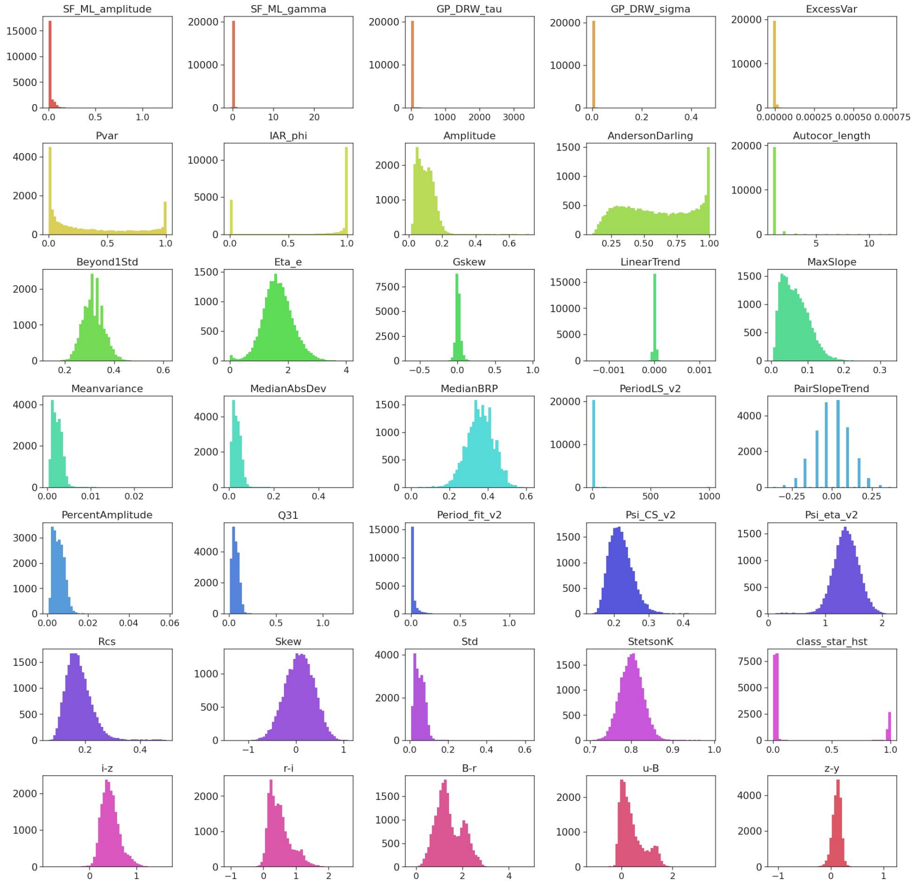
For further details on the dataset: **De Cicco+2021**

Variability features

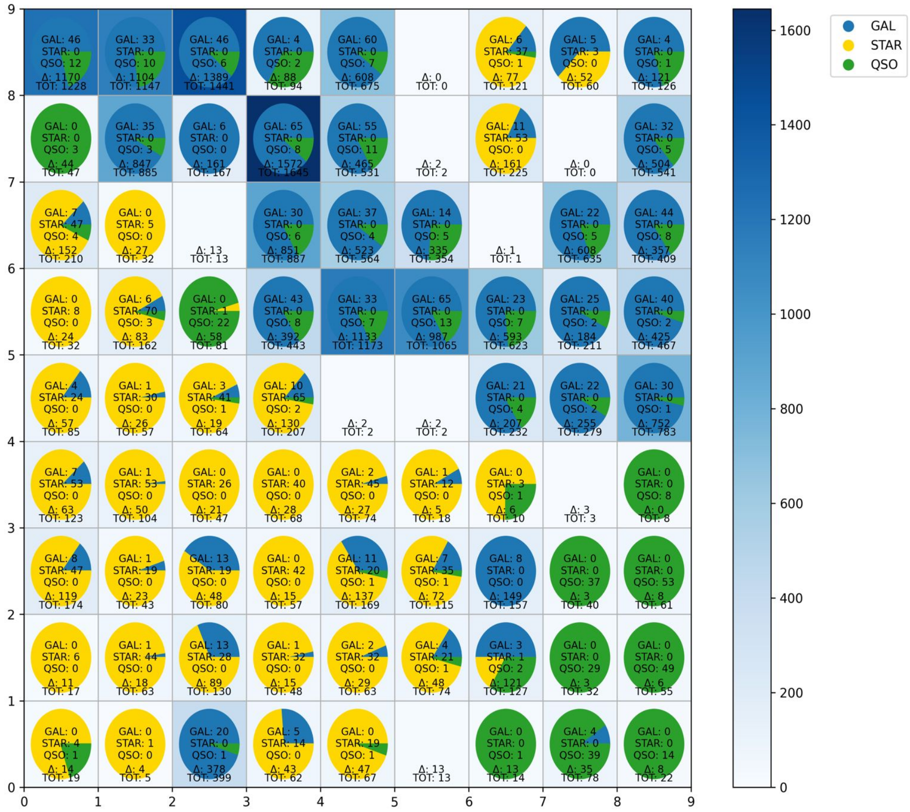
- SF_ML_amplitude
- SF_ML_gamma
- GP_DRW_tau
- GP_DRW_sigma
- ExcessVar
- Pvar
- IAR_phi
- Amplitude
- AndersonDarling
- Autocor_length
- Beyond1Std
- Eta_e
- Gskew
- LinearTrend
- MaxSlope
- Meanvariance
- MedianAbsDev
- MedianBRP
- PeriodLS_v2
- PairSlopeTrend
- PercentAmplitude
- Q31
- Period_fit_v2
- Psi_CS_v2
- Psi_eta_v2
- Rcs
- Skew
- Std
- StetsonK
- class_star_hst

Colors

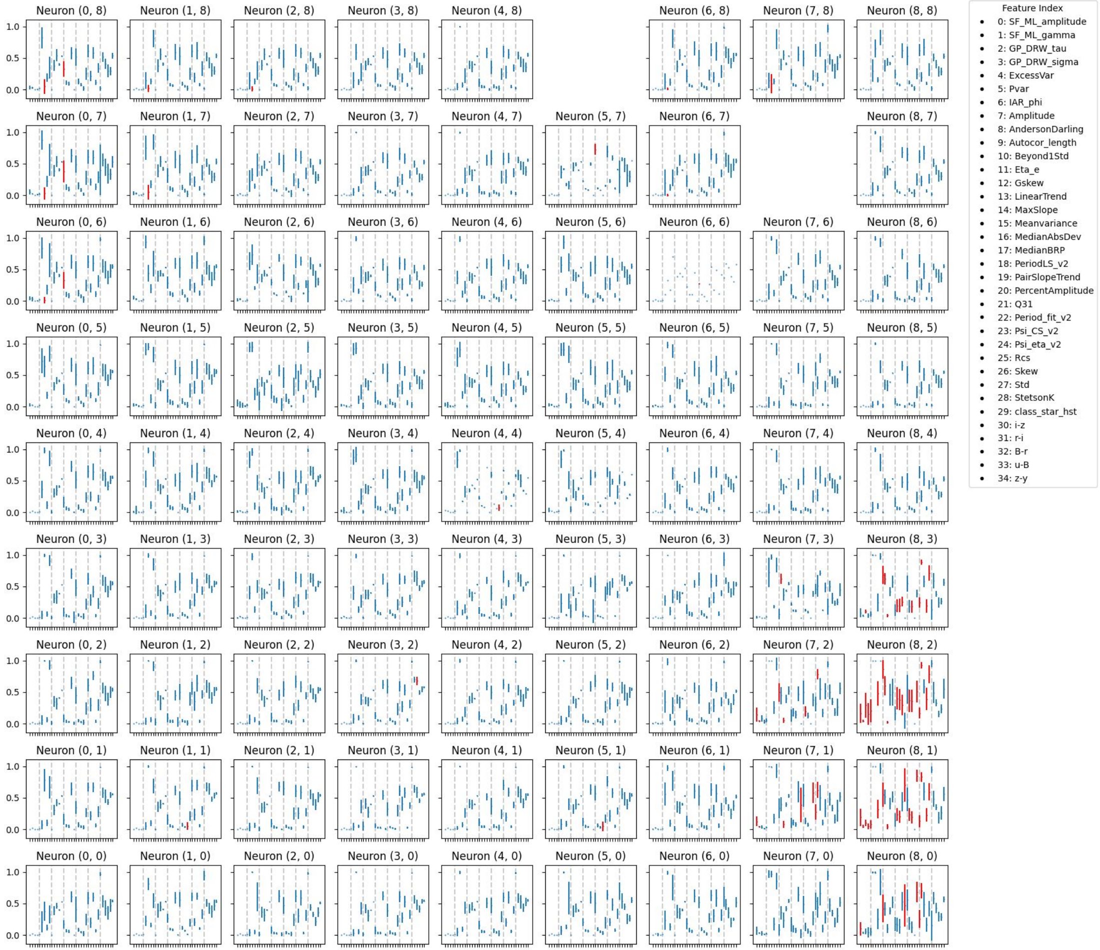
- i-z
- r-i
- B-r
- u-B

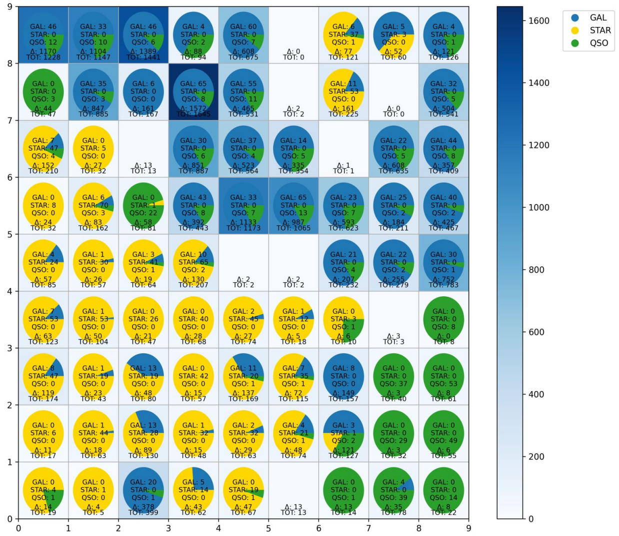


Activation map: All features



All Features





For each neuron, we calculated the global mean and standard deviation. Features are in red if:

