Leveraging IoT and Machine Learning for Improved Monitoring of Water Resources – A Case Study at Upper Ewaso Nyiro River (KENYA).

Jason KABI[1], Ciira MAINA[2]
[1, 2] Dedan Kimathi University of Technology, Department of Electrical and Electronic Engineering, Centre for Data Science and Artificial Intelligence. P.O. BOX - PRIVATE BAG – 10143, Dedan Kimathi - Nyeri, Kenya.
Email: jason.kabi@dkut.ac.ke, ciira.maina@dkut.ac.ke

DSAIL Technical Report

DSAIL-2021-001

August 2021

Contents

1. Abstract

This report outlines the development of an IoT sensor system (Project Ewaso), capable of water-level monitoring in rivers channels, with an aim of ensuring equitable distribution of water and sourcing of data to quantify unsustainable water usage and catchment area destruction. The paper also defines an application scenario in a specific hydrological region of the Ewaso Nyiro basin in Kenya, highlighting the characteristics of data collection and processing used. Fixed position node systems are described along with web-based data acquisition platform developments integrated with IoT techniques to retrieve data. The developed architecture utilizes the LoRaWAN - LoRa protocol to send real-time data packets from nodes deployed to a server that displays, decodes and stores the data. From the server, data can be transferred to a time series database, where it can be accessed and displayed through different customizable queries and graphical representation allowing future use in prediction machine learning systems. All these characteristics are presented along with evidence of the deployment of different devices and of the IoT network infrastructure.

2. Introduction

Competition for clean and safe water has, to a great extent, contributed to water management difficulties around the globe. Overexploitation of water resources is a considerable constraint on sustainable and safe agricultural development practices. Agriculture is one of the main economic activities practised by various communities around the world, hence water resources are an essential factor to the alleviation of poverty. There has been a recognition of water as an essential component of food security [1], with more attention being drawn on the significance of management of water by the United Nations Conference on Sustainable Development 2012, in an attempt to meet the sustainable development goals (SDGs).

The upper Ewaso Nyiro (Ngare Ngiro), found in the Ewaso Nyiro basin, is one of the major rivers in Kenya. The Mt Kenya and Aberdare regions, also found in Kenya and are the main contributors to the Ewaso Nyiro river, have for a long time been the focus areas for water resource management and conservation practises. However, in recent years, there has been an experience of water crises of unknown extent in lower catchments and other areas along the rivers path [2] [3].

The water crises have been brought about by the intensified agriculture, reduced rainfall due to climate change and catchment degradation. These crises, in turn, cause conflicts between water-user communities along the river Ewaso Nyiro path in the lower catchment areas [4].

To ensure equitable distribution and sustainable usage of water available in river channels, effective monitoring is essential. River catchment area degradation in recent years has been severe due to encroachment by people. It is altering run-off and infiltration rates, accelerating soil erosion and increasing sediment transport and deposition. In a quest to protect the catchment from encroachment, monitoring water level can be an important source of data which can be used to quantify the rate of catchment degradation [4].

This report describes the development of a water-level monitoring system for the Ewaso-Nyiro lower catchment that will be the first step in quantifying and discovering the justification statements stated.

3. Objectives

The main objectives of this work are:

1. To design a sensor system to monitor water-level in a river channel.
2. To deploy the LoRaWAN IoT network at Ol-Pejeta conservancy to facilitate data transmission.
3. To integrate the sensor system and the LoRaWAN Radio for long range low power data transmission.
4. To develop web infrastructure to visualize and store sensor data.
5. To utilize machine learning models/algorithms in performing anomaly detection on the water level data collected

4. Methodology

4.1 Development of the system

The IoT sensor system design is based on the Multitech mDot which is an Arm®[1] Mbed™ programmable LoRa module from Multitech[2] which is ideal for rapid prototyping. The modules are programmed using the mbed[3] platform (mbed cli – mbed command line programming tool) which allow development of software in C/C++ and provide drivers/libraries for the peripheral devices connected to the MCU. The design also incorporates a Maxbotix[4] MB1010 ultrasonic sensor for river water-level measurement and a custom PCB designed and etched to house all the components.

Data transmission relies on a LoRa network server able to decode LoRa data packets from fixed position nodes and relay them to a database for storage, awaiting processing. The network server we used is provided by The Things Network[5] (TTN). Through the utilization of a Python-MQTT (Message Queuing Telemetry Transport), data is transferred from TTN to an InfluxDB database located in a Google Cloud[7] (GCP) virtual machine instance (Compute Engine) for storage. From there, data is then visualized on a Plotly-Dash[6] web application https://water-monitoring-258811.wl.r.appspot.com Figure 3 shows a schematic diagram of the system.

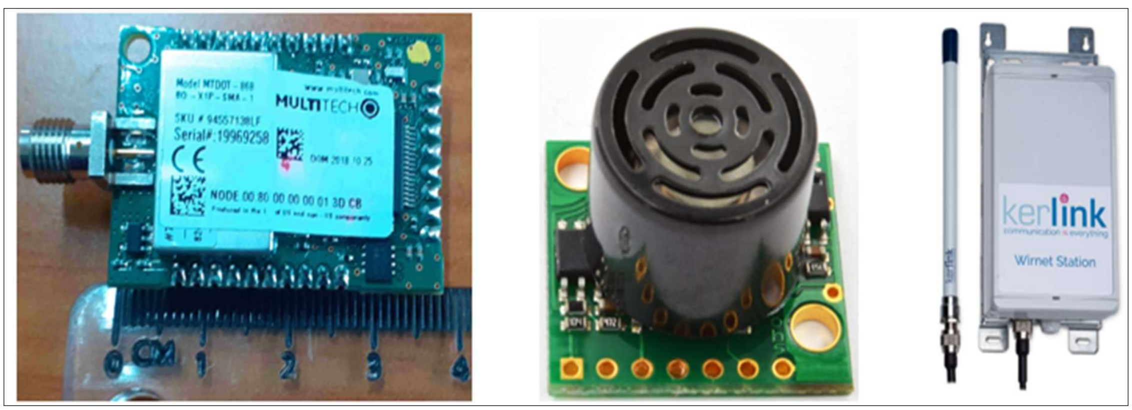

Figure 1: Multitech mDot (left), Maxbotix MB1010 ultrasonic sensor (middle), Kerlink outdoor LoRa gateway (right)

