Applying AI capabilities to address Operations challenges in ECMWF Products Team

Aditya Ahuja, Adithya Niranjan
Mentors: Matthew Manoussakis, Peter Dueben

Use Case

- ECMWF's data services provide massive tailored data to Members States, as well as to commercial and public users.
- The scale and complexity of these services make monitoring and troubleshooting them increasingly difficult. For example, in a typical day, the MARS service processes 1.5 million user requests and delivers 400 TB of data.

<u>Aim of the project:</u>
- To investigate how the application of AI and Machine Learning techniques can be used to diagnose problems during operations of these services (such as crashes of data servers) and to pre-empt developing issues.
- To create predictive insights based on predictive analytics.
- To help ECMWF raise early alarms and take immediate action on impending issues.

The methods developed in this project can potentially be extended to a number of services at ECMWF.
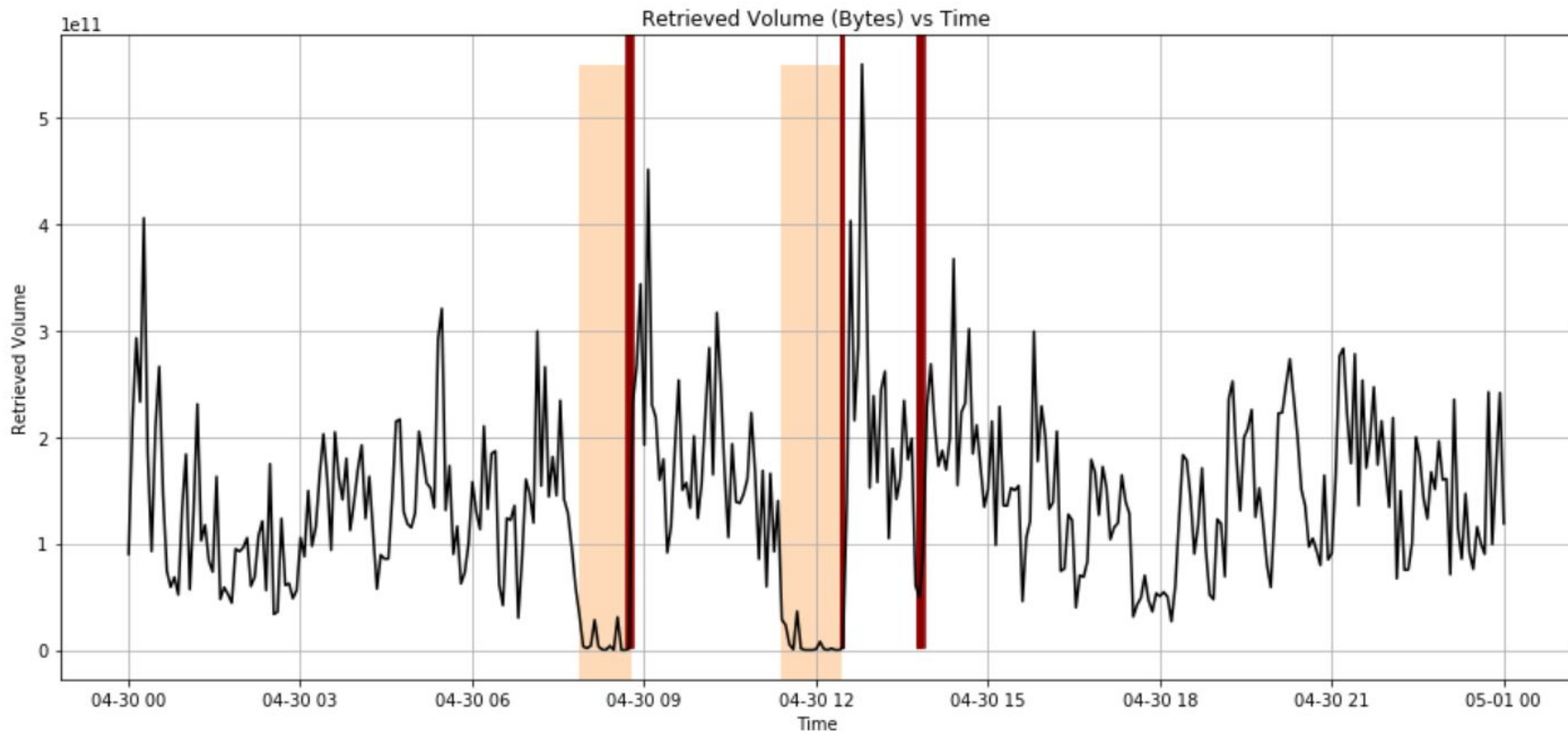
Dataset - The Log Files

$startdate='20200521';$starttime='23:59:59';$verb='retrieve';$version='20200506';$application='mars';$class='od';$type='fc';$stream='oper';$expver='0001';$retdate='20200521';$age='0';$nbdates='1';$bonddataset='0';$reqno='1';$multitarget='new';$fields='1';$database='fdb5_prod';$bytes='13204588';$written='13204588';$interpolated='0';$transfertime='0';$writetarget='0';$cpu='0';$elapsed='0';$status='ok';$stopdate='20200521';$stoptime='23:59:59';$user='o#m#9';$category='l*a*#*e';$account='d*e*#o';$abc='c#*e*o';$environment='batch';$host='0*l#*#0';$pid='40788';$class='od';$type='fc';$stream='oper';$expver='0001';$repres='gg';$levtype='sfc';$param='129.128';$date='20200521';$time='0000';$step='0';$domain='g';$target='/*#a#*b';$resol='1279';$grid='o1280';

Dataset - A Sample Time Series

Retrieved Volume (Bytes) vs Time

Motivation

- While there are several time-series anomaly detection packages online, most have some drawbacks when it comes to real time detection:
    - Often designed and implemented for offline detection.
    - Usually old and deprecated codebases.
    - No common interface that combines different functionality.
    - No estimate of robustness for the task at hand.