mean < global_mean - 2*global_std

or

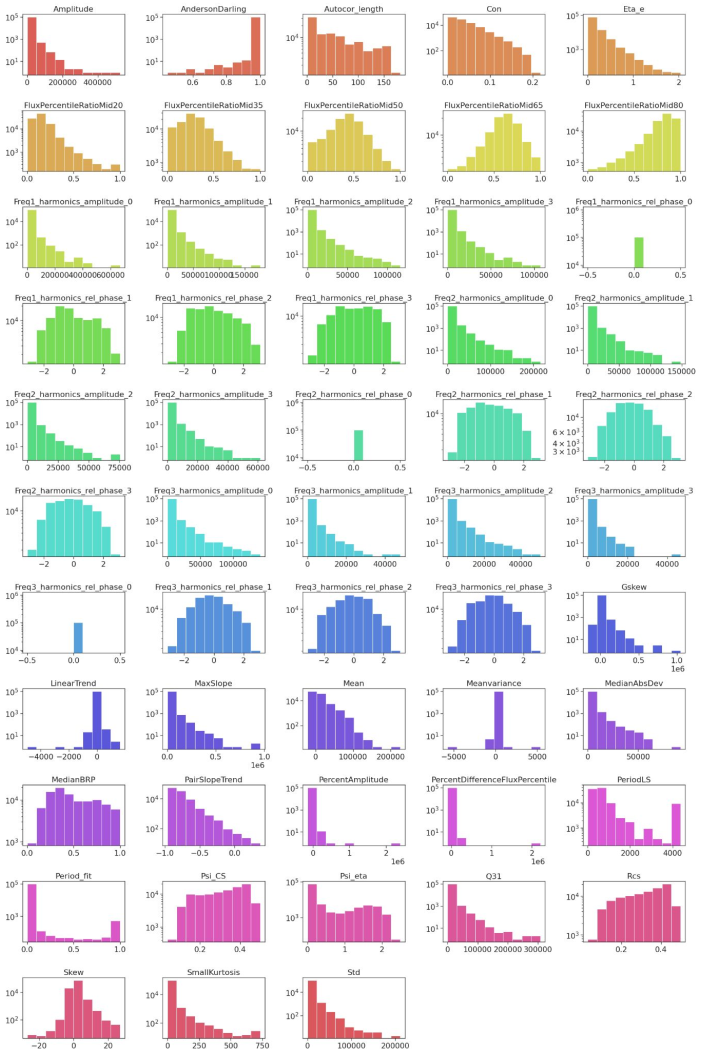
mean > global_mean + 2*global_std

Application: Financial Data

Dataset and features

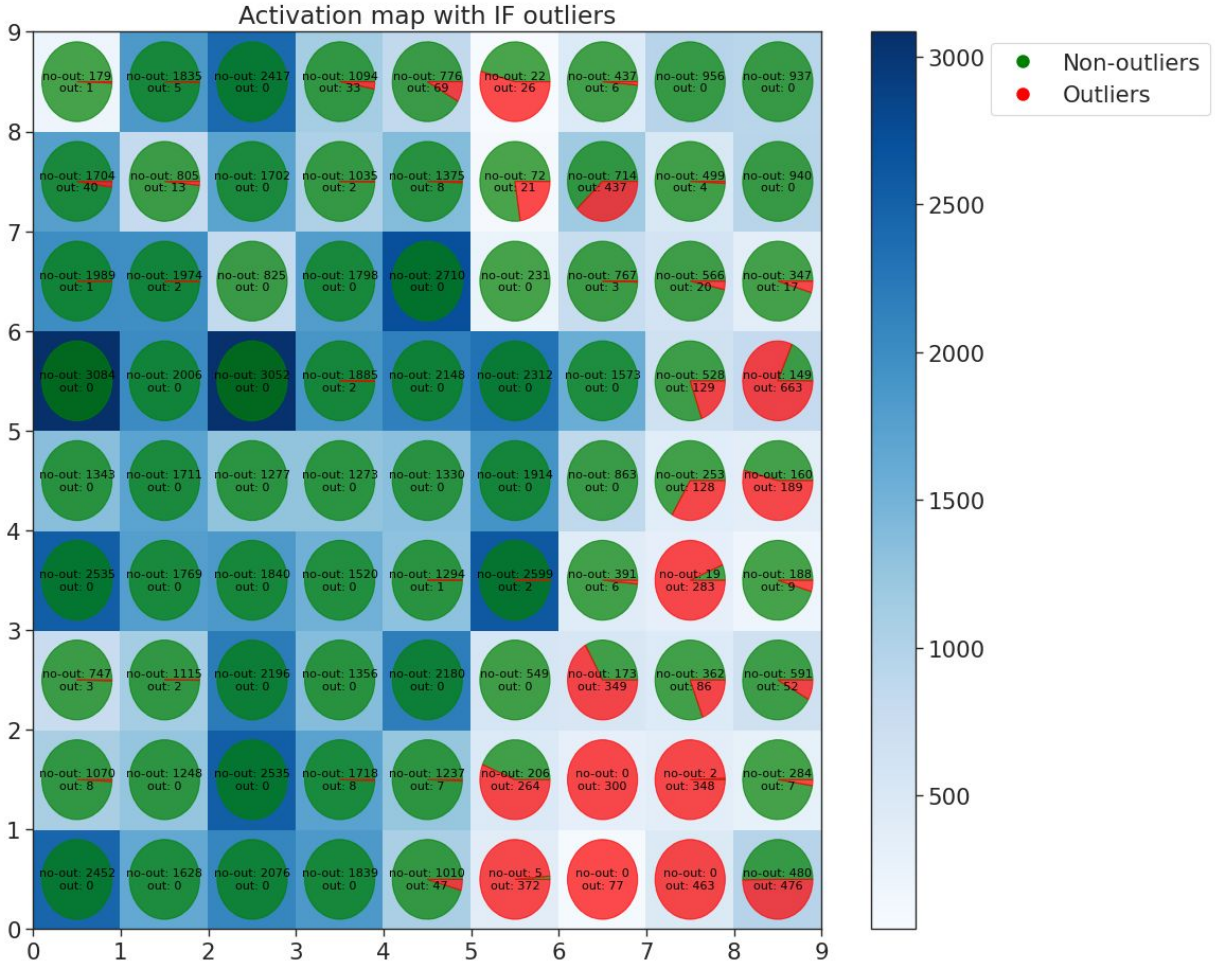
Statistical Features

- Amplitude
- AndersonDarling
- Autocor_length
- Con
- Eta_e
- FluxPercentileRatioMid20
- FluxPercentileRatioMid35
- FluxPercentileRatioMid50
- FluxPercentileRatioMid65
- FluxPercentileRatioMid80
- Freq1_harmonics_amplitude_0
- Freq1_harmonics_amplitude_1
- Freq1_harmonics_amplitude_2
- Freq1_harmonics_amplitude_3
- Freq1_harmonics_rel_phase_1
- Freq1_harmonics_rel_phase_2
- Freq1_harmonics_rel_phase_3
- Freq2_harmonics_amplitude_0
- Freq2_harmonics_amplitude_1
- Freq2_harmonics_amplitude_2
- Freq2_harmonics_amplitude_3
- Freq2_harmonics_rel_phase_1
- Freq2_harmonics_rel_phase_2
- Freq2_harmonics_rel_phase_3
- Freq3_harmonics_amplitude_0
- Freq3_harmonics_amplitude_1
- Freq3_harmonics_amplitude_2
- Freq3_harmonics_amplitude_3
- Freq3_harmonics_rel_phase_1
- Freq3_harmonics_rel_phase_2
- Freq3_harmonics_rel_phase_3
- Gskew
- LinearTrend
- MaxSlope
- Mean
- Meanvariance
- MedianAbsDev
- MedianBRP
- PairSlopeTrend
- PercentAmplitude
- PercentDifferenceFluxPercentile
- PeriodLS
- Period_fit
- Psi_CS
- Psi_eta
- Q31
- Rcs
- Skew
- SmallKurtosis
- Std



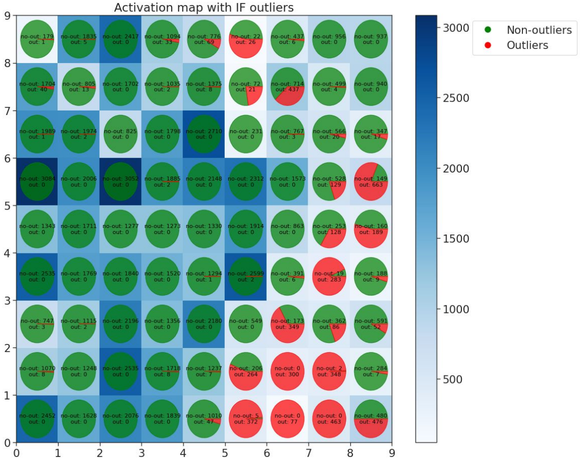
SOM - 1

- Statistical Features
- Preliminary step: **Isolation Forest** with 5% of contamination. 4989 outliers found.
- Neighborhood function: **Gaussian**.
- The pies on the activation map show the distribution of **outliers/not-outliers** in output from the **Isolation Forest**.



