Collaborative Anomaly Detection in ASKAP Monitoring Data: Integrating Machine Learning with Human Expertise

Zhuowei Wang   Ben Harwood   Vanessa Moss   Matt Austin   Marcus Hudson   Aidan Hotan

Space and Astronomy, Data 61, CSIRO

Abstract

We introduce a novel approach to anomaly detection in the ASKAP telescope's monitoring data. Our method combines machine learning with human expertise to effectively identify anomalies across ASKAP's subsystems. Initially, we apply the k-Nearest Neighbors (k-NN) algorithm for unsupervised anomaly detection. We then generate detailed heatmaps that visualize anomalies over different subsystems and timeframes. These visualizations are analyzed by domain experts to cross-reference anomalies between subsystems and accurately diagnose issues within ASKAP. By integrating human expertise, our approach addresses machine learning limitations like false positives, ensuring detected anomalies are statistically robust and contextually meaningful. This collaborative workflow enhances maintenance and optimization of ASKAP's performance. This research contributes to radio astronomy by demonstrating a method to manage and analyze the complex monitoring data from advanced telescope systems like ASKAP.

- We use anomaly detection by recognizing what's 'normal', we can identify the anomalies – the unusual occurrences that could lead to new discoveries.
- Our proposal: a collaborative human-machine approach. Machines process the data and identify anomalies, while humans interpret the results and guide the exploration.

Data from Different Subsystems in ASKAP

ASKAP incorporates several critical subsystems to facilitate its advanced radio astronomy capabilities, each contributing to data exploration and analysis:

- **Phased Array Feed (PAF):** Each ASKAP antenna features a checkerboard phased array feed with 188 active feed elements per antenna. These elements convert radio frequency signals into analog optical signals transmitted via dedicated optical fibers to the central control building, enabling detailed exploration of wide fields of view with high sensitivity.
- **Digitiser (DRX):** Inside the control building, signals from PAF elements undergo digitisation and are processed through oversampled polyphase filter banks. Selected channels are then sent via digital optical links to the beamformers, initiating the initial steps in data transformation and exploration.
- **Beamformer (BMF):** Beamformers compute weighted sums across PAF elements to generate dual-polarization beams. A fine filter bank subsequently divides channels into 18,144 fine channels for detailed data exploration, enhancing the resolution and clarity of observations.
- **Correlator:** The correlator processes fine channels to compute visibilities for each baseline, providing essential data for correlation analysis and image formation. It handles a wide range of fine channels and bandwidth, supporting comprehensive data exploration across different frequencies and resolutions.
- **Chiller (Cooling Infrastructure):** ASKAP's digital signal processing hardware, consuming 280 kW of power, is cooled by a system that circulates chilled water and utilizes a geothermal heat exchange system for efficient heat dissipation, ensuring optimal performance during data-intensive exploration tasks.

These subsystems collectively enable ASKAP to conduct cutting-edge radio astronomy research by facilitating robust data exploration and analysis across various stages of signal processing and observation.