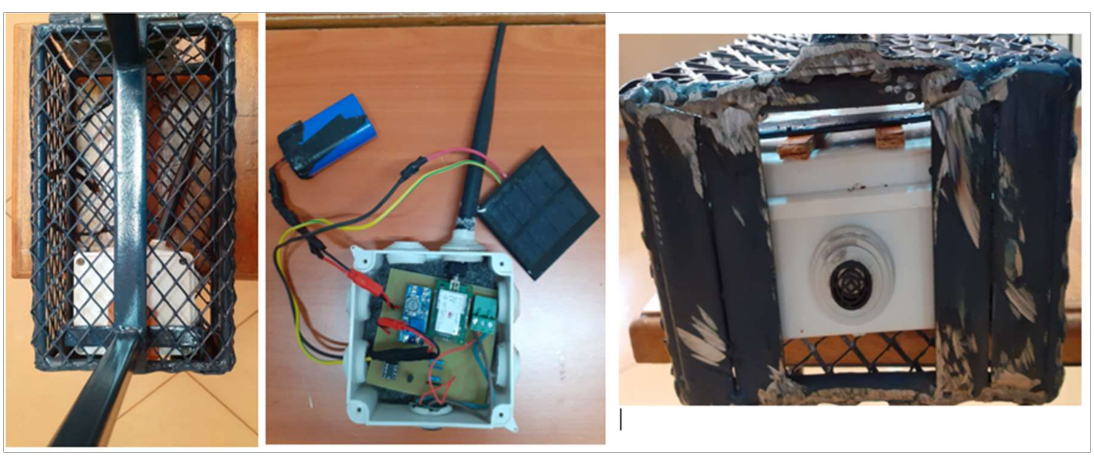

Figure 2: Prototype ready for deployment

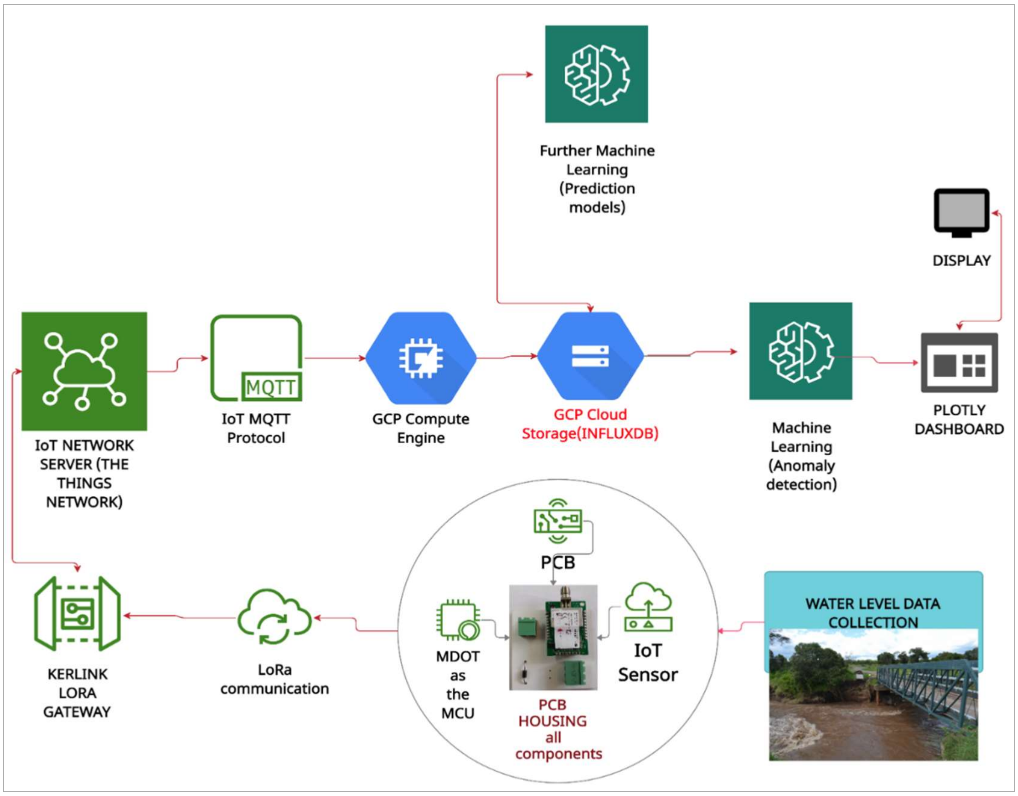

Figure 3: System schematic diagram

The system was deployed along River Ewaso Nyiro at Ol-Pejeta conservancy which is also home to a Wildlife Techlab which was setup to develop and test conservation technology. To avoid damage by primates in the conservancy, we designed a metallic cage, shown on Figure 2 and 4, to house and secure the prototype and we have been receiving river water-level data since June 2020.

Figure 4: Deployment (Prototype on the Left was destroyed by Baboons and the One on the Right has a primate proof cage)

1 https://developer.arm.com
2 https://www.multitech.com
3 https://www.mbed.com/en/
4 https://www.maxbotix.com/Ultrasonic_Sensors/MB1010.htm
5 https://www.thethingsnetwork.org/
6 https://plotly.com/dash/

5. Radio Mapping

The LoRaWAN gateway, also known as the concentrator, is used to relay data packets between the end devices (nodes) and the network server via the internet. It communicates over multi-channels with multi-spreading factors. With this technique, nodes communicate with the gateway using different channels and data-rates without pre-negotiation and enables the gateway to accommodate about 10000 end devices at a go. To facilitate data transmission at Ol-Pejeta, a Kerlink Outdoor gateway was configured and installed on a WIFI tower at a height of approximately 16 meters above the ground. The height provided adequate radio coverage of a large section of the upper Ewaso Nyiro River running through the conservancy.

The robust operation and efficient deployment of many IoT systems rely on the deployment of gateways and relays to ensure quality wireless coverage. Radio mapping aims to predict network coverage extent based on a small number of link measurements from sampled locations [6].

We conducted radio mapping at Ol-Pejeta conservancy[8] to determine transmission range of LoRa[9] enabled prototypes developed, and also to test the deployment along river Ewaso Nyiro within the conservancy. We mounted two LoRaWAN gateways at Ol-Pejeta house, a Kerlink Gateway at approximately 16 meters and LoRix One Gateway at approximately 13 meters.

We deployed 3 devices at various points within the conservancy. The devices and gateways were already connected to The Things Network and the network server was relaying radio propagation data to an InfluxDB[10] database for storage, awaiting processing.