Feature Index

- 0: Amplitude
- 1: AndersonDarling
- 2: Autocor_length
- 3: Con
- 4: Eta_e
- 5: FluxPercentileRatioMid20
- 6: FluxPercentileRatioMid35
- 7: FluxPercentileRatioMid50
- 8: FluxPercentileRatioMid65
- 9: FluxPercentileRatioMid80
- 10: Freq1_harmonics_amplitude_0
- 11: Freq1_harmonics_amplitude_1
- 12: Freq1_harmonics_amplitude_2
- 13: Freq1_harmonics_amplitude_3
- 14: Freq1_harmonics_rel_phase_1
- 15: Freq1_harmonics_rel_phase_2
- 16: Freq1_harmonics_rel_phase_3
- 17: Freq2_harmonics_amplitude_0
- 18: Freq2_harmonics_amplitude_1
- 19: Freq2_harmonics_amplitude_2
- 20: Freq2_harmonics_amplitude_3
- 21: Freq2_harmonics_rel_phase_1
- 22: Freq2_harmonics_rel_phase_2
- 23: Freq2_harmonics_rel_phase_3
- 24: Freq3_harmonics_amplitude_0
- 25: Freq3_harmonics_amplitude_1
- 26: Freq3_harmonics_amplitude_2
- 27: Freq3_harmonics_amplitude_3
- 28: Freq3_harmonics_rel_phase_1
- 29: Freq3_harmonics_rel_phase_2
- 30: Freq3_harmonics_rel_phase_3
- 31: Gskew
- 32: LinearTrend
- 33: MaxSlope
- 34: Mean
- 35: Meanvariance
- 36: MedianAbsDev
- 37: MedianBRP
- 38: PairSlopeTrend
- 39: PercentAmplitude
- 40: PercentDifferenceFluxPercentile
- 41: PeriodLS
- 42: Period_fit
- 43: Psi_CS
- 44: Psi_eta
- 45: Q31
- 46: Rcs
- 47: Skew
- 48: SmallKurtosis
- 49: Std



For each neuron, we calculated the global mean and standard deviation. Features are in red if:

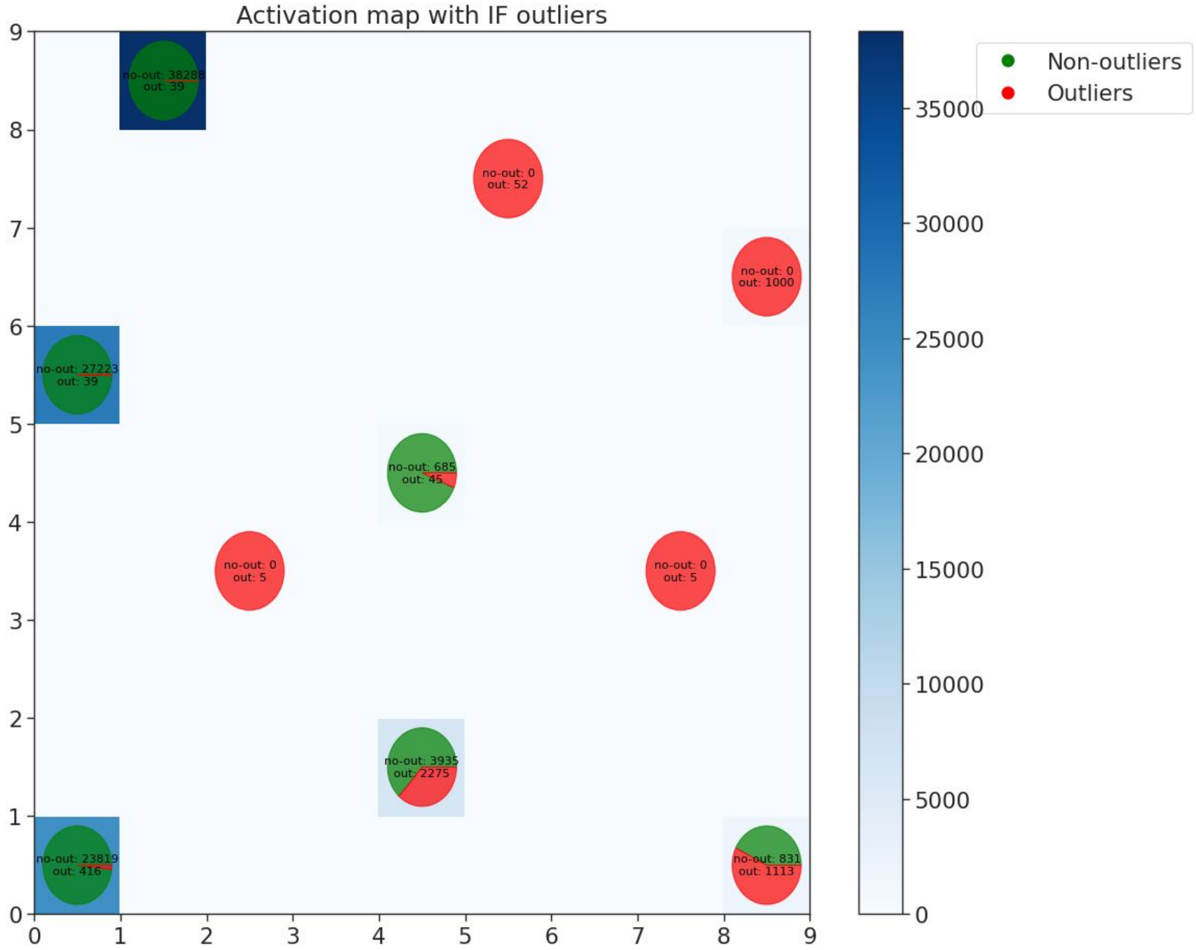
mean < global_mean - 2*global_std

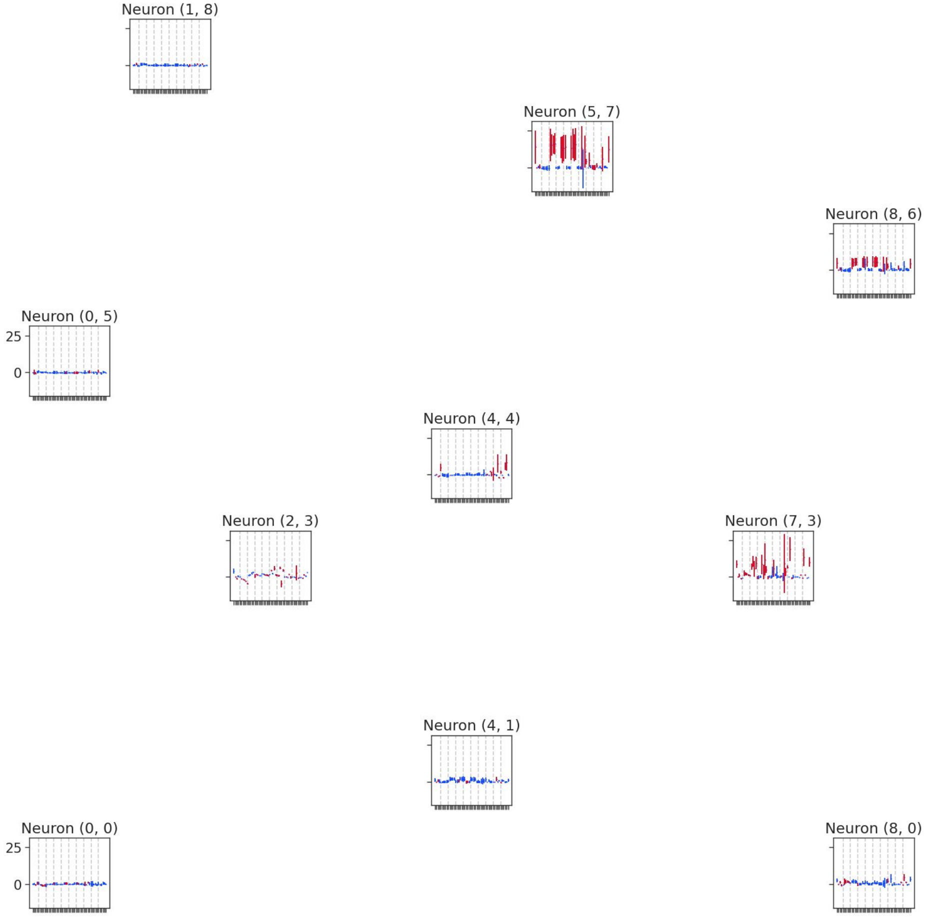
or

mean > global_mean + 2*global_std

SOM - 2

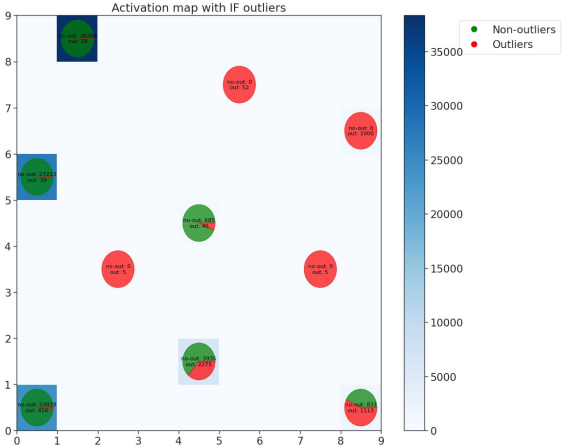
- Statistical Features
- Preliminary step: **Isolation Forest** with 5% of contamination. 4989 outliers found.
- Neighborhood function: **Mexican Hat**.
- The pies on the activation map show the distribution of **outliers/not-outliers** in output from the **Isolation Forest**.





Feature Index

- 0: Amplitude
- 1: AndersonDarling
- 2: Autocor_length
- 3: Con
- 4: Eta_e
- 5: FluxPercentileRatioMid20
- 6: FluxPercentileRatioMid35
- 7: FluxPercentileRatioMid50
- 8: FluxPercentileRatioMid65
- 9: FluxPercentileRatioMid80
- 10: Freq1_harmonics_amplitude_0
- 11: Freq1_harmonics_amplitude_1
- 12: Freq1_harmonics_amplitude_2
- 13: Freq1_harmonics_amplitude_3
- 14: Freq1_harmonics_rel_phase_1
- 15: Freq1_harmonics_rel_phase_2
- 16: Freq1_harmonics_rel_phase_3
- 17: Freq2_harmonics_amplitude_0
- 18: Freq2_harmonics_amplitude_1
- 19: Freq2_harmonics_amplitude_2
- 20: Freq2_harmonics_amplitude_3
- 21: Freq2_harmonics_rel_phase_1
- 22: Freq2_harmonics_rel_phase_2
- 23: Freq2_harmonics_rel_phase_3
- 24: Freq3_harmonics_amplitude_0
- 25: Freq3_harmonics_amplitude_1
- 26: Freq3_harmonics_amplitude_2
- 27: Freq3_harmonics_amplitude_3
- 28: Freq3_harmonics_rel_phase_1
- 29: Freq3_harmonics_rel_phase_2
- 30: Freq3_harmonics_rel_phase_3
- 31: Gskew
- 32: LinearTrend
- 33: MaxSlope
- 34: Mean
- 35: Meanvariance
- 36: MedianAbsDev
- 37: MedianBRP
- 38: PairSlopeTrend
- 39: PercentAmplitude
- 40: PercentDifferenceFluxPercentile
- 41: PeriodLS
- 42: Period_fit
- 43: Psi_CS
- 44: Psi_eta
- 45: Q31
- 46: Rcs
- 47: Skew
- 48: SmallKurtosis
- 49: Std



For each neuron, we calculated the global mean and standard deviation. Features are in red if:

mean < global_mean - 2*global_std

or

mean > global_mean + 2*global_std

Conclusions

Conclusions

Cross-Disciplinary Insights

- Time series can be adapted across various domains, demonstrating that innovations in one field can drive advancements in others.

Identification of Similar Objects

- This method allows for the identification of objects with similar characteristics, such as classifying Active Galactic Nuclei based on specific features.

Anomaly Detection and Outlier Identification

- Time series analysis is crucial for detecting anomalies, particularly useful in fields like banking for identifying outliers that may indicate fraud or other irregular activities.

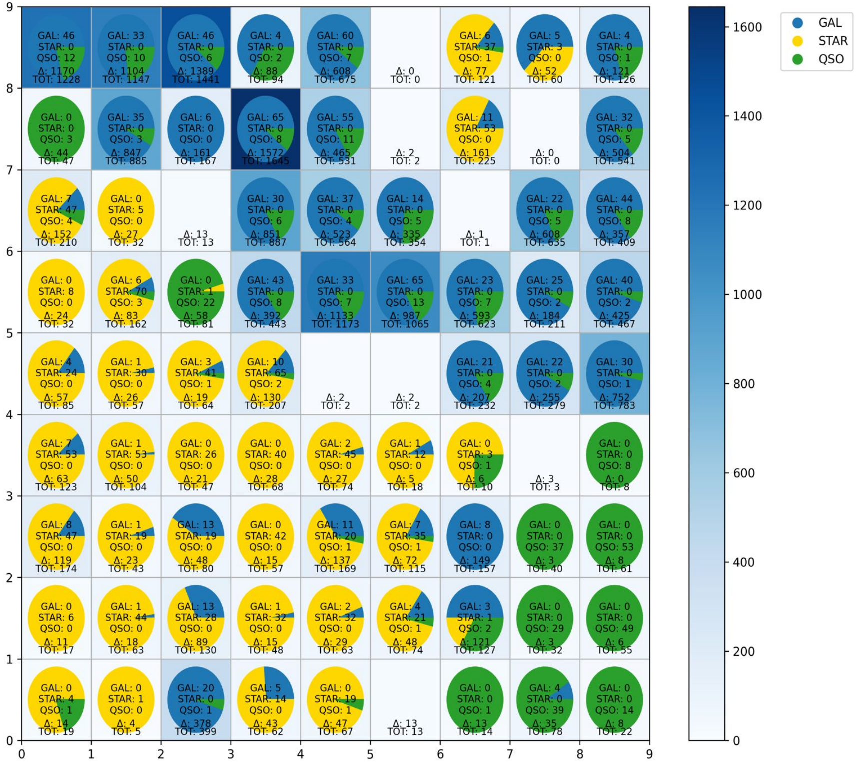
Thank you!

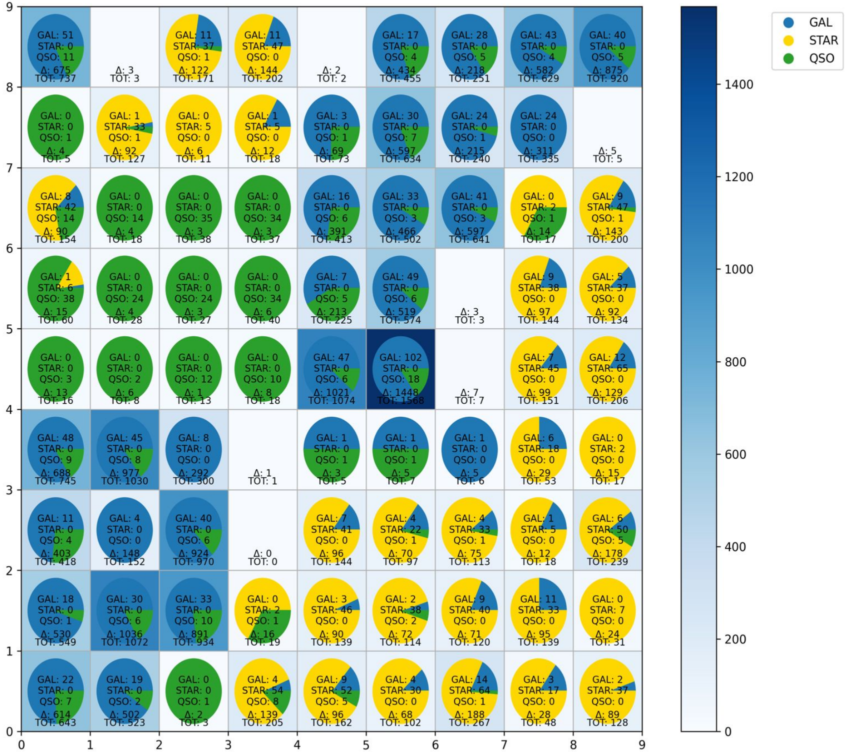
Ylenia Maruccia
E: ylenia.maruccia@inaf.it

