

Machine Learning for Quality Assurance of Lutz Catchment Runoff Data

Final Report

Gillian McGinnis, Project Manager

ADVISORS:

Dr. Sriram Iyengar (*The University of Arizona College of Medicine, Phoenix*),
Steven Paton (*Smithsonian Tropical Research Institute*)

Date: December 10, 2025

Contents

1	Introduction	2
	Background	6
	Failure Modes	6
2	Methods	7
	Data Wrangling	7
	XGBoost	7
3	Results	7
4	Discussion	8
5	Literature Review/Market Research	10
6	Research Project Deliverables	11
	Final Presentation Format	11
	What Analysis Is Being Run?	11
	What Accuracy Is Expected?	12
	What if the Analysis doesn't work?	12
7	Appendix	13
A.	My Title	13
	1) My subtitle	13
	2) My subtitle	13
B.	Ground Rules	14

Abstract

1 Introduction

The primary goal of this project is to create models that can automate quality assurance of temporal runoff data from the Lutz catchment of Barro Colorado Island (BCI), Republic of Panama (Figure 1). This will remove the need for manual data corrections, which not only use the valuable time and energy of researchers, but also have the potential to be imprecise or inconsistent. Rainfall and soil moisture measurement data will be integrated to produce a more well-informed model.

Runoff data has been continually collected from the weir since 1972, at first through hand-written records (which have since been digitized) in intervals ranging from three minutes to three hours, and electronically measured every five minutes since 1989. Quality issues in the runoff data come about due to environmental factors (such as partial blockages of the outflow channel), sensor failures, and calibration issues, to name a few. Currently, quality assurance is conducted manually using a customized program in Visual FoxPro (VFP) to allow for visual inspection and correction of the data. Not only does this approach run the risk of imprecise or inconsistent corrections, but the deprecated status of the language and interface leaves it vulnerable to inaccessibility. Past attempts at automated quality assurance have failed due to the failure modes' resilience against traditional stochastic methods.

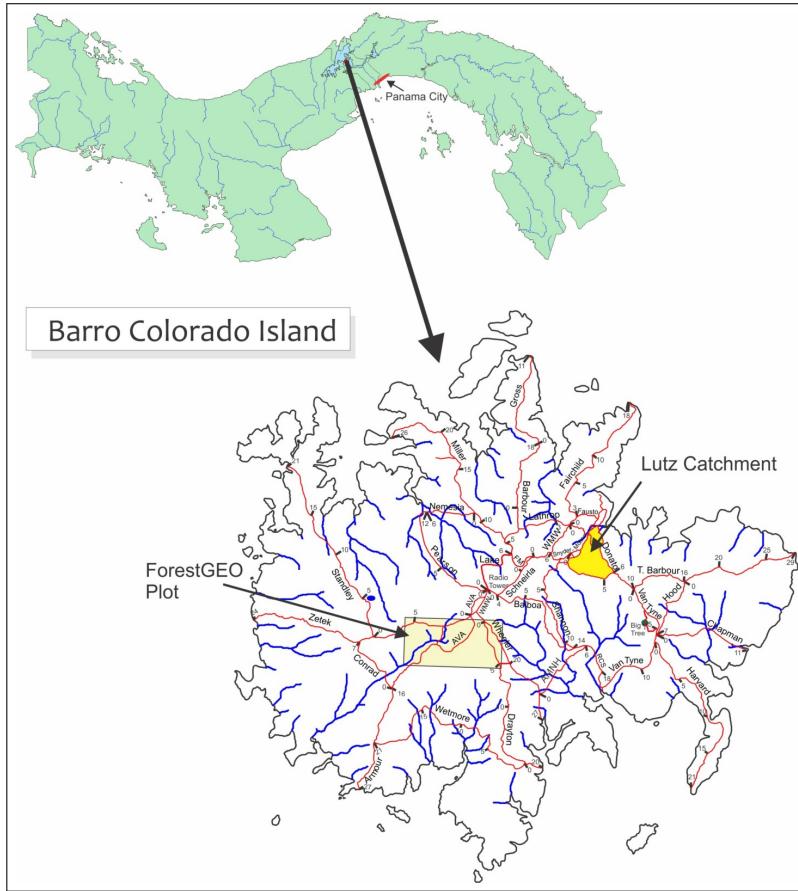


Figure 1: The Lutz Catchment on Barro Colorado Island in the Republic of Panama.¹