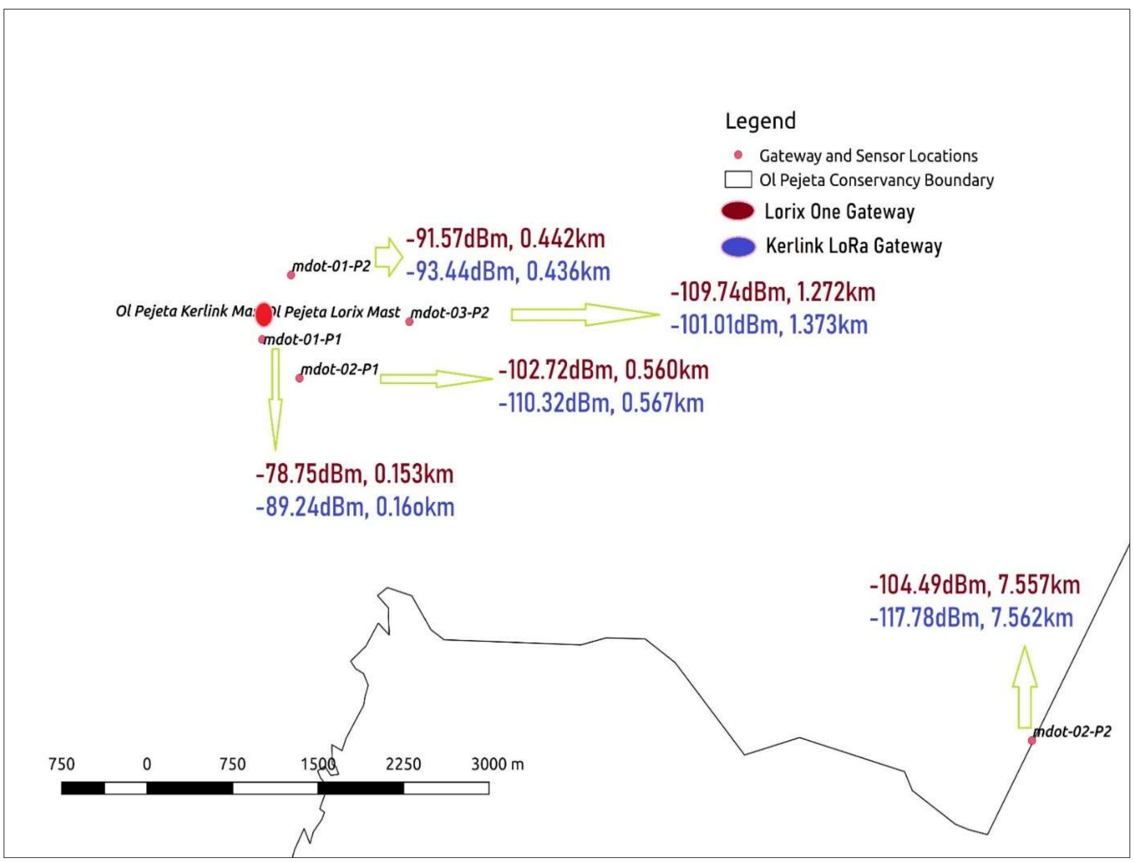

Figure 5: Map of Ol-Pejeta conservancy showing radio mapping test points

Table 1: The LoRix One and Kerlink Lora Gateways Specifications and Operations (subject to environmental factors and placement of nodes/sensors and gateways).

	LORIX ONE	KERLINK LORA
Antenna	Indoor 20cm Inclinable Antenna, 2dBi, 27dBm max output	Indoor 20cm Inclinable Antenna, 2dBi, 28dBm max output
Receiver Sensitivity	-140dBm	-141dBm
Operating Temperature	Min: -30 °C; Max: +55 °C	Min: -20 °C; Max: +55 °C
Communicating Range	Line of sight(*Antenna): +10kms Urban: up to 1km	Line of sight(*Antenna): +15kms Urban: up to 2kms
Installation	Wall or Pole mounting Metallic Strapping	Wall or Pole mounting Metallic Strapping

5.1 Radio Mapping Results

The Received Signal Strength Indication (RSSI).

This refers to the signal power that is received in mill watts (mW), and it is measured in dBm. How clear a receiver can "hear" from a sender can be measured using this value. Received signal strength indication, i.e., RSSI, is usually a negative value; hence, the signal is better when it is more positive (closer to 0). The value ranges of typical LoRa RSSI is -140 dBm to -30dBm.

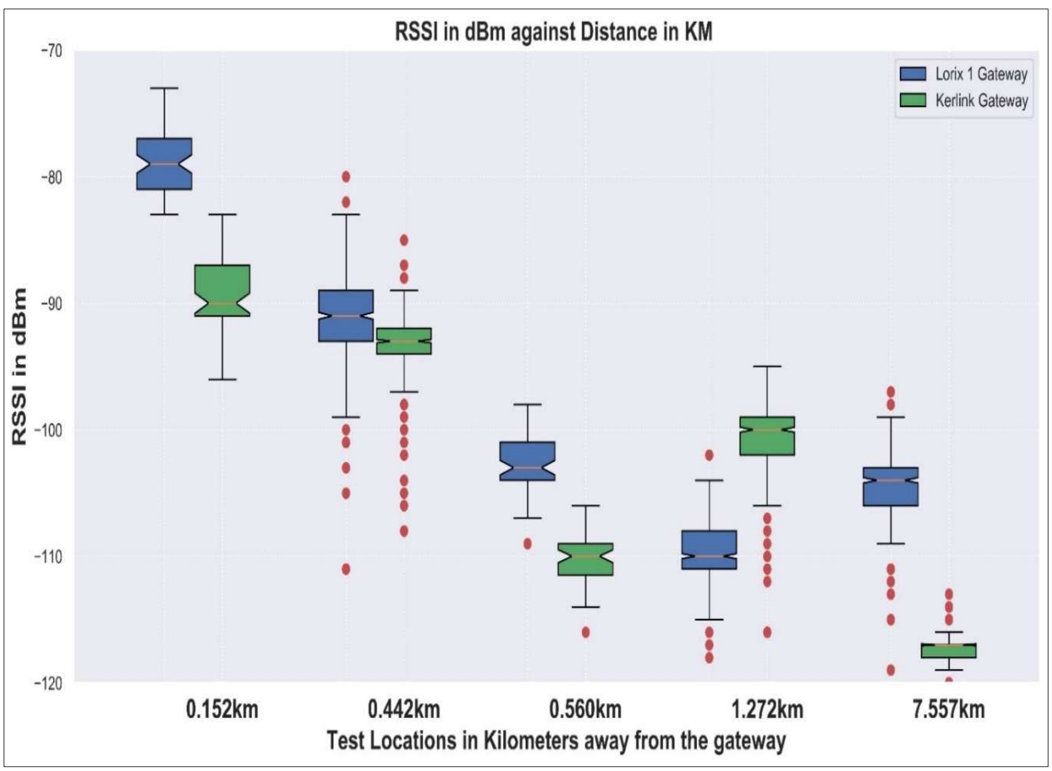

Figure 6: The Received Strength Plots for the 5 Test Locations.

For our radio mapping experiments, we computed the mean RSSI for each of the 5 test locations that were used. At approximately 150m away from the gateway, the LORIX One gateway outperformed the Kerlink gateway by a margin of 10dBm as well as at the furthest distance (approximately 7.5km) by 13dBm. The best strength was realized at the nearest test location and reduced as we moved to test locations far away from the gateway, as depicted by **Figure 6.** Plots in **Figure 6** provides a quick graphical examination of the RSSI for each of five (5) test locations for each gateway. Outlier RSSIs were highly realized in test location 2 and they are plotted as individual points, while none were realized at location 1 (nearest to the gateways). However, location 1 depicts the highest notable degree of dispersion (spread) and skewness for both gateways. There is a general non-linear variation of the median positions of the RSSI, usually determined by various parameters, which include free space loss, shadowing, reflection and transmission, diffraction, among others.