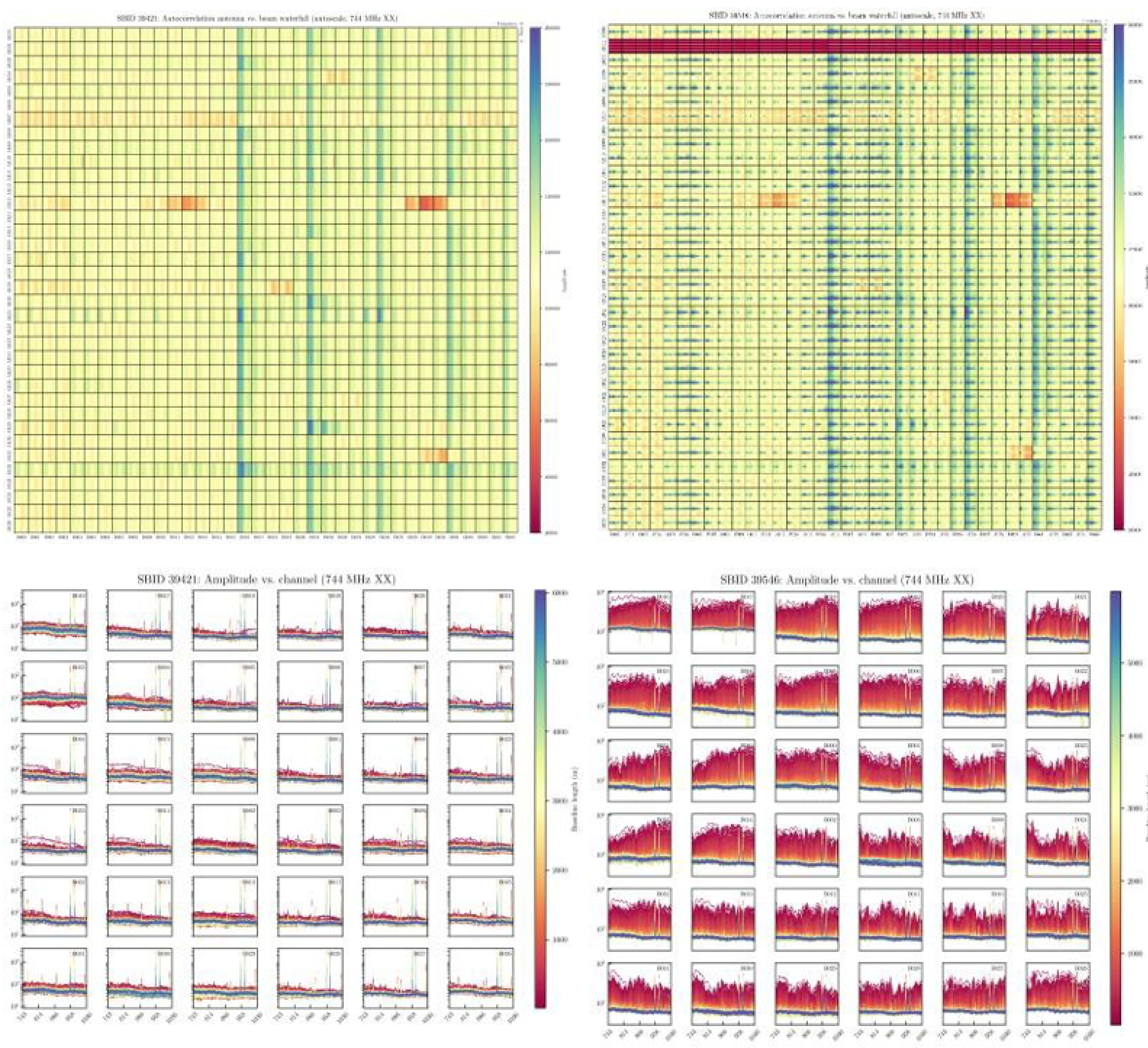

Figure 1:Diagnostic plot comparison, made routinely for every observation by Dr. Vanessa Moss's software.

Half The Solution: Anomaly Detection

In addressing the challenge posed by the 'data explosion', we turn to the potential of anomaly detection. Anomaly detection is the process of identifying patterns in the data that do not conform to an expected behavior, known as anomalies. These anomalies, essentially the 'unknown unknowns', could pave the way for new scientific discoveries or even highlight potential system errors. Thresholding method, trained to discern 'normal' patterns, can automate the detection of these anomalies, even amidst vast datasets.