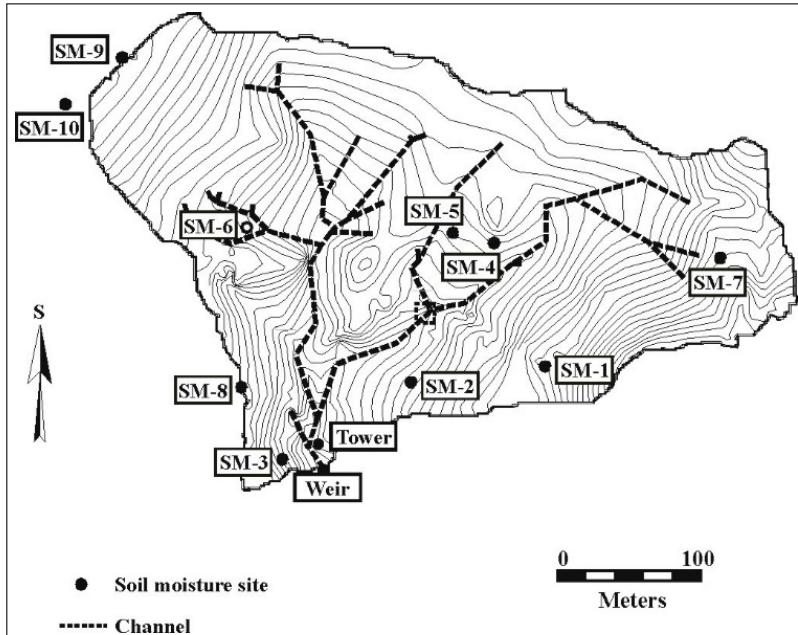


Figure 2: A fine-scale topographic map of the Lutz catchment showing the locations of the weir, tower, and ten soil moisture sampling sites.²



Figure 3: The Lutz Weir.³

Relevant data sets are available on the Smithsonian Research Data Repository with CC BY 4.0 licenses.⁴ Steven Paton—director of the Physical Monitoring Program at the Smithsonian Tropical Research Institute (STRI)—will be the primary contact regarding background information of the data, knowledge of exceptional data points, and any hydrological questions that come about. Project work will be conducted remotely, with regular advisor check-ins via teleconference.

The goals by the end of the project are to have a system for multiclass classification to flag periods of time with failure mode(s), and models to correct flagged data points. Measured success of the models can provide information into the feasibility of applying machine learning (ML) models in the context of micro-catchment runoff data quality assurance, as well as identify the most significant difficulties which can be addressed in future refinements of the models. To maintain a feasibility & realistic goal for the semester, it is anticipated that accurate correction models can be made for at least two failure modes.