- We wanted to build an Anomaly Detection system that works out of the box, combines several different methods, gives the end user multiple choices (accessible through config files) and handles streaming data, with a GUI that allows for real time monitoring.
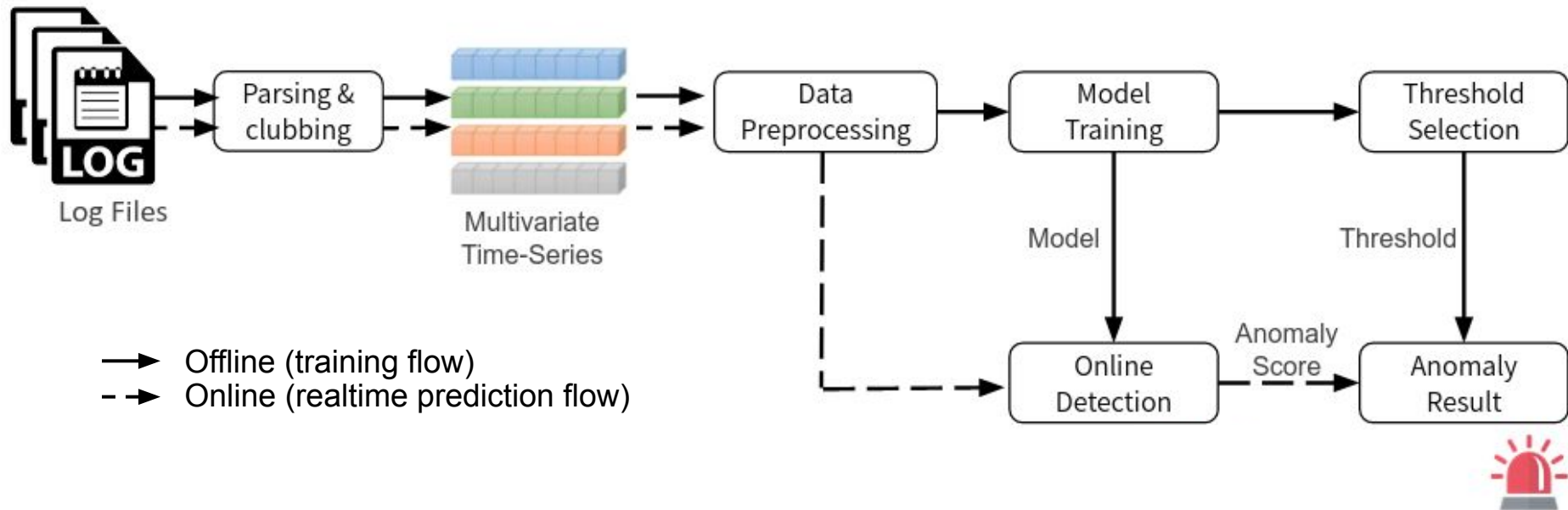
What we Built