INAF
OSSERVATORIO ASTRONOMICO
DI CAPODIMONTE

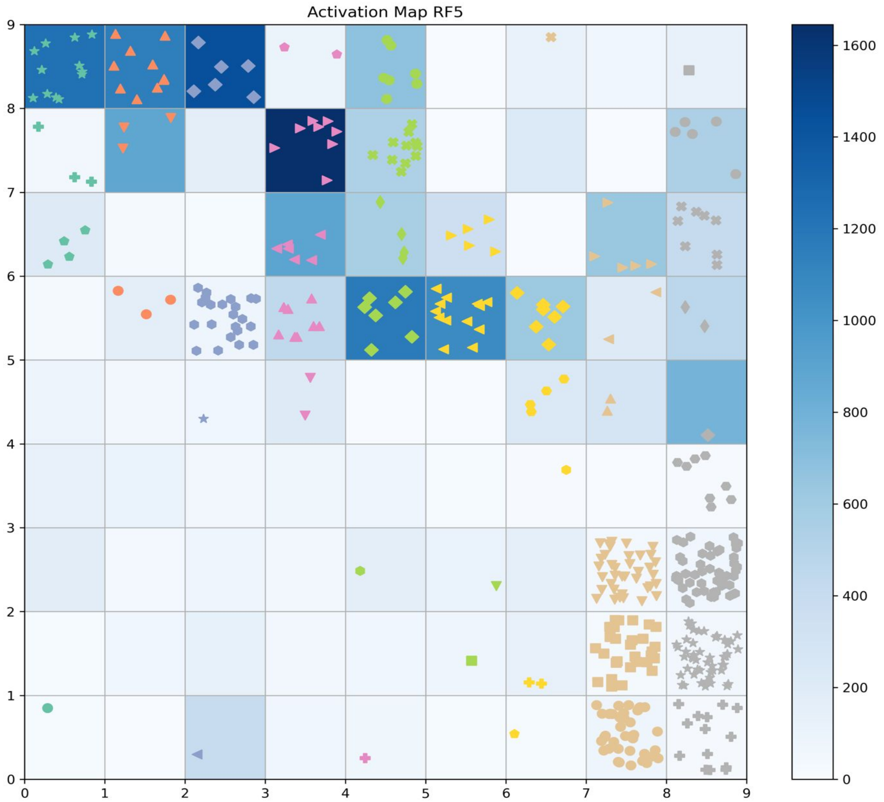
All features vs No-colors

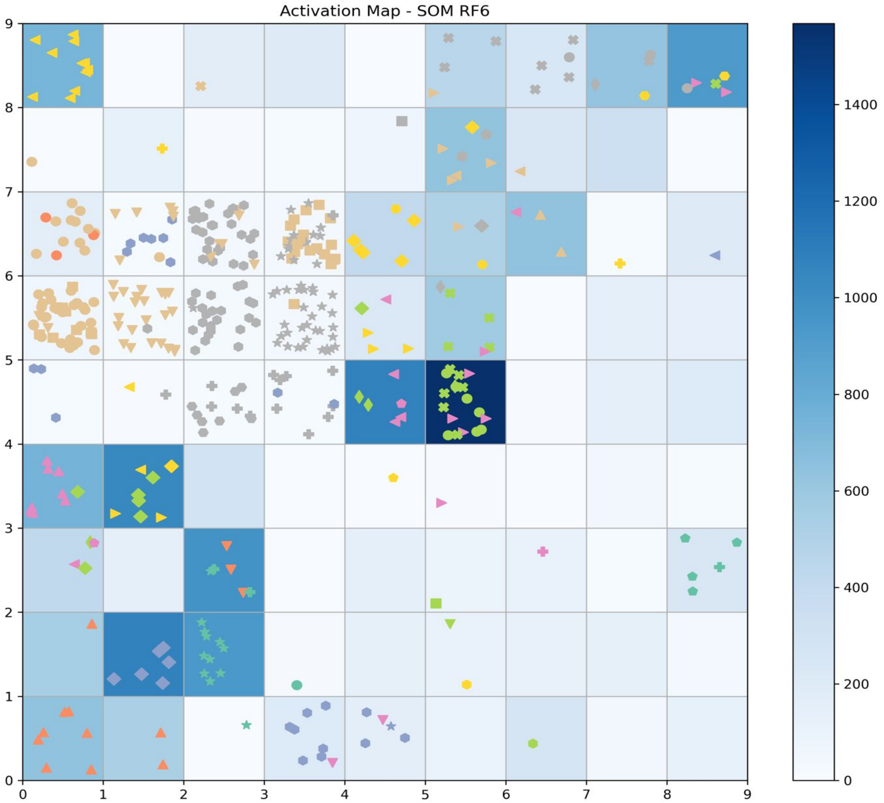
Activation map: All features vs No-colors