6. Sensor Calibration

Figure 7: MB1010 Ultrasonic sensor location

The MB1010 was calibrated with the help of the MB1010 datasheet[7]. With 2.5V-5.5V power, the MB1010 provides very short to long range detection and ranging in a very small package. The MB1010 detects objects from 0 inches to 254 inches (6.45 meters) and provides sonar information from 6 inches out to 254 inches with an inch resolution. The output formats included are **pulse width output, analog voltage output** and **RS232 serial output**. According to the datasheet, the analog pin (**pin 3-AN**) outputs analog voltage with a scaling factor of (Vcc/512) per inch.

We utilized the Analog Voltage Output as the interface output format in Project Ewaso. In this format, distance / detection range is converted to a voltage signal that can be read by a microcontroller and calibrated later. The MB1010 Pin3: AN - outputs an analog voltage signal with a scaling factor of (V_{CC}/512) V per Inch. A supply of 5V yields ~ 9.8mV/inch and a 3.3V yields ~ 6.4mV/inch. The output is buffered and corresponds to the most recent range data**.** **Example**: Assuming the payload (HEX: 16E1) is in hexadecimal form sent to TTN as an **unsigned 16-bit integer** so we have to convert it to decimal form to get (DEC: 5857). The system runs on a V_{CC} of about 3.7V hence, if 3.3V yields 6.4mV/Inch, 3.7V should yield 7.2mV/Inch or (2.834mV/Cm).

(1) According to the datasheet: Voltage per Inch(V/In) $= \frac{VCC}{512} = \frac{3.7(\text{System Vcc value})}{512} = 7.2\text{mV/in}$

(2) Voltage per inch - conversion - to voltage per cm(V/cm) $= (1 \text{ inch} = 2.54\text{cm}) = \frac{7.2mV / In}{2.54} = 2.8\text{mV/cm}$

(3) Converting the payload back to the analog pin voltage $= \frac{5857}{2^{16}-1} \times 3.0V = 0.26811V$

(i) $2^{16}-1 =$ Unsigned 16 bit integer limit

(ii) $3.0\text{V} =$ Multitech Mdot analog input pin voltage limit

(4) payload to distance (cm) conversion $= \frac{\text{Payload to voltage}}{\text{Voltage per centimeters}} = \frac{0.26811V}{2.8mV / cm} = 95.75cm$

(5) Hence 16E1(HEX) represents or is equal to $= 95.75$cm

We also carried out a preliminary experiment to verify the outlined MB1010 detection range and also to establish the precision and accuracy of the sensor. In the experiment, we measured several sets of distances using the sensor and produced plots to showcase the results.

	Trials	50 cm	100 cm	150 cm	200 cm	300 cm	400 cm	500 cm	600 cm	645 cm
0	Trial 1	51.11	100.66	149.88	202.29	302.96	400.45	501.86	603.40	646.90
1	Trial 2	48.89	100.98	148.93	199.11	302.64	401.72	501.95	603.08	645.31
2	Trial 3	51.11	101.30	149.88	202.29	303.27	401.72	503.68	603.08	647.22
3	Trial 4	50.16	102.25	150.83	199.11	300.05	400.39	503.36	605.62	646.58
4	Trial 5	49.21	100.03	149.56	202.29	303.27	401.40	502.09	604.99	647.53

Figure 8: Measured values for every test distance

Tabulated values on Figure 8 provides a quick graphical examination of the accuracy of the sensor for each actual distance measured. The precision and accuracy were found to be high for the short distances but a little lower for the longer measurements. Accuracy is basically how close the measured value is to the actual distance coincided to what is stated on the datasheet. The small deviation from the actual measurement could have also been brought about by the calibration process and human error. Precision is the measure of reproducibility – how close the measured values are to the actual value - at long distances, the sensor precision was impressive as shown on Figure 8.The longest distance measured was 647.53cm. River Ewaso Nyiro channel, at the point of deployment, is 4.9 meters, hence, the channel depth was within the sensor range.

7. PCB Design

Since the components needed could not be soldered directly onto the Multitech mDot pins, we developed an etched circuit to harbour all the components and also facilitate deployment. The circuit design was developed using the KICAD PCB development and design software. The circuit consisted of the battery/power socket, a socket to harbour the Multitech mDot and an ultrasonic sensor socket. The mDot is powered by a lithium 4400mAH rechargeable battery which is charged by a 150mAh Solar panel through a diode. The analog pin of the MB1010 is connected to one of the analog pins on the mDot to facilitate data collection.