- We built on existing open-source time-series anomaly detection libraries, such as Telemanom, Banpei and DeepADoTS, re-implementing state of the art methods to work with real time streaming data.
- We then implemented our own forecasting methods and combined them into this loop to get estimates of the anomaly scores.
- We worked on implementing a multithreaded training pipeline, so that training can happen in the background, while the algorithm continues to generate real time predictions.
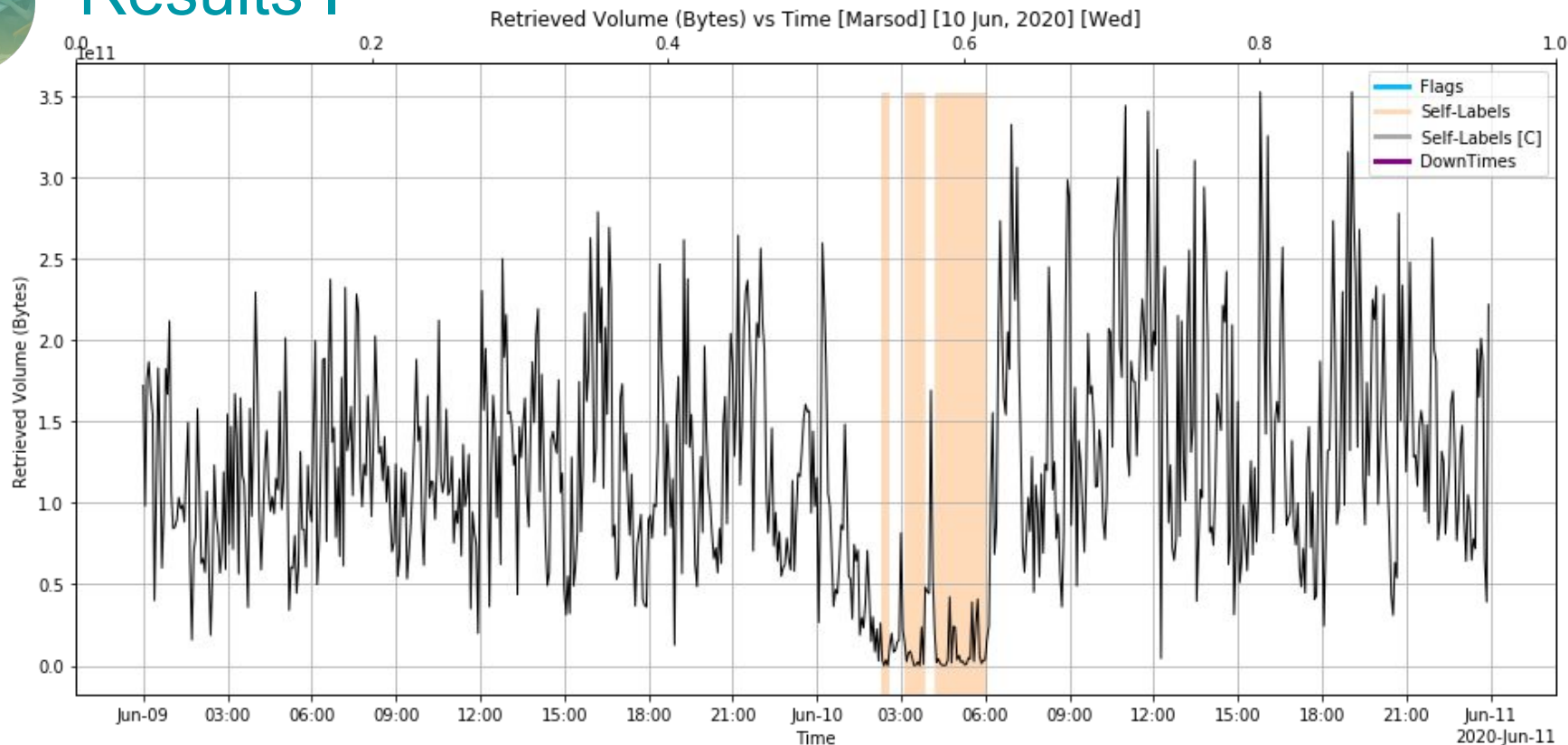- Finally, we implemented a GUI that allows real time visualization of the estimated anomaly score.
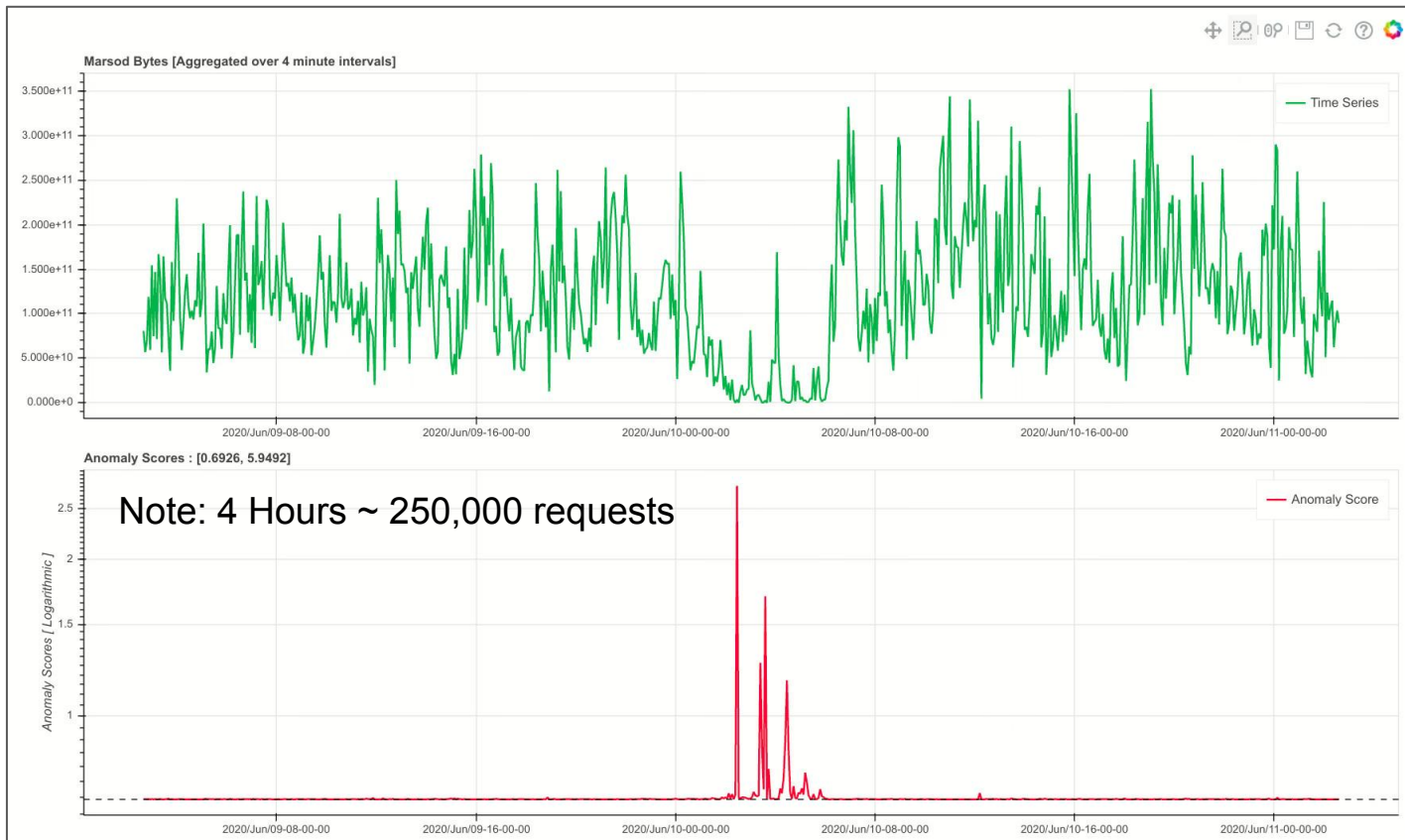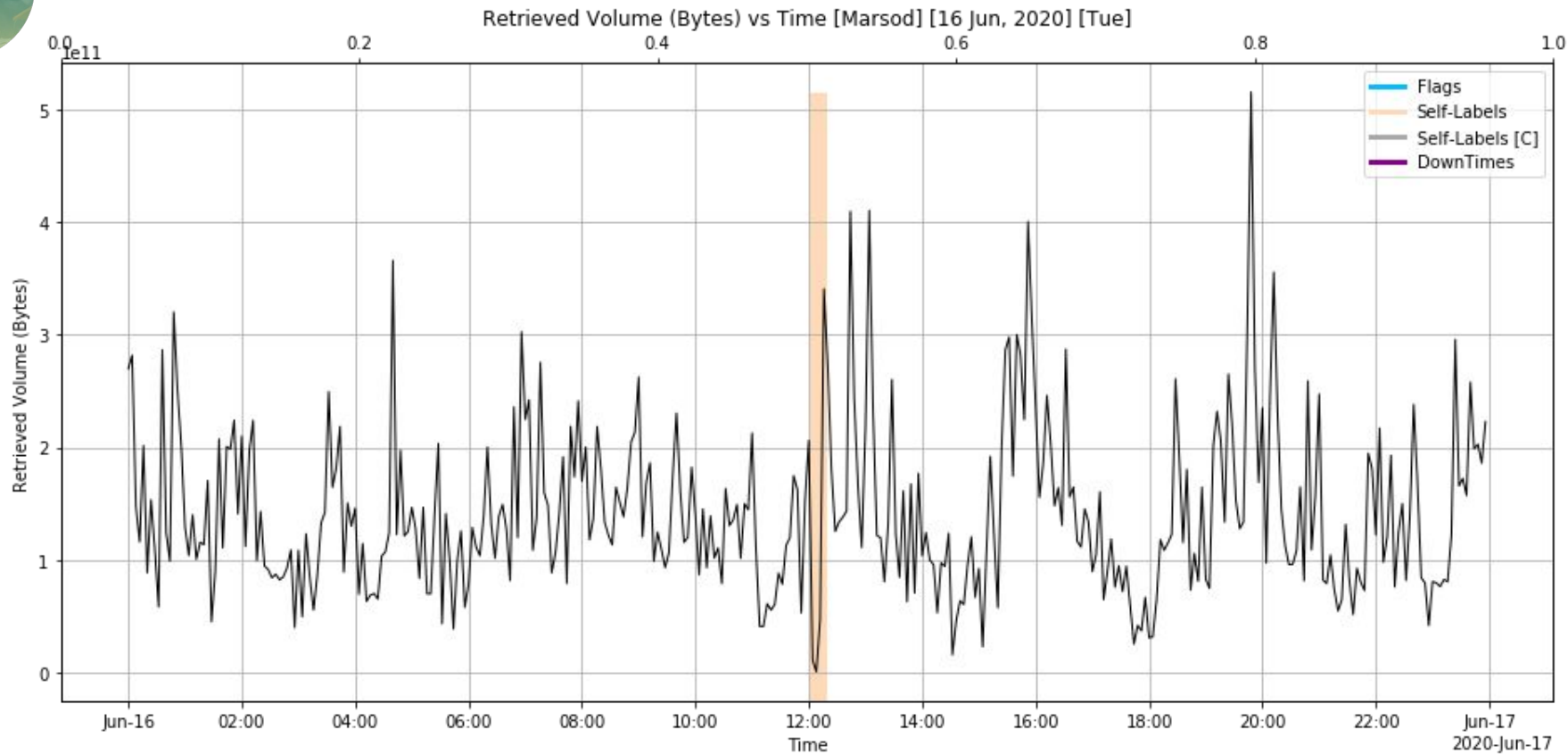
Realtime Pipeline