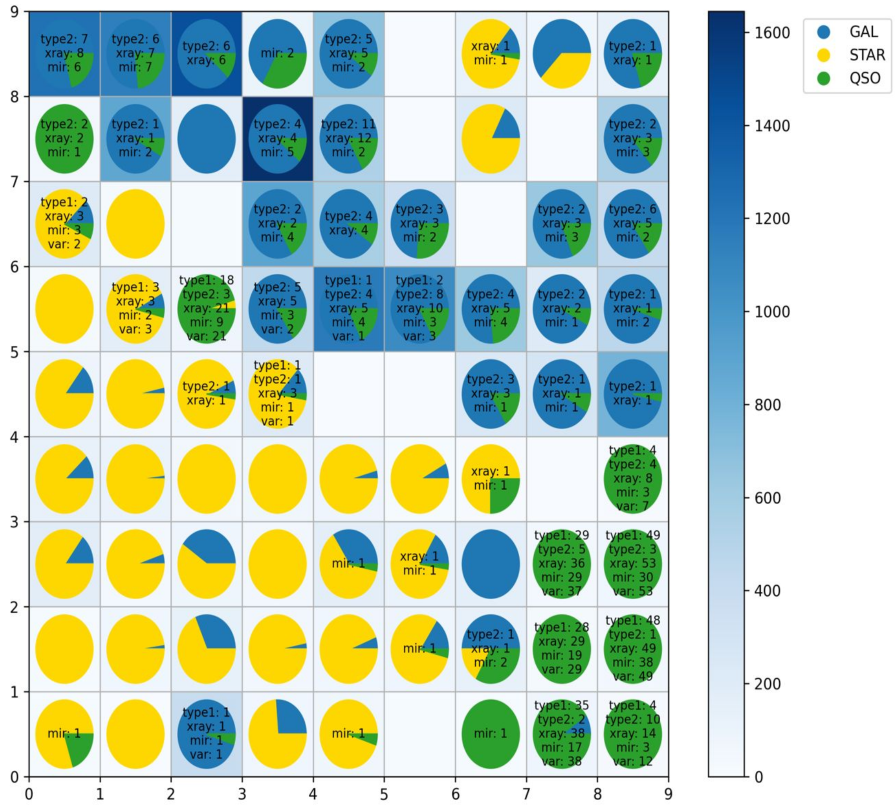


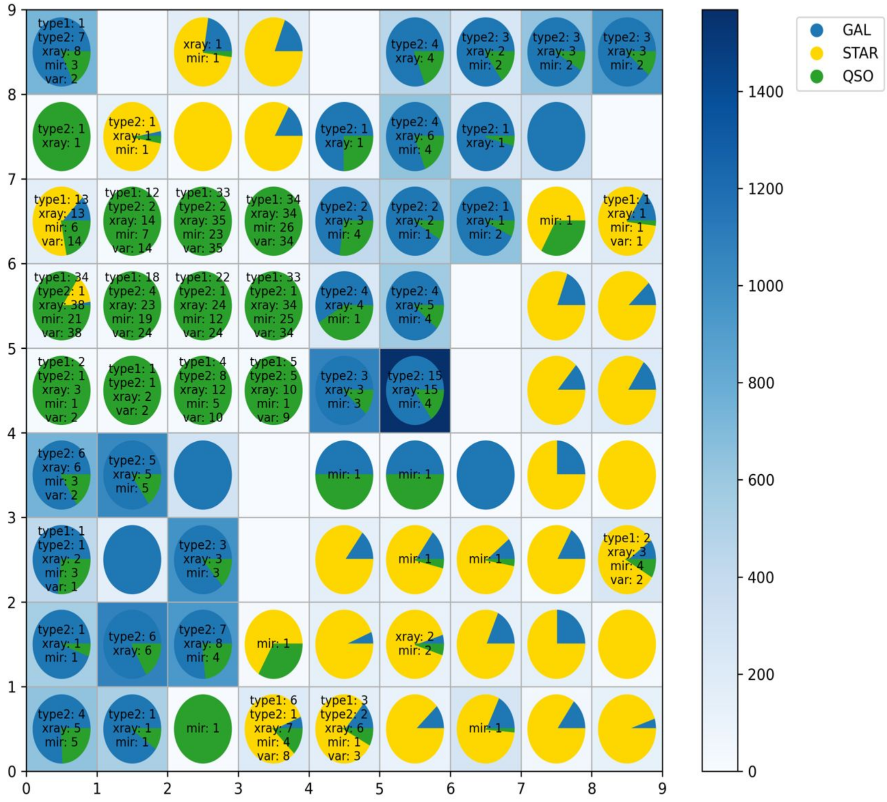
Mapping QSO: All features vs No-colors



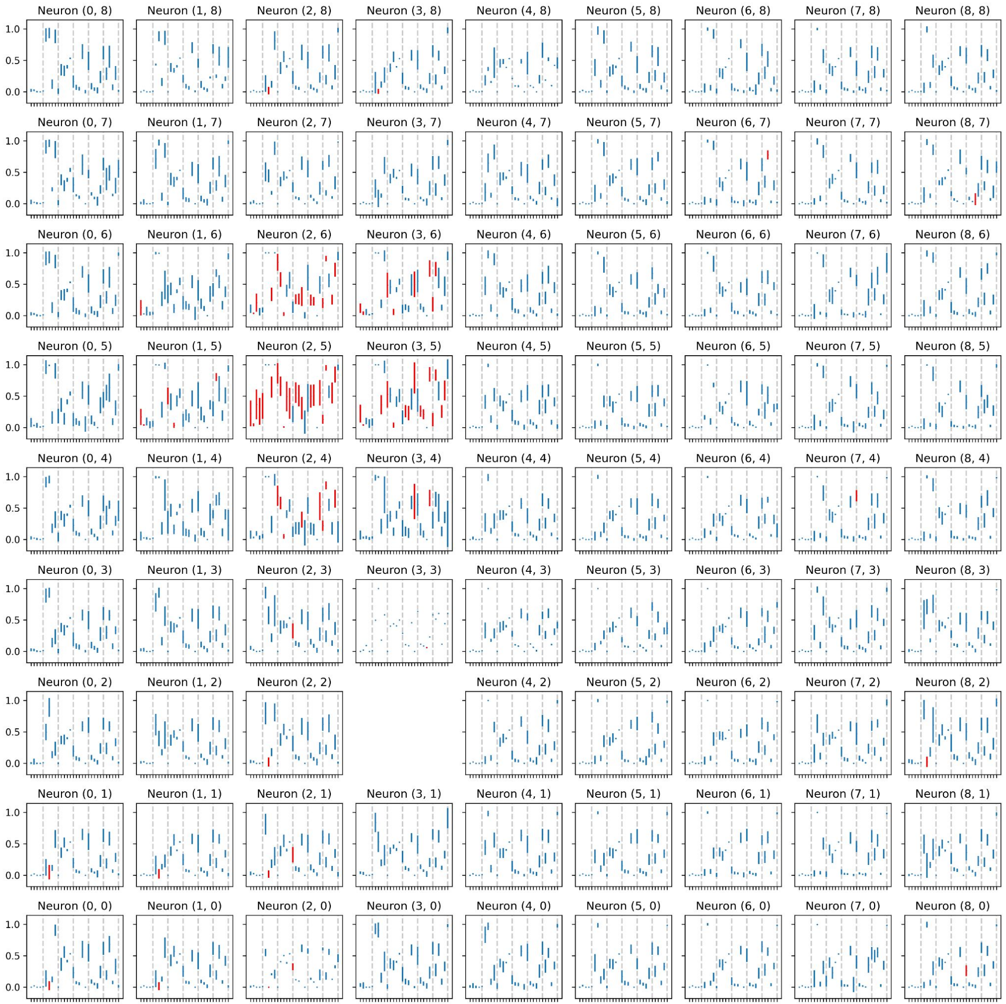


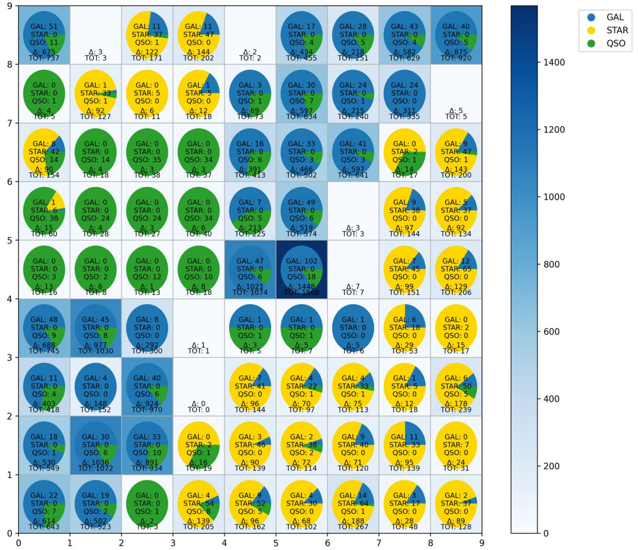
Activation map and AGN Types: All features vs No-colors





No-colors





For each neuron, we calculated the global mean and standard deviation.
Features are in red if:

mean < global_mean - 2*global_std

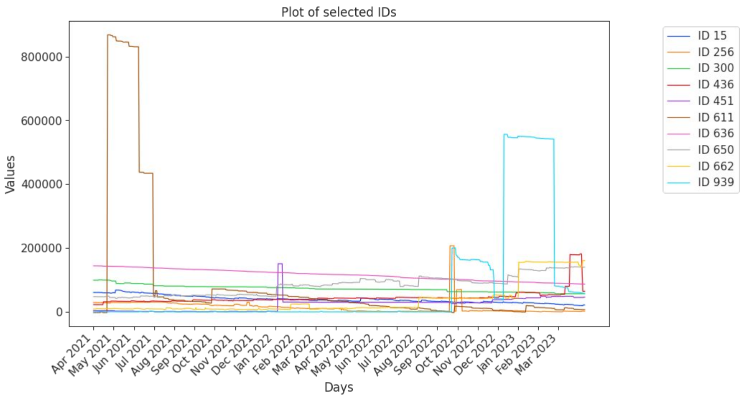
or

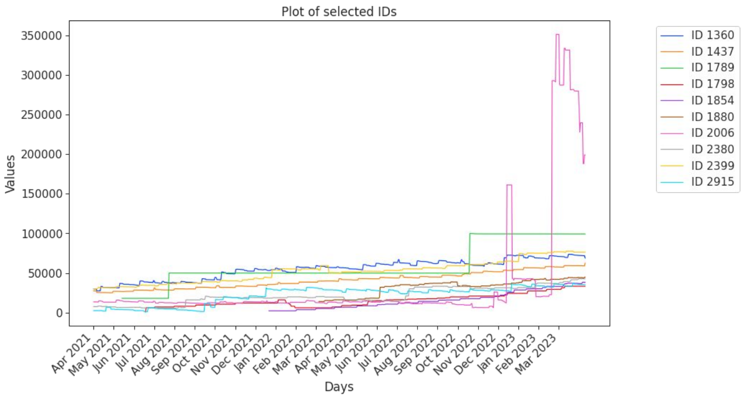
mean > global_mean + 2*global_std

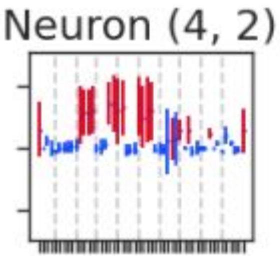
Examples of time series in specific neurons

Cella (4,2)

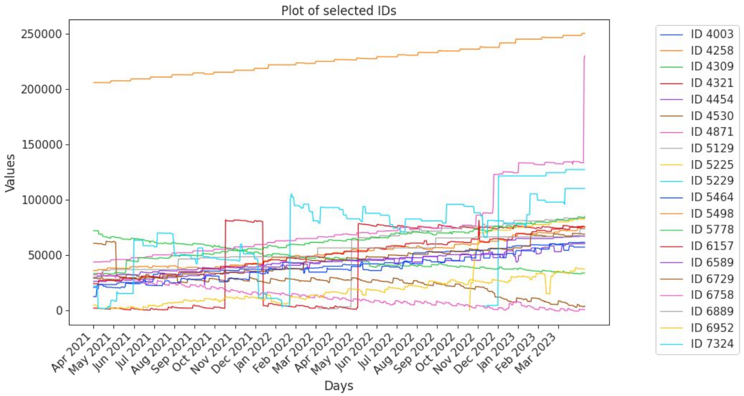
Only 763 IF outliers

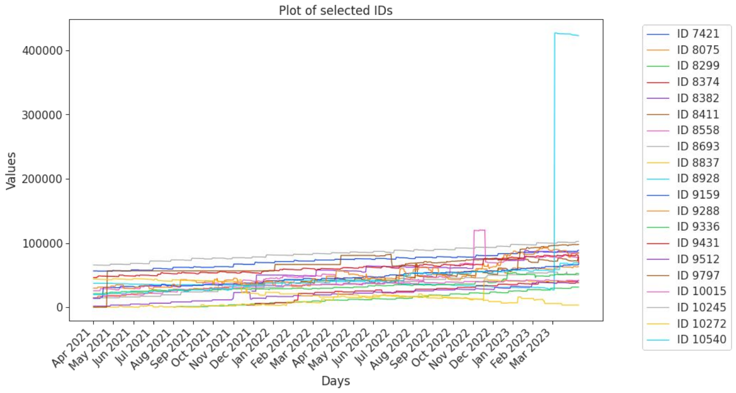




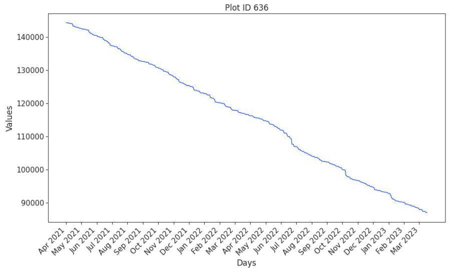


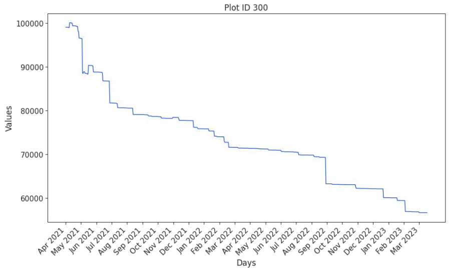

```
---------------------------
Neuron (4, 2)
Amplitude
Freq1_harmonics_amplitude_0
Freq1_harmonics_amplitude_1
Freq1_harmonics_amplitude_2
Freq1_harmonics_amplitude_3
Freq2_harmonics_amplitude_0
Freq2_harmonics_amplitude_1
Freq2_harmonics_amplitude_2
Freq2_harmonics_amplitude_3
Freq3_harmonics_amplitude_0
Freq3_harmonics_amplitude_1
Freq3_harmonics_amplitude_2
Freq3_harmonics_amplitude_3
LinearTrend
Mean
MedianAbsDev
PeriodLS
Std
---------------------------
```