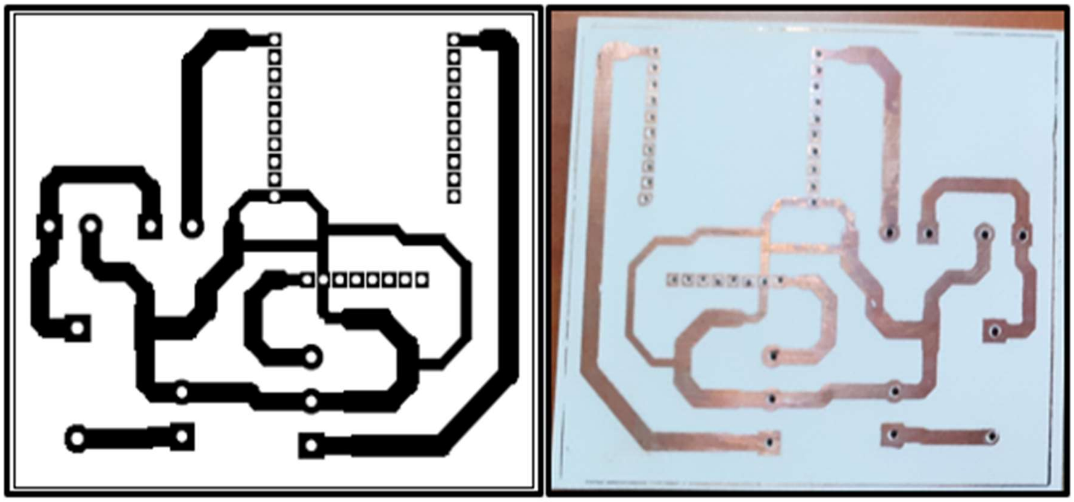

Figure 9: (Left) PCB design & (Right) Finished product

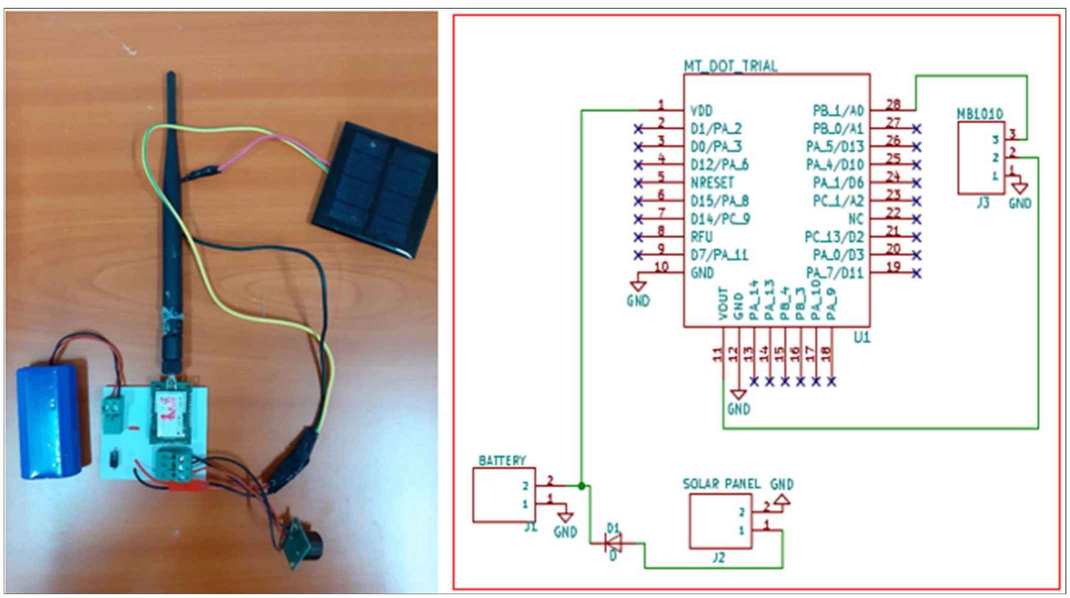

Figure 10: (LEFT) Complete system, (RIGHT) Circuit diagram of the system

8. Power Analysis/Management Sensors

To conserve the power in the battery, we took advantage of the power management scheme available in mbed OS which allows devices to be placed in a low power sleep mode between data acquisitions. In the setup, we acquire data at 30 minute intervals and put the device to sleep in between these sampling times. In particular, the board draws 0.2mA when in sleep mode and 68mA when taking measurements and transmitting data. The data acquisition period lasts for between 1-3 seconds. Also, to analyse (keep track of) the amount of power being generated by the solar, we developed **Project Ewaso Prototype 2** shown on Figure 11.

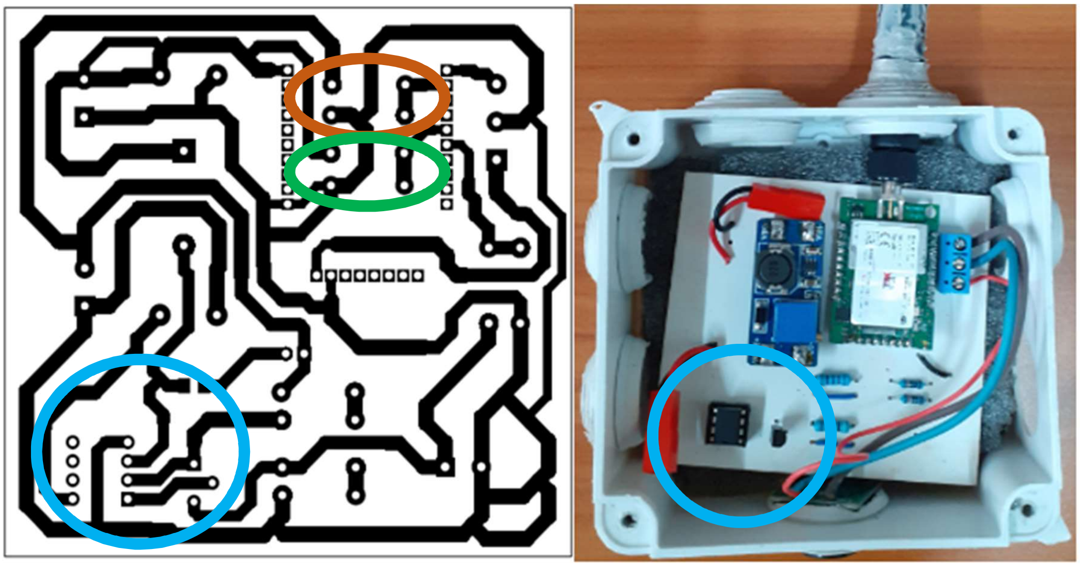

Figure 11: Power Analysis prototype

Apart from the main (data collection circuit), the prototype also included:

Solar Voltage Sensor: This was basically a voltage divider circuit that could lower the solar voltage to a level that could be fed to the Multitech mDot analog pin for conversion and measurement.

Solar Current Sensor: The solar current sensor was based on a current sense resistor and a differential amplifier. The amplifier amplifies the voltage drop across the resistor, which is dependent on the amount of current passing through the resistor. The voltage signal produced by the amplifier is therefore directly proportional to the amount of current from the solar panel through the resistor. The current sensor circuit is as shown on Figure 12. Additional design information and circuit details can be found at: https://www.allaboutcircuits.com/technical-articles/how-to-monitor-current-with-an-op-amp-a-bjt-and-three-resistors/.

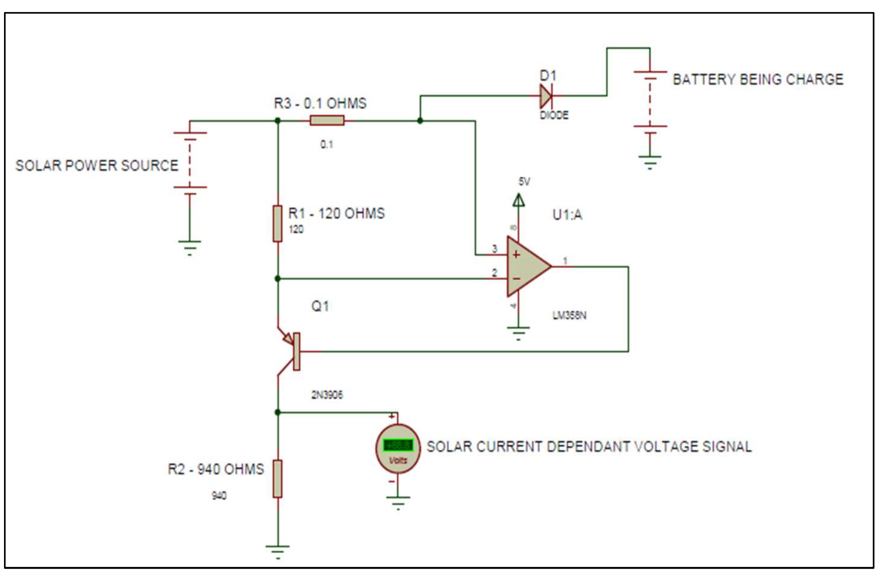

Figure 12: Solar current sensor

Battery voltage sensor: This was also a voltage divider circuit that could lower the battery voltage to a level that could be fed to the Multitech mDot analog pin for conversion and measurement.