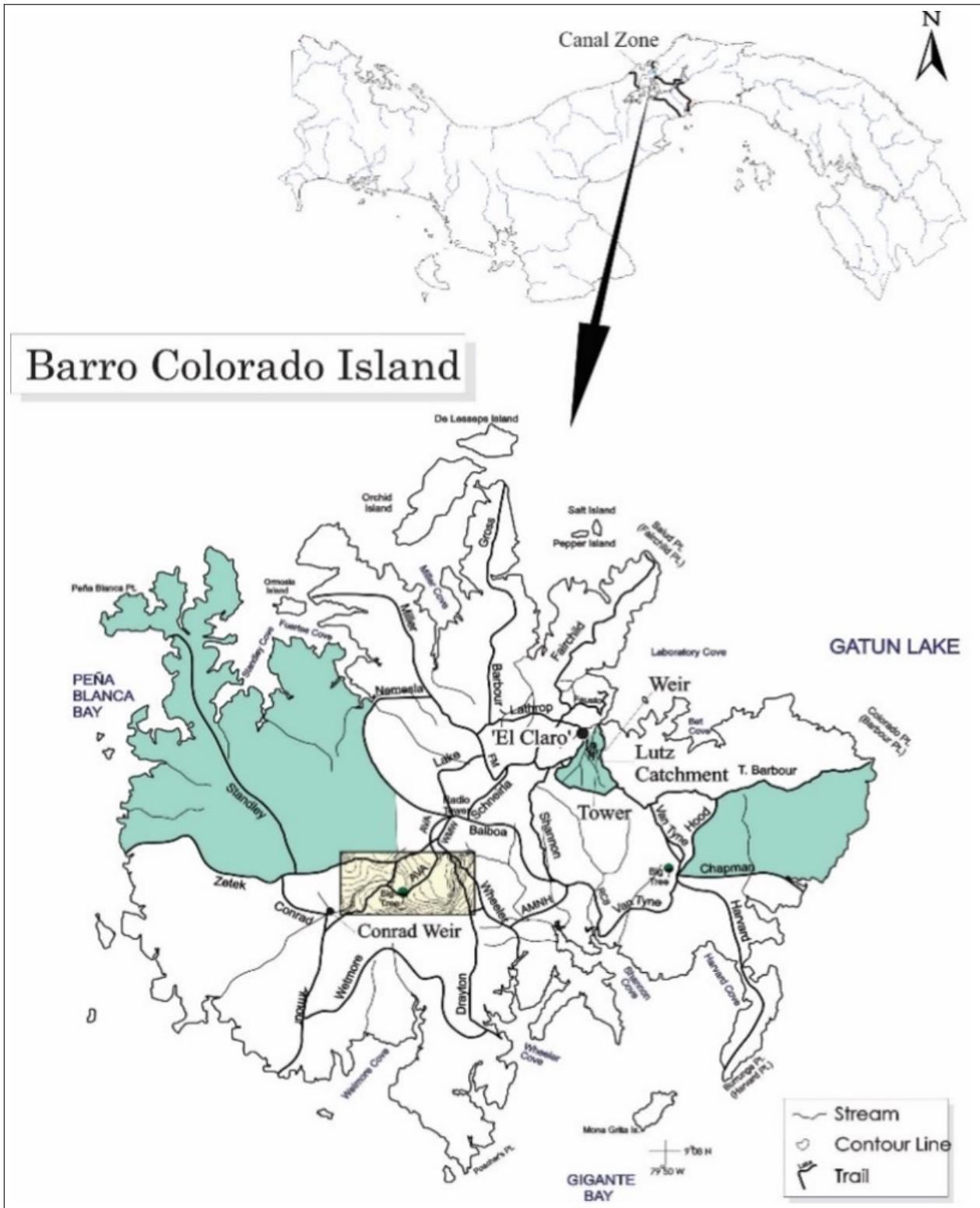


Figure 4: The location of BCI, Republic of Panama, showing the locations of the Lutz catchment, weir, and tower.⁵

1.1 Background

1.2 Failure Modes

2 Methods

2.1 Data Wrangling

Data cleanup information here.

2.2 XGBoost

XGBoost information here.

3 Results

Annotated code, data files, and project reports can be found in the project's GitHub repository.¹
Performance metrics information here.

¹Project repository: <https://github.com/gmcginnis/info-698-capstone>.

4 Discussion

Text.

5 Literature Review/Market Research

Overall research landscape

ML has been used in hydrology for applications such as predictive runoff modelling⁶ and modelling agricultural drought.⁷ Rainfall runoff data collected by citizen scientists in Ethiopia has been visually (via graphical inspections) and quantitatively (via statistical methods) analyzed for quality compared to nearby professional references.⁸ In the field of remote sensing and geospatial analysis, recurrent neural networks—especially long short-term memory (LSTM)—are successful for analyzing multi-temporal data, however scalability is limited and they can be prone to overfitting.⁹ In environmental ML, improper missing data management techniques can also worsen data leakage.¹⁰

The Lutz catchment (Figure 2) is a 9.73 ha area located in the secondary low land tropical rainforest on BCI, a 15 km² island located in Lake Gatun in the Republic of Panama.⁵ BCI is operated by STRI, a branch of the Smithsonian Institution.³ Runoff at the Lutz catchment—as well as other meteorological and hydrological data—have been continually monitored since 1972, making it one longest, continually monitored micro-catchment datasets for the neotropics.⁵

External variables of rainfall and soil moisture content are necessary for a well-informed model. In the context of modelling expected droughts, Houmma et al.⁷ highlights how incorporation of local and microclimate variables are becoming increasingly necessary due to climate change and “exceptional situations” which otherwise make extreme peaks of variables & interdependencies no longer able to be statistically modelable.