Log Files → Parsing & clubbing → Multivariate Time-Series → Data Preprocessing → Model Training → Threshold Selection

Data Preprocessing → Model Training → (Model) → Online Detection → (Anomaly Score) → Anomaly Result

Threshold Selection → (Threshold) → Anomaly Result

→ Offline (training flow)
- → Online (realtime prediction flow)

Results I

Retrieved Volume (Bytes) vs Time [Marsod] [10 Jun, 2020] [Wed]

Results I

Marsod Bytes [Aggregated over 4 minute intervals]

Anomaly Scores : [0.6926, 5.9492]

Note: 4 Hours ~ 250,000 requests

Results II

Retrieved Volume (Bytes) vs Time [Marsod] [16 Jun, 2020] [Tue]

Results II

Marsod Bytes [Aggregated over 4 minute intervals]

Anomaly Scores : [0.6926, 5.9492]

Challenges and Learnings

- Challenges
    - Working with noisy data and mostly self-supervised labels.
    - Working with information available only from completed logs.
    - Implementing with real-world constraints in mind.

- Learnings:
    - We explored current Time-Series anomaly detection and forecasting research as well as their applications.
    - We gained valuable experience working with real world data and settings.
    - We realized the limitations of current methods and learnt how we can extend them.
    - We also learnt a lot about software development - Developing an end to end system.

Usage and Demo

- How to extend/use this framework for similar tasks - new log files generated from other services or other time series datasets?
    - Modify the parser function to generate time series from the new logs files.
    - Tune hyperparameters for the newly generated Time Series.
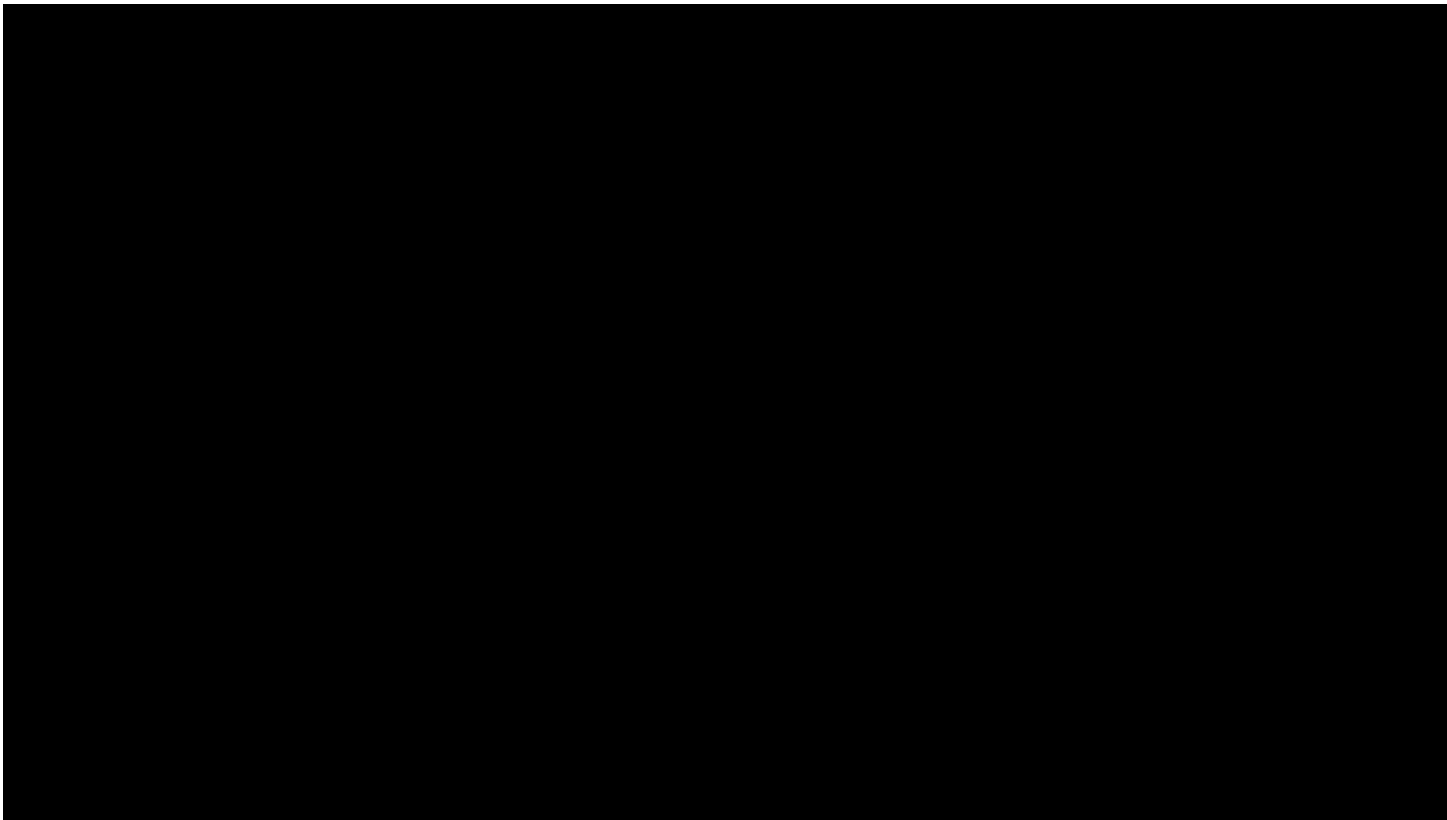
- Time for a Demo!