Using the data collected by the solar sensors we were able to measure and keep track of the amount of power coming in from the solar as shown by Figure 13. The battery voltage sensor data plot helps in keeping track of the amount of power available in the 4400mAh Lithium ion battery (Figure 14). The data also served as evidence that the charging system was operating as desired.

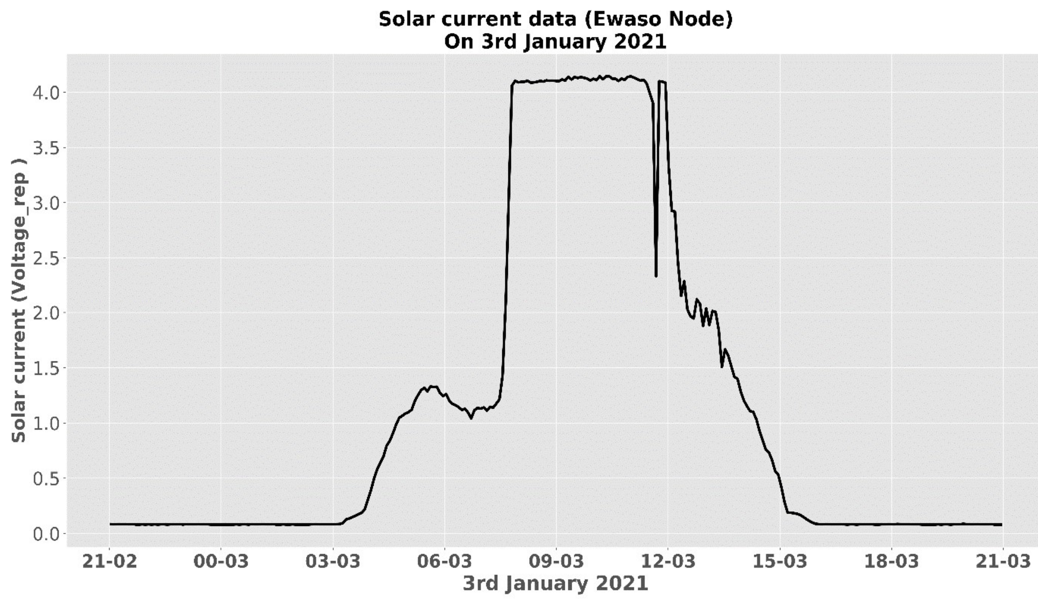

Figure 13: Solar current plot on 3rd January 2021

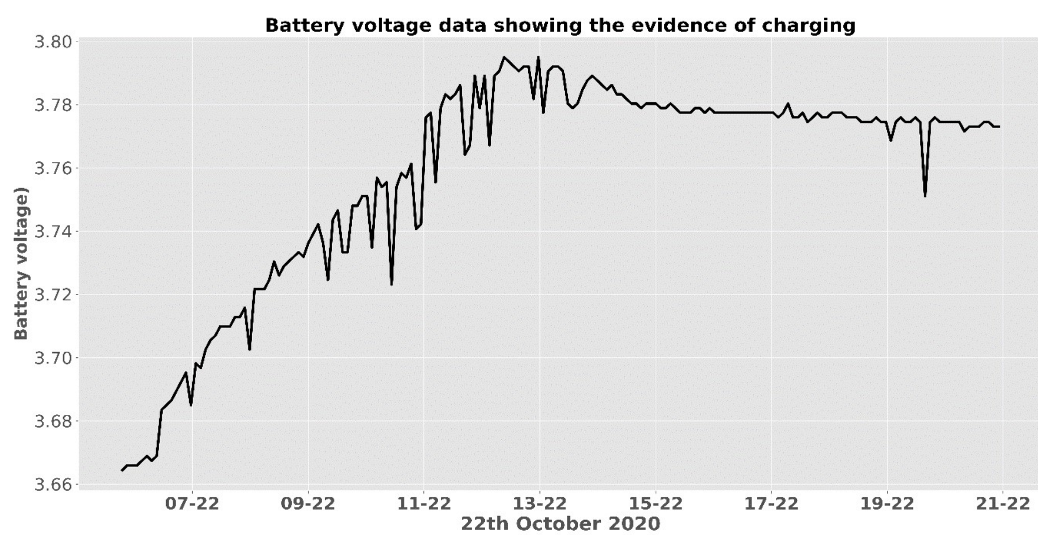

Figure 14: Battery voltage data showing the evidence of charging

8.1 Evidence of charging

Using the battery voltage sensor on PROTOTYPE 2 (Figure 11), we were able to discover some evidence of charging of the lithium ion battery. On 22th October 2020, the battery voltage level was 3.66V at 9.00am (EAT) and after 9hrs of being charged by the solar panel installed, the battery voltage level went up to 3.79V at 6.00pm(EAT) as shown on Figure 14.

8.2 Comparing the Solar current data to the short- wave radiation.

We acquired access to the short radiation data from TAHMO (Trans-African Hydro-Meteorological Observatory) and managed to compare it to our solar current data. The similarity was striking as shown on Figure 13 and Figure 15.