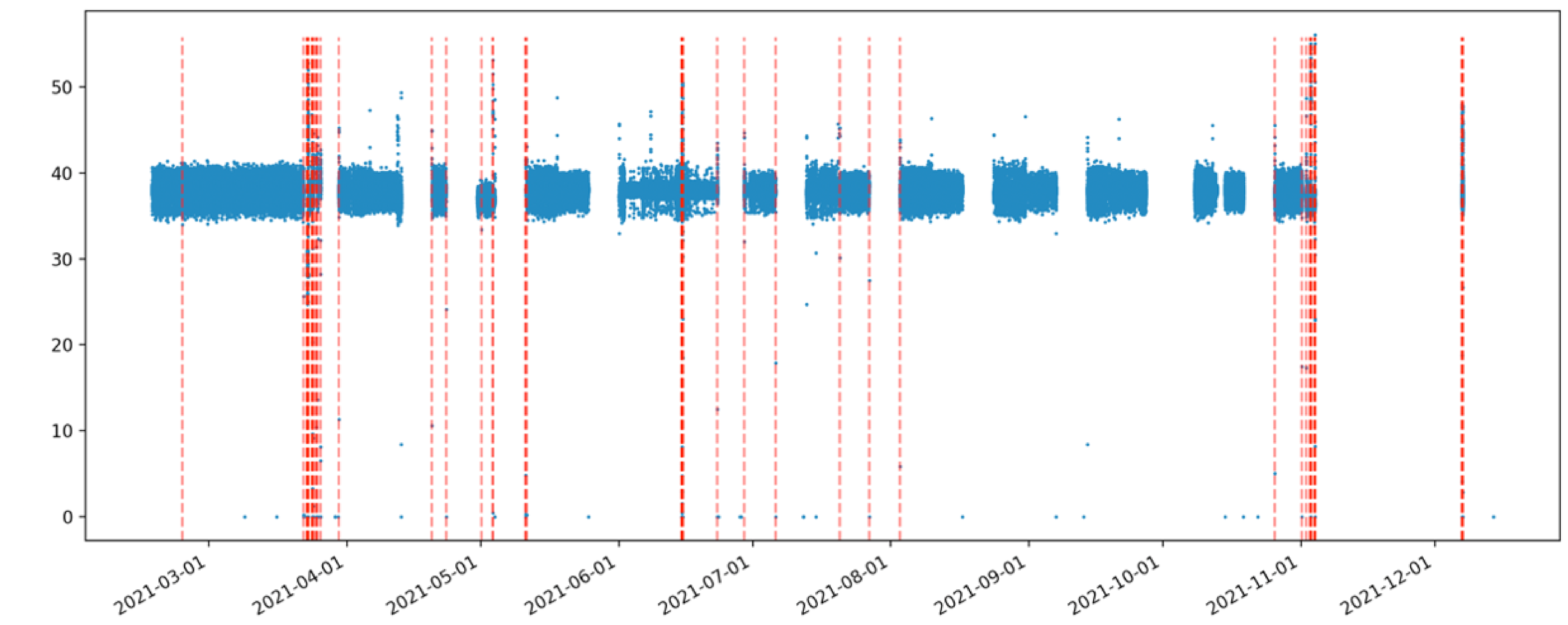

Figure 2:Anomalies detection in water cooling system of ASKAP. Red dotted lines represent the anomalies detected by thresholding algorithm.

Better Visualizatio using Heatmaps

We devised a systematic approach to summarize anomalies across ASKAP's different subsystems subsystems using heatmaps. Initially, we synchronized anomaly data from each subsystem based on a selected time frame. This synchronized data was then segmented into smaller time intervals, ranging from fine-grained 10-minute windows to broader hourly intervals, depending on the desired analysis granularity. To visualize anomaly patterns effectively, we employed heatmaps. These heatmaps provided a clear, color-coded representation of anomaly counts, with darker shades indicating higher frequencies of anomalies and lighter shades indicating fewer occurrences. Each row in the heatmap represented a subsystem, while columns corresponded to specific time intervals. This structured visualization approach enabled us to discern correlations and patterns among anomalies across subsystems and time periods. For instance, simultaneous spikes in anomaly counts across multiple subsystems within the same time window suggested potential systemic issues or external factors impacting multiple components concurrently. This methodological framework supports informed decision-making in anomaly detection, facilitating targeted interventions for optimizing ASKAP's operational reliability and performance.