ECMWF
EUROPEAN WEATHER CLOUD
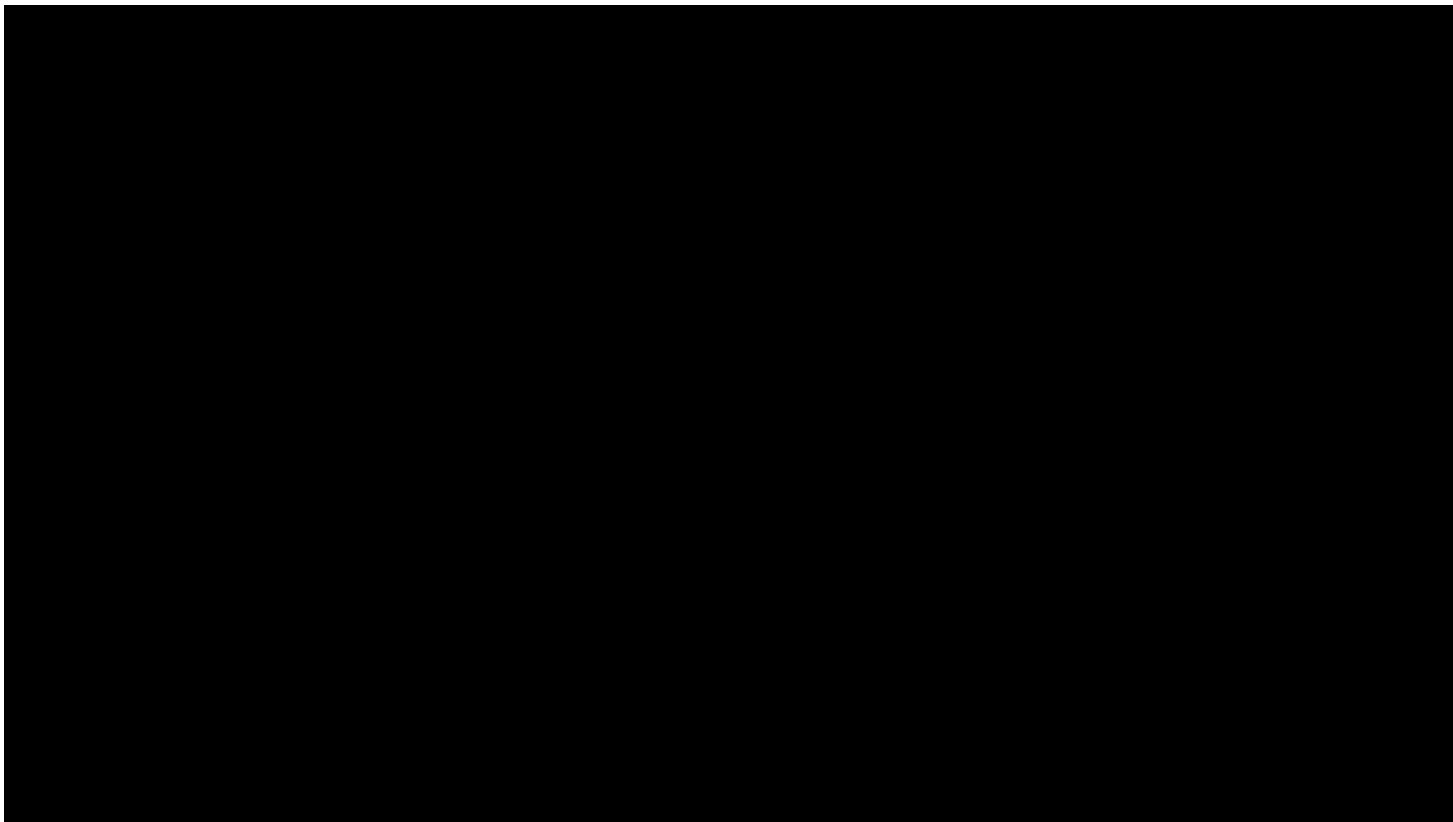CLOUD COMPUTING-BASED INFRASTRUCTURE, FOCUSED
ON THE NEEDS OF THE METEOROLOGICAL COMMUNITY
WEkEO
Copernicus
Europe's eyes on Earth
European
Commission

Usage and Demo - Sample GUI

Additional Demo I

Additional Demo II

Additional Demo III [Terminal]