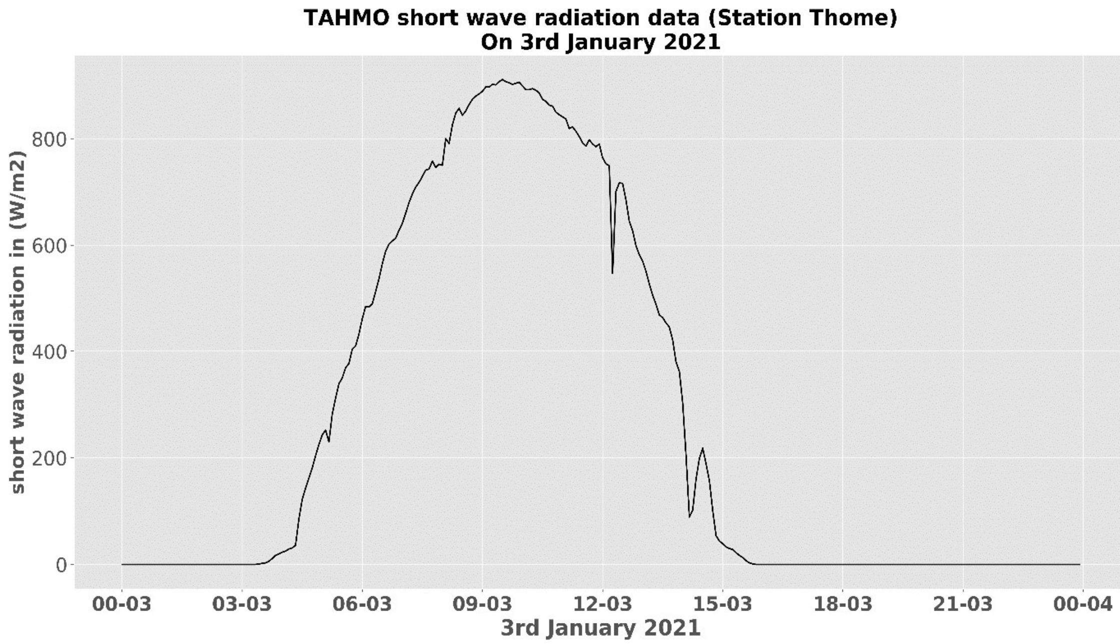

Figure 15: TAHMO short wave radiation data on 3rd January 2021

9. Programming the Multitech mDot

Mbed enabled modules such as the Multitech mDot used in this work can be programmed in various ways including using an online program compiler and the offline mbed command line interface (CLI). I used the CLI to compile the program used. Code used in this can be accessed at https://github.com/ciiram/mdot-maji. The program to collect power analysis data (solar current, solar voltage and battery voltage) can be accessed at https://github.com/ciiram/mdot-maji. The power analysis parameters were sampled at intervals of 30 minutes, just like the water level data, and relayed to the network servers and database. The data collected provided a clear picture of how the system was utilizing the power available and how much was coming in from the solar panel.

10. Results and Discussion

10.1 Machine Learning (Anomaly Detection)

Sensor prototypes deployed under potentially harsh weather conditions for tasks like environmental forecasting are prone to breakage and damage. The high probability of erroneous readings or data corruption during transmission brings up the problem of ensuring quality of the data collected by sensor. Since WSNs (wireless sensor networks) have to operate continuously and therefore generate very large volumes of data every day, the data quality process has to be automated, scalable and fast enough to be applicable to streaming data. The most common approach to ensure quality of data is anomaly detection. It consists of automatic detection of erroneous readings or anomalous behaviour of sensors. From a high level and in generic way, anomaly detection can be done by three main ways

By predictive confidence approach: this method involves using historical data to train a model that can predict the value of the next measurement, if the actual measurement is far from the predicted it is labelled anomalous. We carried out some preliminary experiments using FacebookProphet (at its core it is an additive regression model) and we were able to get rid of the anomalies in the data stream. As we proceed with the project, we aim to implement a neural network predictive model in the anomaly detection phase. For building a predictive anomaly detection model, popular time series modelling like ARIMA, SARIMA, VAR or any regression or machine learning and deep learning based algorithm like LSTM can also be used effectively.

The clustering based unsupervised approach: this method involves clustering the data using various clustering techniques like K-means and DBSCAN. If new measurements are assigned to small clusters very far from the big cluster centroid they are labelled as anomalies. We used this approach to detect and eliminate anomalies from our time series, water level data. KMeans is able to cluster correct data points and anomalies in different clusters and with a few python operations we eliminate the anomalies.

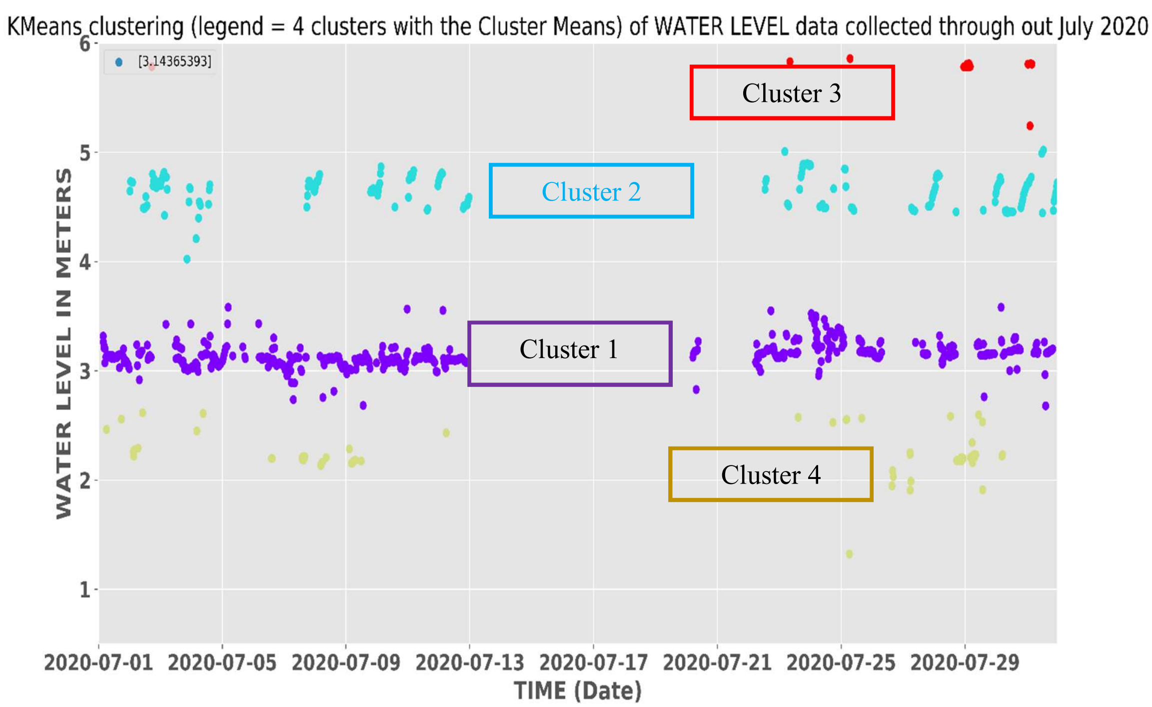

Figure 16: How the data is clustered by the KMeans clustering algorithm

Since **Purple Cluster 1 on figure 16** (cluster 1 between the 3 and 4 meter mark) has the accurate and highest number of data points it is retained by executing simple Python operations[7] and the points (anomalies), in the other clusters, Cluster 2, Cluster 3 and Cluster 4 are scrapped off the dataset.

Statistical profiling approach: this is done by calculating the statistical profile of the historical data and using a standard deviation to come up with a band of statistical values which can define the boundaries and anything falling outside these boundaries is regarded as anomalous. Following the preliminary experiment using IQR (interquartile range) analysis, we aim to test and implement other statistical based methods in anomaly detection.

10.2 Anomaly Detection Results

We used anomaly detection to generate a data frame of daily water level figures. Figure 17 shows the daily water level figure from 12th May 2020 to 9th January 2021. The plot on Figure 17 is also available on a visualization dash web application to display the water level data from the database. (DASH web app

Link: (https://water-monitoring-258811.wl.r.appspot.com).