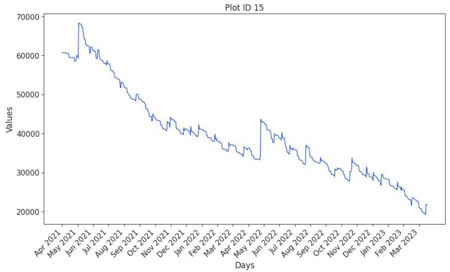


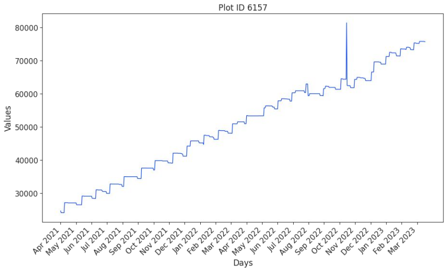


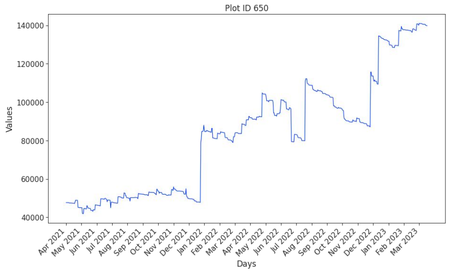
Cella (4,2)

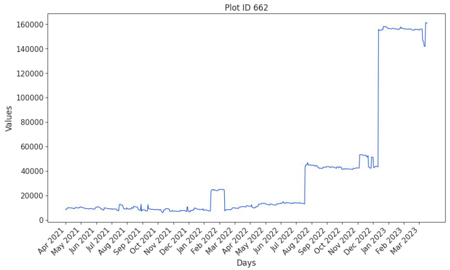

Plot ID 636
Values
Days

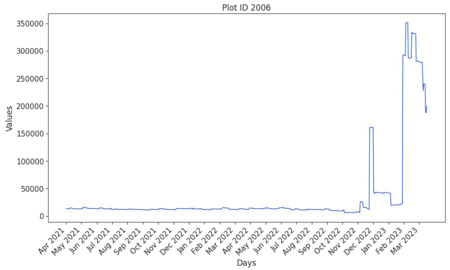

Plot ID 300
Values
Days

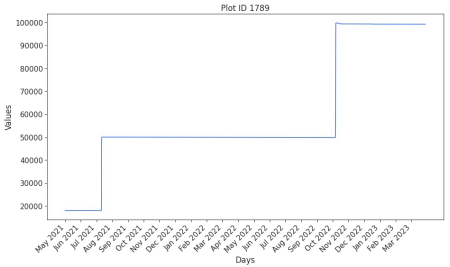

Plot ID 15
Values
Days

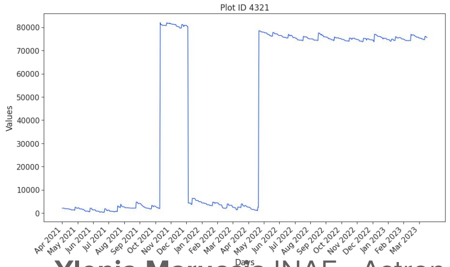

Plot ID 6157
Values
Days

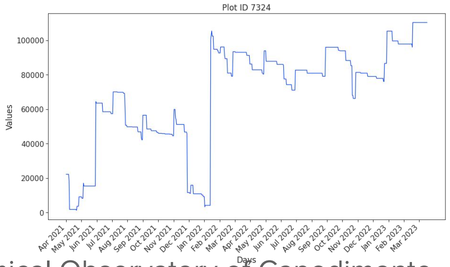

Plot ID 650
Values
Days

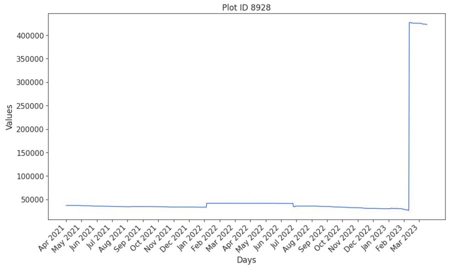

Plot ID 662
Values
Days

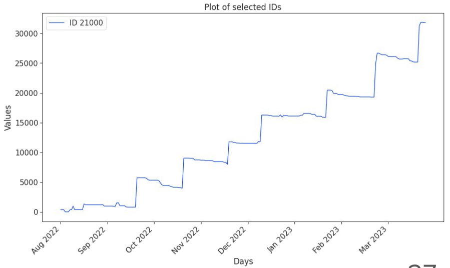

Plot ID 2006
Values
Days


Plot ID 1789
Values
Days


Plot ID 4321
Values
Days

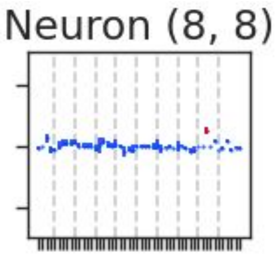

Plot ID 7324
Values
Days

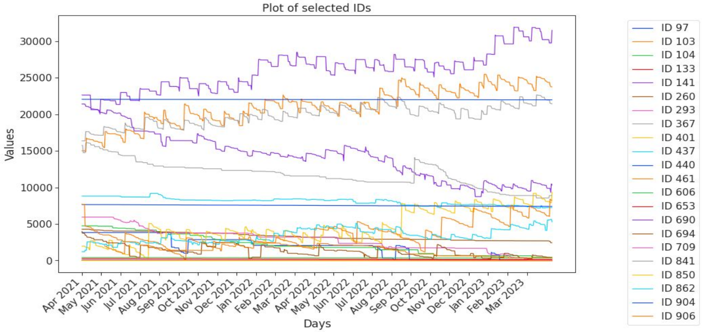

Plot ID 8928
Values
Days

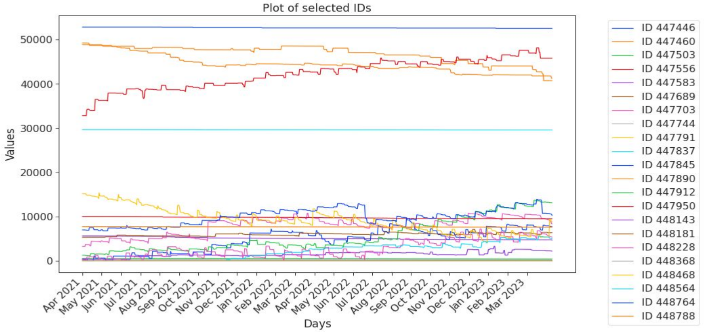

Plot of selected IDs
ID 21000
Values
Days

Cella (8,8)

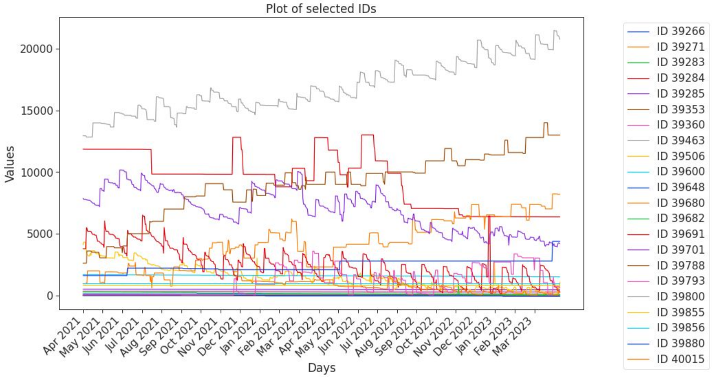
No IF outliers

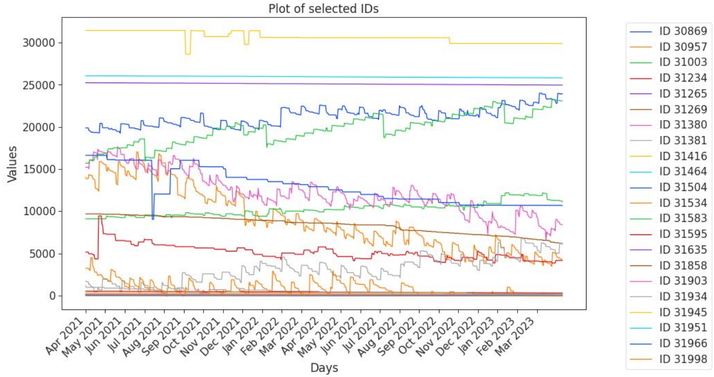




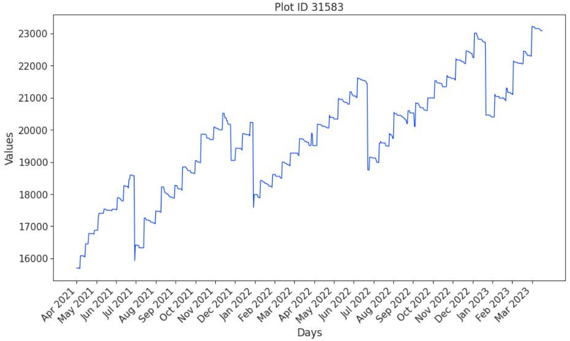


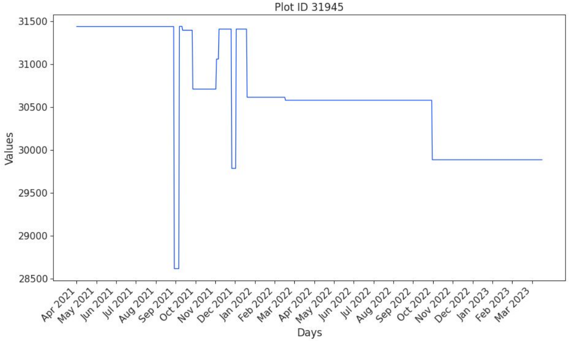

```
---------------------------
Neuron (8, 8)
PeriodLS
---------------------------
```