We would like to thank Makis, Peter, Manuel & Victoras for their amazing mentorship and guidance as well as everyone from the wider ECMWF community who helped make this project possible.
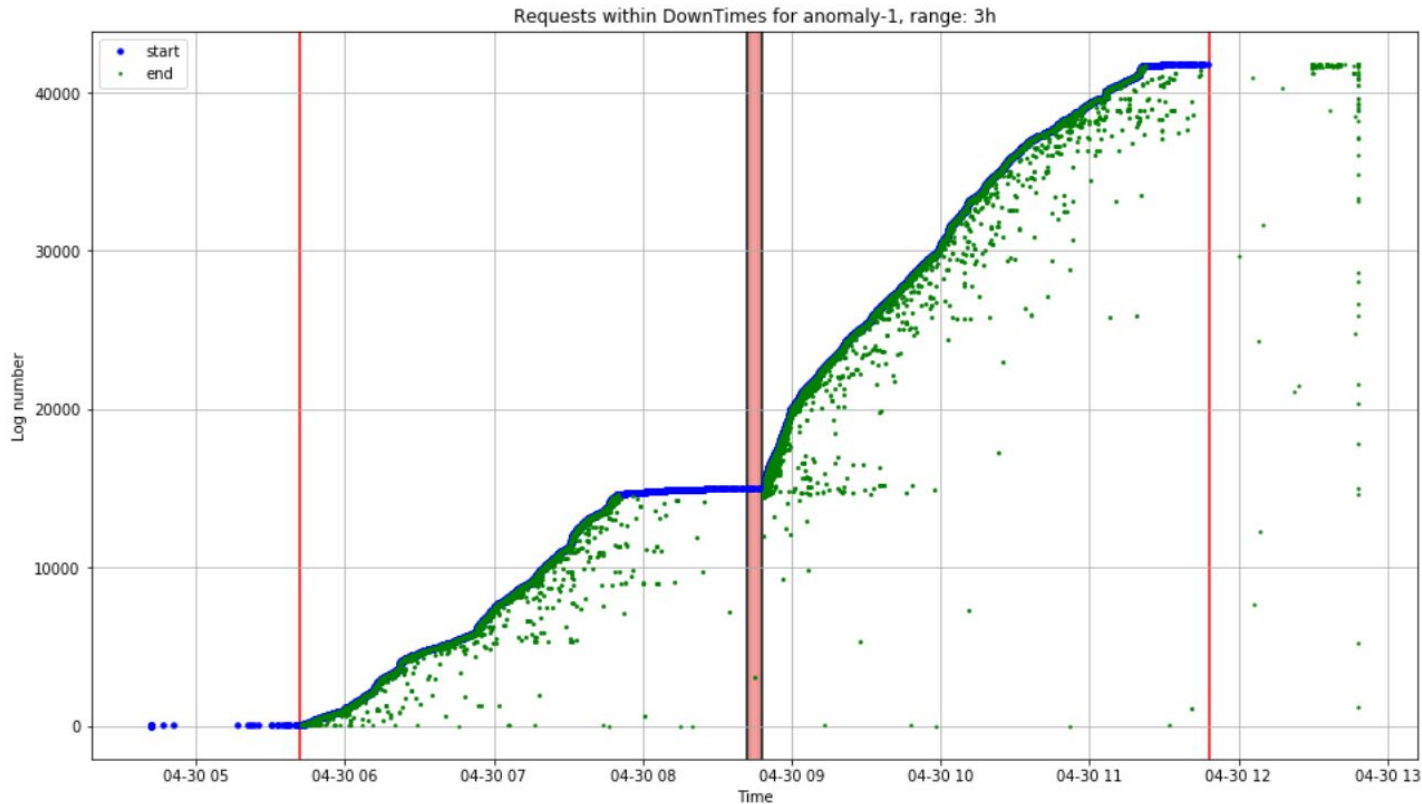
Thank You!
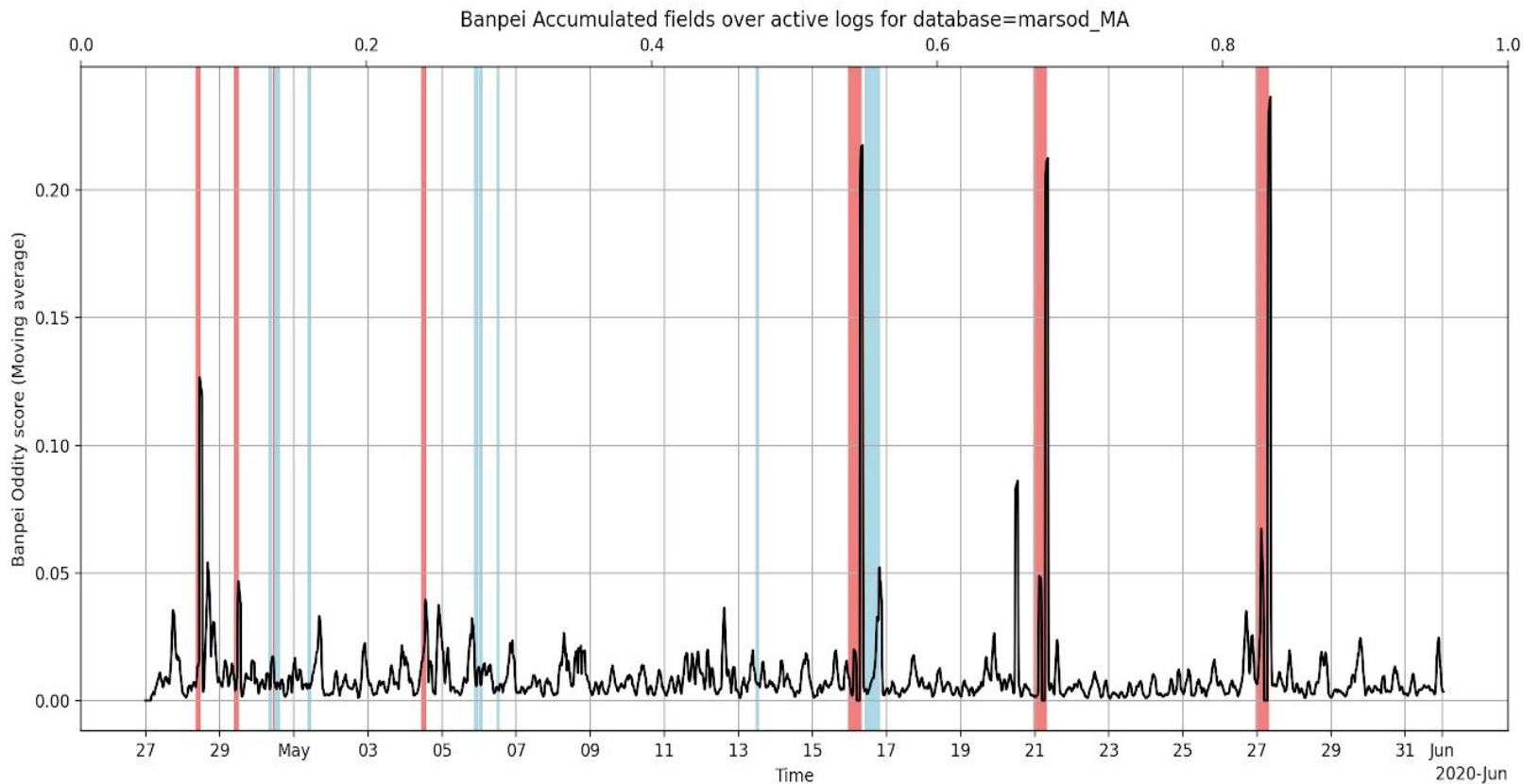# Questions?