User insights

Paton currently uses Visual FoxPro to manipulate the data. While the interface has been sufficient for completing the tasks at hand, the software is vulnerable due to its depreciation. Furthermore, current quality assurance methods rely on subjective decisions by the individual analyzing the data, meaning that results are potentially biased to some unknown degree.⁵ Additionally, it is not currently set up to automatically detect problematic portions of the data. Because data monitoring is ongoing, new data must be regularly checked for occurrence of any type of failure mode.

According to Paton, there have been few—if any—attempts to apply ML algorithms to singular catchment runoff data quality assurance. Previously, stochastic methods were attempted for automation of the Lutz catchment data quality assurance, however the team concluded it was unsuccessful in being more accurate than Paton’s manual corrections. Although water flow is a complex phenomenon, it is anticipated that the quantity of raw and manually-corrected data as well as inclusion of additional variables will enable an algorithmic approach to correct the data to a reasonably-accurate degree.

6 Research Project Deliverables

Final Presentation Format

The final project will be submitted in the form of a GitHub repository containing code that is relevant, readable, and thoroughly annotated to maximize accessibility and reproducibility. In addition to code files, a written report with background information, methods/approach, and final model performance statistics & visualizations will also be created.

What Analysis Is Being Run?

Classification of Failure Mode

Periods of time in the data can be flagged as containing “failure modes”, as explained in Table 1. The system will determine where there are readings containing a particular failure mode. This also requires the system to determine the time range in which a failure mode is occurring. A challenge for this feature is that some failures occur in extremely short intervals of time (even as few as one reading), while others can spoil data spanning multiple weeks.

To flag data points, supervised classification can utilize rolling statistics and the external environmental variables (which provide some seasonal insights). Gradient boosted trees or LSTM can be used, but more literature review and exploratory data analysis will need to be conducted prior to finalizing which model(s) to utilize.

Failure Mode	Description	Complicating Factors
Calibration	Standard checks that require baseline correction or slight data pivot.	Recovery after blockage clearing can render calibration points ineffective.
Spike	Short and abrupt changes in level typically caused by equipment issues.	Extremely short-term issue that could be skipped over by random sampling.
Sub-zero	The stream runs dry, or the pond is being drained for cleaning.	Rain during a pond draining can render new data unrecoverable.
Signal noise	Equipment failure results in impossible variability of reported values.	Impossible to manually fix.
Blockage	Debris blocks the weir’s ‘V’, resulting in gradual increases in water level, and later abrupt decreases following blockage removal.	Rainfall events before, during, or shortly after a blockage can interrupt the base flow and decay curves.

Table 1: Overview of Weir Failure Modes²

Quality Assurance on Periods of Failure

Based on the classification result(s), additional models will attempt to adjust the raw outputs to the appropriate values had one of the above-mentioned failure modes not occurred. This will involve creation of multiple separate models, as the method in which the data is corrected differs based on the classification of failure type—for example, an interrupted drainage caused by a blockage in the weir can be fixed by the expected decay curve, while a spike is simply “flattened” to the level of adjacent data points using a simple linear regression model.

What Accuracy Is Expected?

Accuracy for certain failure mode classifications are expected to be high—for instance, “sub-zero” values are simple to flag. Others, such as “spikes”, are also expected to be easily found. Accuracy for multiple failure modes may be more difficult, especially in the dry season when there are non-erroneous fluctuations in flow caused by evapotranspiration. Multiple, large rainfall events also pose particular challenges, as well as overlapping failure modes (e.g., simultaneous blockage and calibration problems).

