Automated Offset Detection in Global Positioning System Time Series Using Sliding-Window based Machine and Deep Learning Methods

Nikil Shyamsunder

Background/Research Question/Engineering Goal

- GPS data is disrupted by offsets
- Many of these offsets are recorded, but some are missed
- These offsets are laborious to account for, and if not properly done, can cause wild variances in the results of scientific studies that utilize GPS Data
- Critical in quantifying crustal stress changes
- Currently, there is a lack of effective automated methods to detect these offsets

Objective:
Develop efficient, autonomous, and accurate machine learning (ML) models to detect the occurrence of offsets given an input of time series data

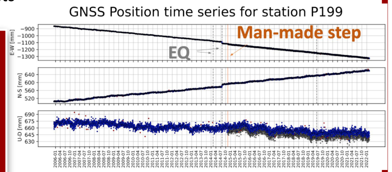


Figure 1. Example Time Series data showing both earthquakes and man-made offsets and how it affects the data in all three directions.

Methods & Materials

We used synthetic GPS data from King et al. and manually calculated the date of offsets to use as ground truth. A sliding window algorithm, inspired by the object detection technique, was developed to create a dataset of time series intervals in the North, East, and Up directions for all stations. The algorithm takes two parameters, the size of the window, and the time delta between windows. Then, a data preprocessing step was taken to account for data imbalance and scaling. A variety of time series classification methods were trained, and the final models were assessed with a 40% hold-out test dataset.

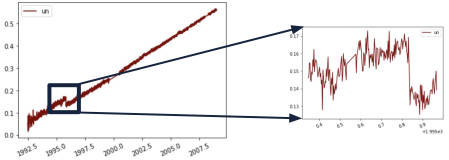
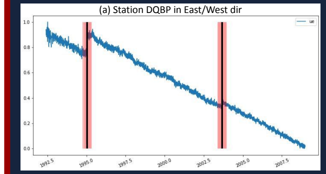
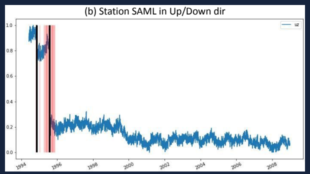


Figure 2. Plot on the left visualizes the North/South time-series data for Station AJGJ. Plot on the right pictures an example window created from a portion of that data containing an offset.

Visualization of final classifier performance





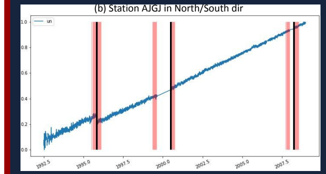


Figure 3. Visualizes how the trained RocketCLF performs on three different stations in three different directions: the vertical black line shows the ground truth offsets, and the red rectangle window shows the window in which the classifier detects an offset. (B) shows how the classifier can miss offsets and provide false negatives. (C) shows that false positives may occur, much of the time due to missing data.

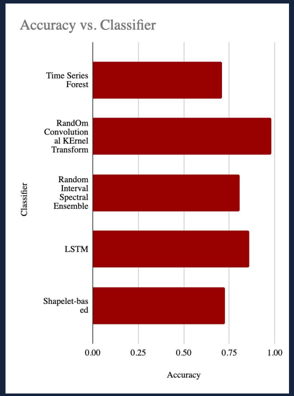


Figure 4. Compares the accuracy of 5 different time series classifiers.

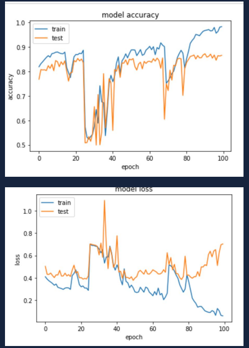


Figure 5. LSTM model accuracy and loss during training over 100 epochs. The model's best parameters were chosen using Kerastuner. The model reached an 87% accuracy on the test set before beginning to over fit on the train set, evidenced by the divergence between train and test loss past ~Epoch 80.

Results

- Synthetic data from a total of 50 stations over 16 years in all three spatial dimensions were included
- The sliding window algorithm with a window size of 200 and time delta of 10 in order trained the highest accuracy classifiers
- The RocketCLF model based on the combined dataset had the best performance with an accuracy and F1 scores of 0.98, 0.98, and 0.98 respectively
- Other models such as Time series forest, Shapelet-based, and Random Interval Spectral Ensemble failed to perform nearly as well
- Neural networks using LSTM and GRU units also failed to achieve performance similar to the RocketCLF

Time Δ	Window Size	Mean Accuracy	Mean F1 Score for Class 0	Mean F1 Score for Class 1
10	20	65.109	0.75	0.46
10	40	82.250	0.86	0.77
10	100	93.576	0.94	0.93
30	100	81.271	0.84	0.77
100	100	65.787	0.74	0.48
10	200	98.176	0.98	0.98

Future Directions for Research

- Investigation of the use of larger datasets in order to train more robust LSTMs
- Use of deep learning image classification on time series visualizations
- Examining the efficacy of the models on real GPS datasets
- Finding solutions for the false positives that arise from missing data, such as time series imputation
- Classifying between man-made and seismic-driven offsets
- Developing an algorithms to detect offsets and automatically account for those which are classified as man–made

References

Gazeaux, J., Williams, S., King, M., Bos, M., Dach, R., Deo, M., ... & Webb, F. H. (2013). Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment. *Journal of Geophysical Research: Solid Earth*, *118*(5), 2397-2407.
King, Matt, and Simon Williams. "Detection of Offsets in GPS Experiment (DOGEx)." *Newcastle University/National Oceanographic Centre, Liverpool* (2011).
Aubet, F. X., Zügner, D., & Gasthaus, J. (2021). Monte Carlo EM for Deep Time Series Anomaly Detection. *arXiv preprint arXiv:2112.14436*.