Appendix - Going Deeper.
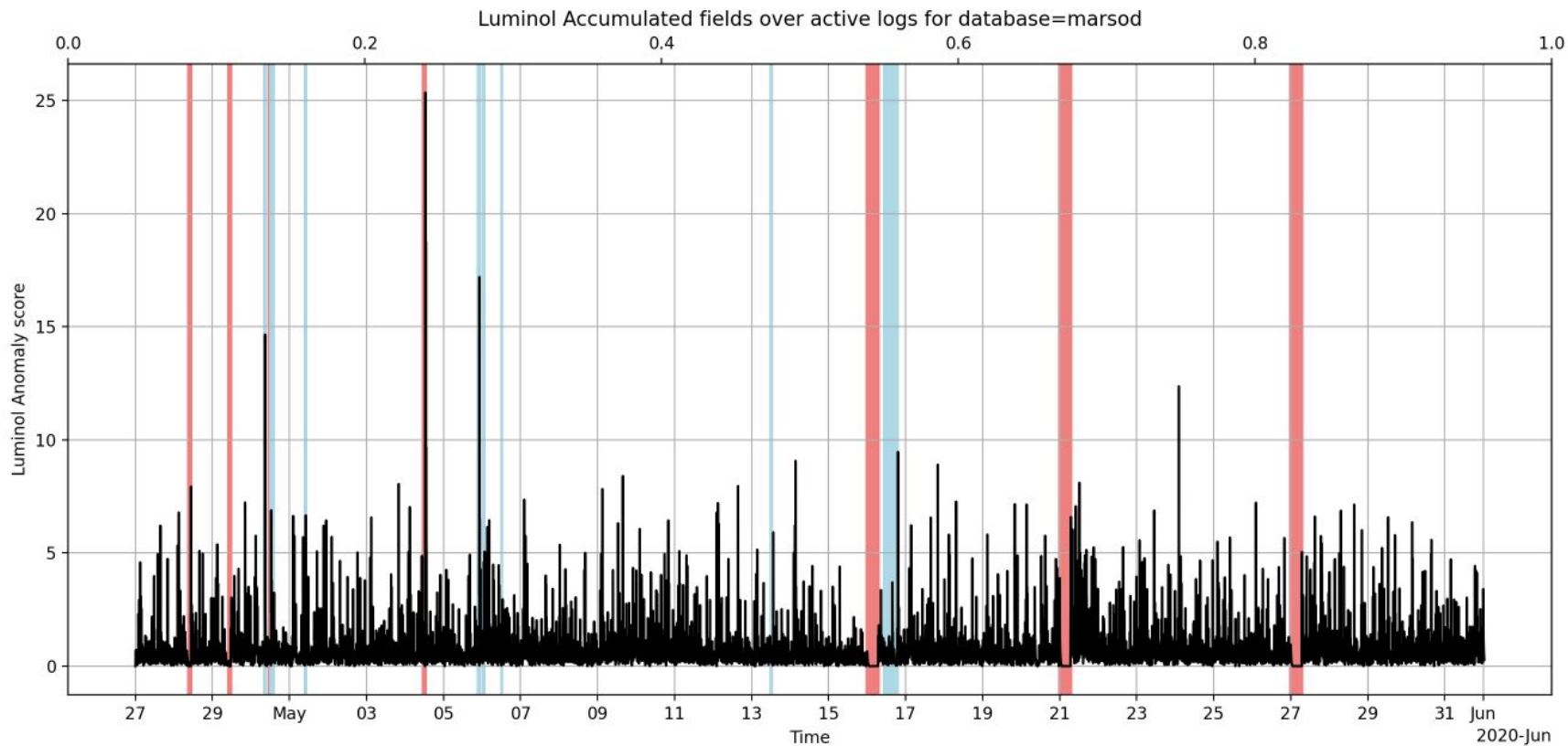
What Happens During a DownTime?