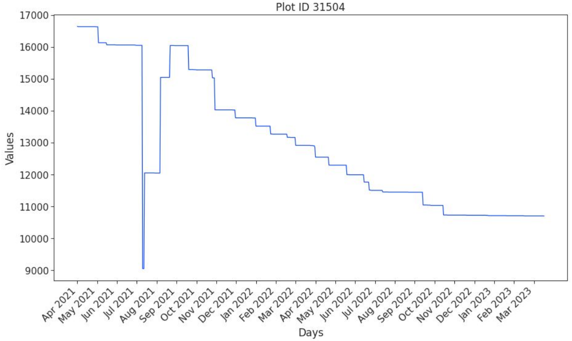


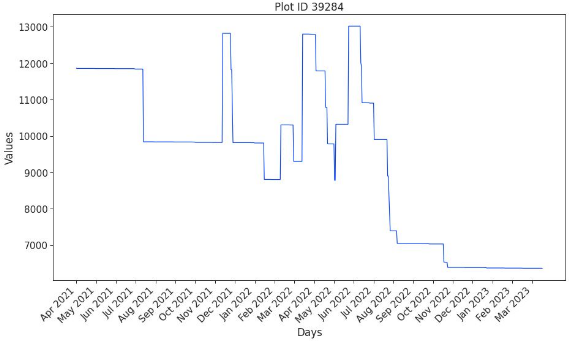


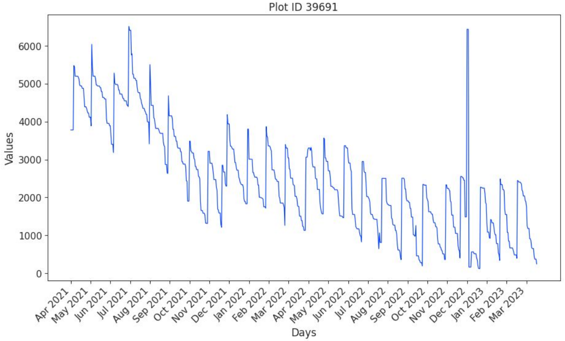
Cella (8,8)

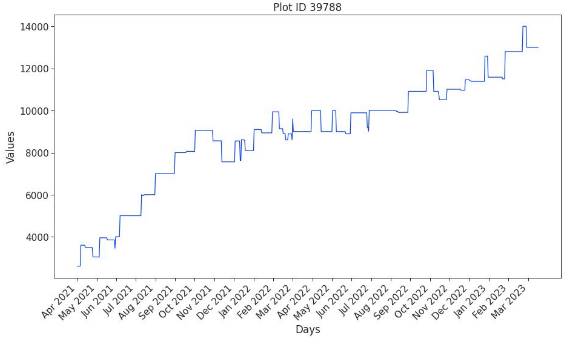

Plot ID 31583
Values
23000
22000
21000
20000
19000
18000
17000
16000
Apr 2021 May 2021 Jun 2021 Jul 2021 Aug 2021 Sep 2021 Oct 2021 Nov 2021 Dec 2021 Jan 2022 Feb 2022 Mar 2022 Apr 2022 May 2022 Jun 2022 Jul 2022 Aug 2022 Sep 2022 Oct 2022 Nov 2022 Dec 2022 Jan 2023 Feb 2023 Mar 2023
Days


Plot ID 31945
Values
31500
31000
30500
30000
29500
29000
28500
Apr 2021 May 2021 Jun 2021 Jul 2021 Aug 2021 Sep 2021 Oct 2021 Nov 2021 Dec 2021 Jan 2022 Feb 2022 Mar 2022 Apr 2022 May 2022 Jun 2022 Jul 2022 Aug 2022 Sep 2022 Oct 2022 Nov 2022 Dec 2022 Jan 2023 Feb 2023 Mar 2023
Days


Plot ID 31504
Values
17000
16000
15000
14000
13000
12000
11000
10000
9000
Apr 2021 May 2021 Jun 2021 Jul 2021 Aug 2021 Sep 2021 Oct 2021 Nov 2021 Dec 2021 Jan 2022 Feb 2022 Mar 2022 Apr 2022 May 2022 Jun 2022 Jul 2022 Aug 2022 Sep 2022 Oct 2022 Nov 2022 Dec 2022 Jan 2023 Feb 2023 Mar 2023
Days

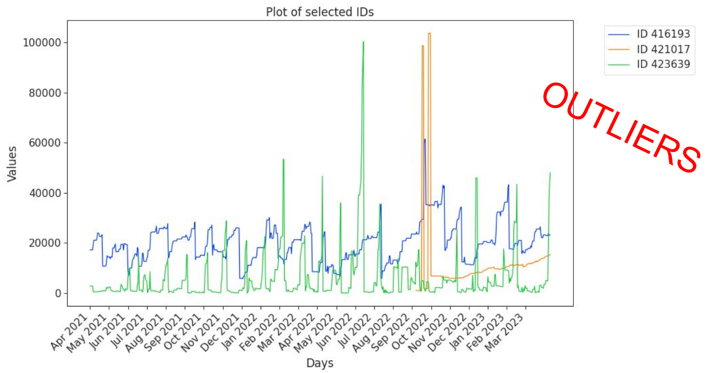

Plot ID 39284
Values
13000
12000
11000
10000
9000
8000
7000
Apr 2021 May 2021 Jun 2021 Jul 2021 Aug 2021 Sep 2021 Oct 2021 Nov 2021 Dec 2021 Jan 2022 Feb 2022 Mar 2022 Apr 2022 May 2022 Jun 2022 Jul 2022 Aug 2022 Sep 2022 Oct 2022 Nov 2022 Dec 2022 Jan 2023 Feb 2023 Mar 2023
Days


Plot ID 39691
Values
6000
5000
4000
3000
2000
1000
0
Apr 2021 May 2021 Jun 2021 Jul 2021 Aug 2021 Sep 2021 Oct 2021 Nov 2021 Dec 2021 Jan 2022 Feb 2022 Mar 2022 Apr 2022 May 2022 Jun 2022 Jul 2022 Aug 2022 Sep 2022 Oct 2022 Nov 2022 Dec 2022 Jan 2023 Feb 2023 Mar 2023
Days

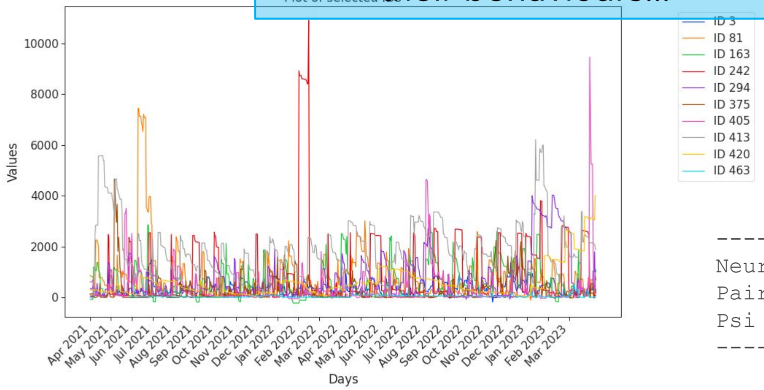

Plot ID 39788
Values
14000
12000
10000
8000
6000
4000
Apr 2021 May 2021 Jun 2021 Jul 2021 Aug 2021 Sep 2021 Oct 2021 Nov 2021 Dec 2021 Jan 2022 Feb 2022 Mar 2022 Apr 2022 May 2022 Jun 2022 Jul 2022 Aug 2022 Sep 2022 Oct 2022 Nov 2022 Dec 2022 Jan 2023 Feb 2023 Mar 2023
Days

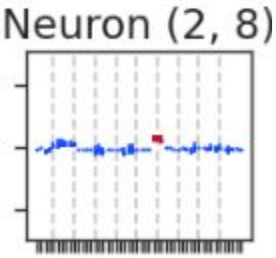
Cella (0,8)

3 IF outliers
1813 not-outliers, nevertheless their behaviours…



OUTLIERS






```
---------------------------
Neuron (0, 8)
PairSlopeTrend
Psi eta
---------------------------
```