Similarly, correction models for simpler failure modes are expected to be highly accurate and precise, since they are often simple linear gap-fills. Corrections for the failure mode of a blockage in combination with a rain event may have reasonable accuracy but lower precision.

What if the Analysis doesn't work?

If the error type categorization has poor performance or is unable to consistently perform, and the needs of the correction models become more pressing as the semester continues, the data set does include the manually-flagged error types. The results of the ML-assisted corrections will be compared against the results of the manual corrections to data not included in the training sets. Large deviations from the expected values will be considered “failures”.⁵ If the correction models do not work, it will still provide valuable insights into which failure modes are most difficult to automate or most “resistant” to certain algorithms.

7 Appendix

A. My Title

1) My subtitle

Text

2) My subtitle

- The Faculty Advisor will provide knowledge and expertise to help the group stretch their skills.
- The Faculty Advisor will participate in a weekly or bi-weekly call/meeting with the Project Team to review the project status, upcoming deliverables, priorities, issues, and progress to the agreed Project Plan.
- The Faculty Advisor will provide document review, feedback and approval, rejection, approval with contingencies with adequate time for the Project Team to meet the course due dates.
- The Faculty Advisor will provide feedback to requested support requirements from the Project Team. This includes feedback and guidance on design implementations decisions, design files, test plans, test procedures and test results.
- The Faculty Advisor shall provide technical advice and guidance to the Project Team answering inquiries approximately 1 hour per week.
- Modifications to the Project Plan by the Project Team will be resolved and documented within 1 week of the request.
- Grade the finalized project using a skill-based rubric
- Attend iShowcase in May.

B. Ground Rules

Text.

References

- (1) Paton, S.; Larsen, M. C.; Stallard, R. The Lutz Creek watershed of Barro Colorado Island, Panama, 2021, <https://blogs.egu.eu/divisions/hs/2021/06/16/the-lutz-creek-watershed-of-barro-colorado-island-panama/> (accessed 11/18/2025).
- (2) Paton, S. BCI Hydrology Introduction, personal communication, 2025.
- (3) Smithsonian Tropical Research Institute Barro Colorado (Clearing, Lutz, Conrad weir), 2025, <https://striresearch.si.edu/physical-monitoring/barro-colorado/> (accessed 09/13/2025).
- (4) Smithsonian Institution Smithsonian Research Data Repository, <https://smithsonian.dataone.org/> (accessed 09/19/2025).
- (5) Paton, S. Runoff proposal suggestions, personal communication, 2025.
- (6) Mohammadi, B. A review on the applications of machine learning for runoff modeling. *Sustainable Water Resources Management* **2021**, 7, 98, DOI: [10.1007/s40899-021-00584-y](https://doi.org/10.1007/s40899-021-00584-y).
- (7) Houmma, I. H.; Mansouri, L. E.; Gadal, S.; Garba, M.; Hadria, R. Modelling agricultural drought: a review of latest advances in big data technologies. *Geomatics, Natural Hazards and Risk* **2022**, 13, 2737–2776, DOI: [10.1080/19475705.2022.2131471](https://doi.org/10.1080/19475705.2022.2131471).
- (8) Kebede Mengistie, G.; Demissie Wondimagegnehu, K.; Walker, D. W.; Tamiru Haile, A. Value of quality controlled citizen science data for rainfall-runoff characterization in a rapidly urbanizing catchment. *Journal of Hydrology* **2024**, 629, DOI: [10.1016/j.jhydrol.2024.130639](https://doi.org/10.1016/j.jhydrol.2024.130639).
- (9) Dritsas, E.; Trigka, M. Remote Sensing and Geospatial Analysis in the Big Data Era: A Survey. *Remote Sensing* **2025**, 17, 550, DOI: [10.3390/rs17030550](https://doi.org/10.3390/rs17030550).
- (10) Zhu, J.-J.; Yang, M.; Ren, Z. J. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environmental Science & Technology* **2023**, 57, PMID: 37384597, 17671–17689, DOI: [10.1021/acs.est.3c00026](https://doi.org/10.1021/acs.est.3c00026).