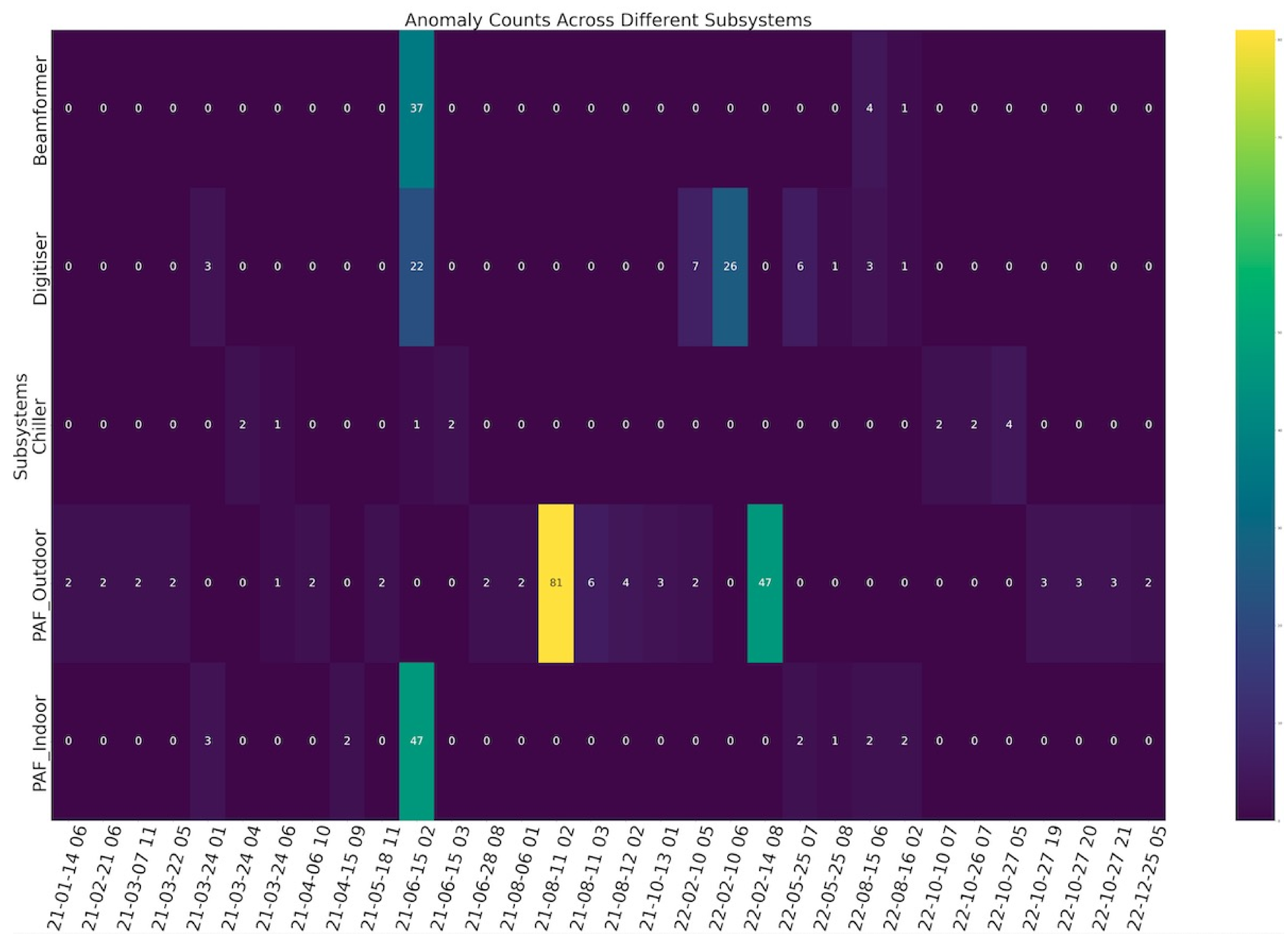

Figure 3:Heat map visualization of anomaly counts across five different subsystems. X axis represent the one-hour time window where anomalies are detected. The color intensity represents the number of anomalies detected, with darker shades indicating a higher frequency of anomalies. Notably, there are specific time stamps where a significant concentration of anomalies is observed, suggesting periods of heightened irregular activity.

Conclusion

We propose a novel collaborative human-machine approach to address the data explosion from ASKAP. By harnessing the computational prowess to process the vast datasets and detect anomalies, we allow human experts to focus on interpreting these anomalies, guided by their domain-specific knowledge and expertise. In doing so, we not only manage the surge of data efficiently but also heighten our potential for significant astronomical discoveries. This collaborative approach sets a promising trajectory for the future of astronomical research, providing us with robust strategies for tackling the challenges of data explosion and opening new avenues for unprecedented discoveries in our universe.