Requests within DownTimes for anomaly-1, range: 3h

Experiments - Banpei

Banpei Accumulated fields over active logs for database=marsod_MA

Experiments - Luminol

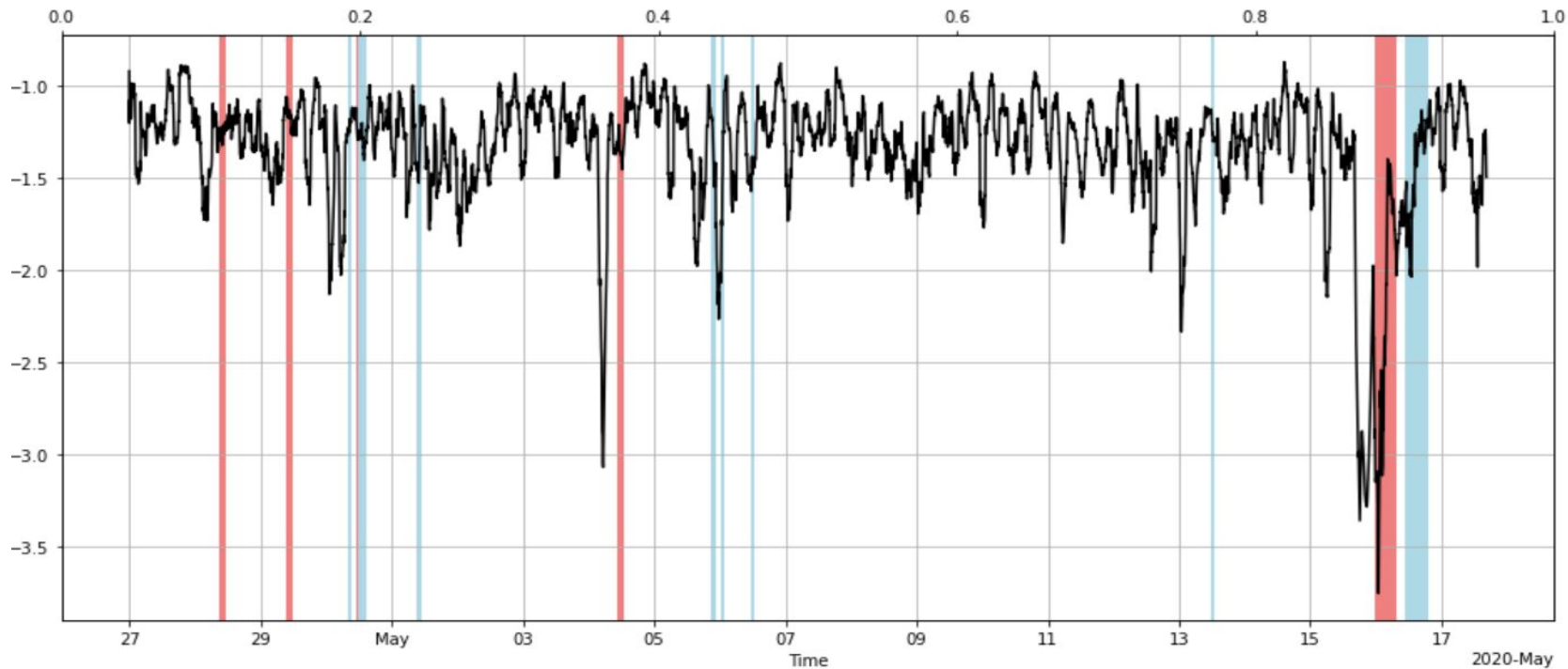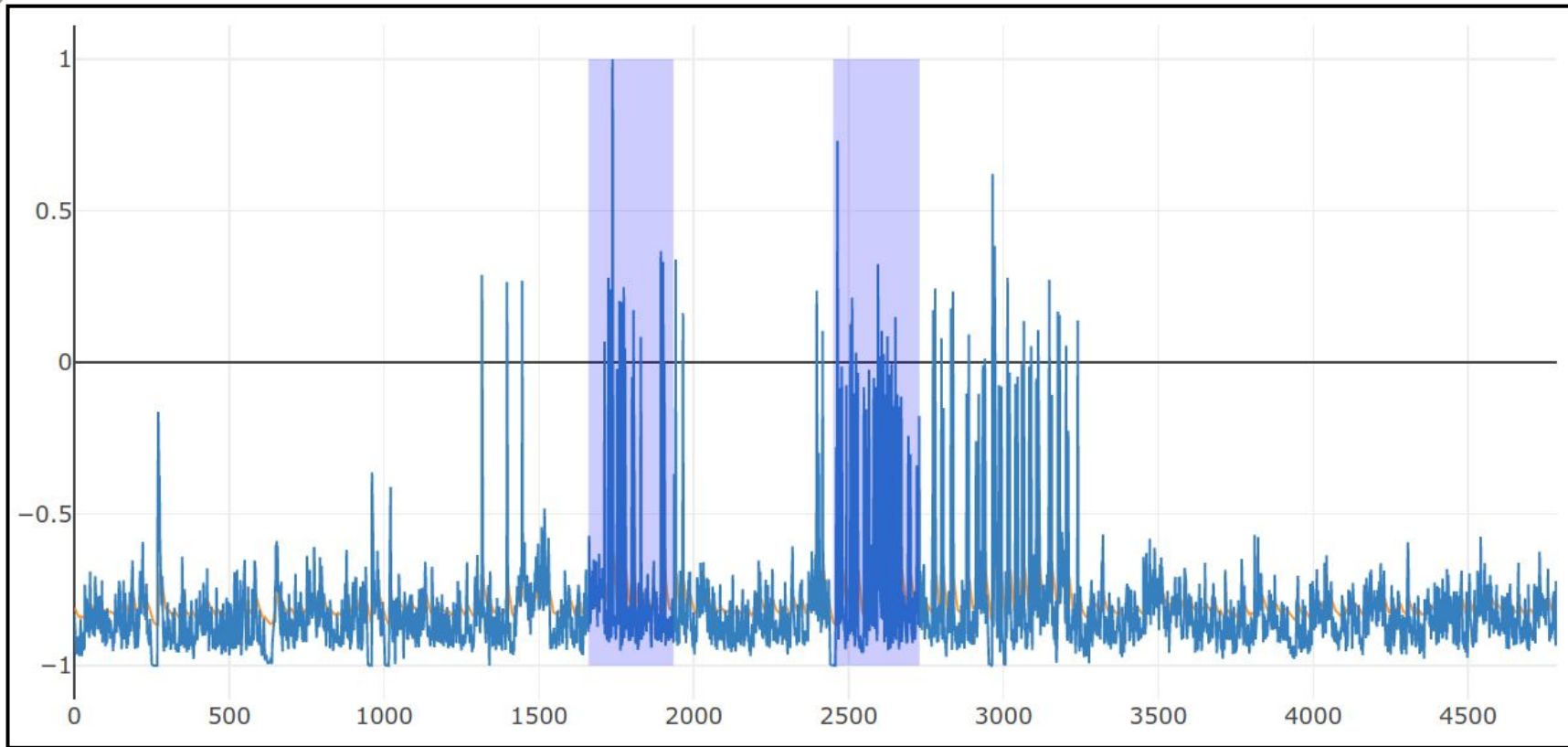Luminol Accumulated fields over active logs for database=marsod
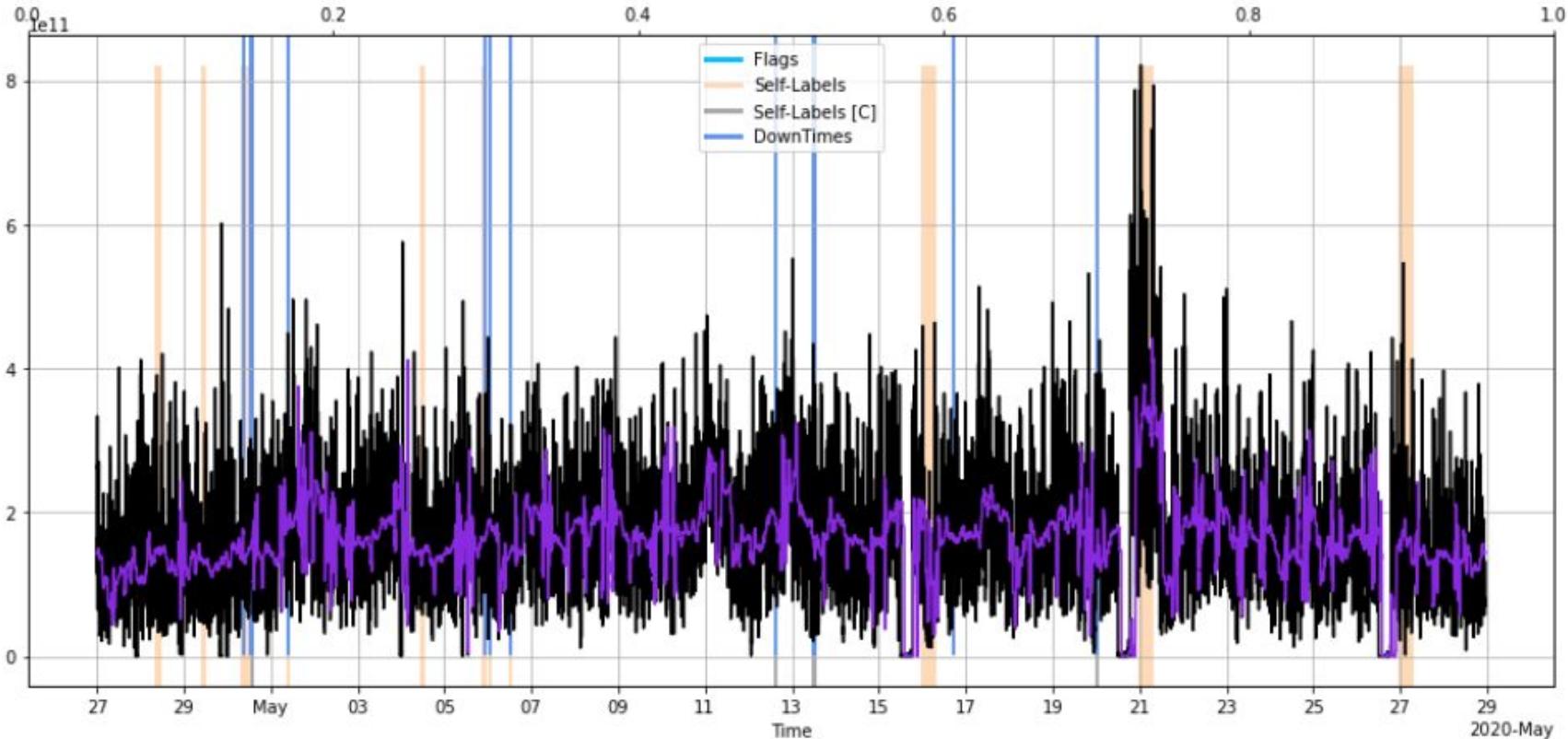
Experiments - Donut

Experiments - Telemanom

Experiments - ForecastX

Experiments - Clustering