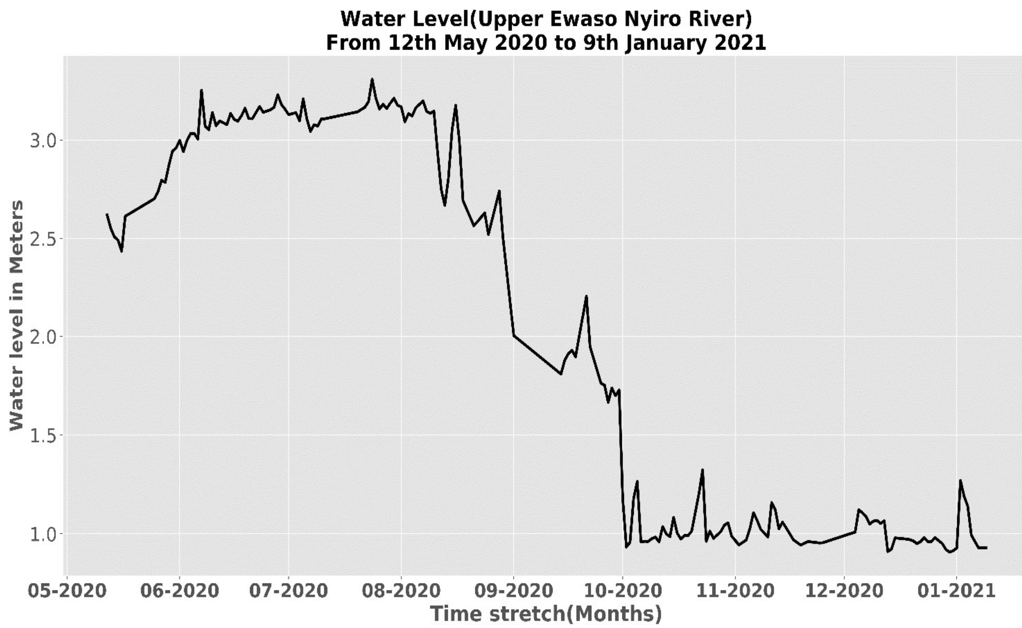

Figure 17: water level plot

11. Catchment Analysis.

Analysis of a river catchment is an efficient method to process our understanding of how varying climatic conditions and catchment parameters/characteristics (vegetation, soils, and topography) interact to define hydrological response. The hydrological response and status of catchment can be quantified by means of specific signatures of catchment behaviours like water flow rate, run-off co-efficient, water level, water capacity in the channels and flow duration curve. Quantifying the mentioned signatures is the main way to accomplish catchment analysis and classification [8] [9]. To analyse the catchment, we decide to check how long a spike in water level takes to appear after a spike in rainfall occurs. According to the test part we considered on Figure 18, the spike in water level occurred two days after a spike in rainfall. This shows the catchment area is possibly not degraded to a large extent. If the spike in water level was picked up by the sensor hours after the spike in rainfall, it would have meant that the catchment is possibly degraded to a large extent and the water runoff had nothing to obstruct it. In the future we aim to develop prediction machine learning models for the water level profile with the help of the rainfall data from TAHMO. By predicting and collecting water level data various aspects of the river and the basin can be quantified.

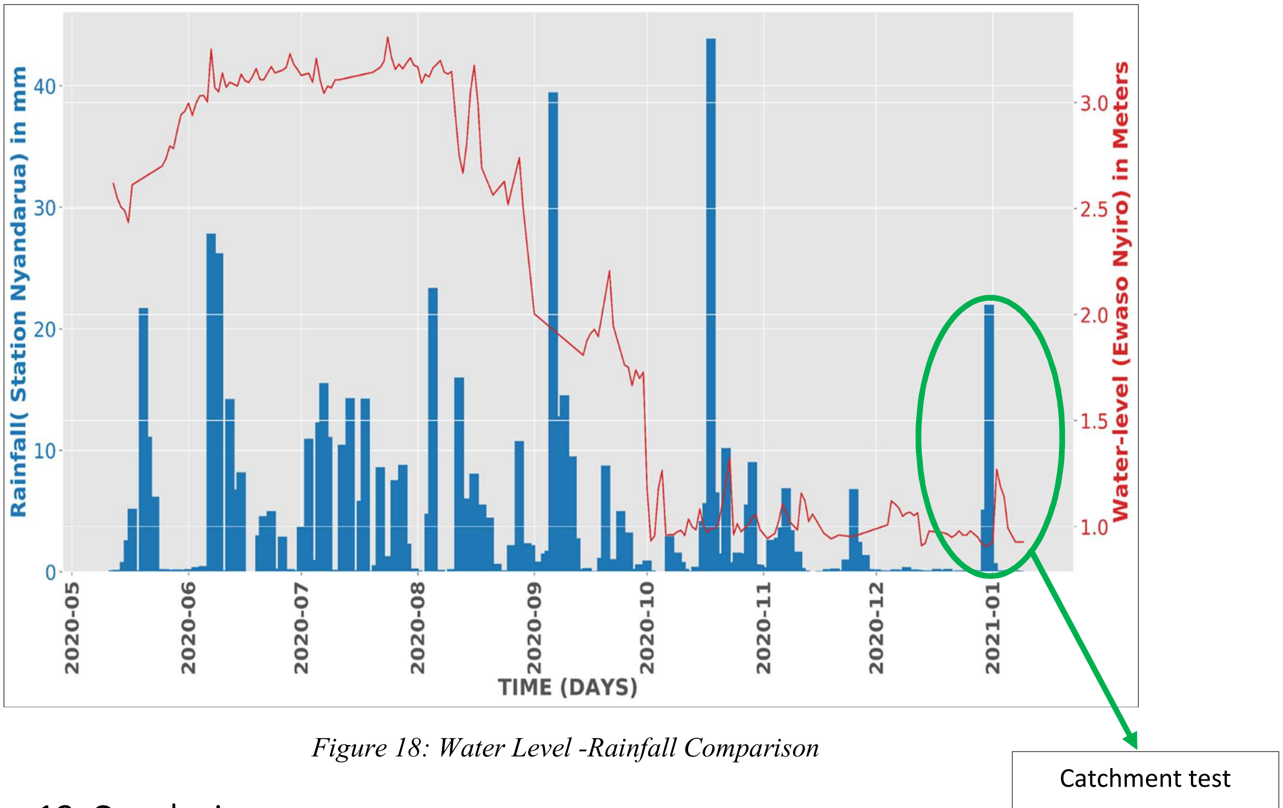

Figure 18: Water Level -Rainfall Comparison

12. Conclusion

This report has outlined the real-time monitoring of water-level in rivers by leveraging Internet of Things (IoT) and Machine Learning. We collected useful data from the prototypes deployed and in turn we were able to track water levels in the River Ewaso Nyiro channel. This data can be used by other governmental or non-governmental bodies to predict the future behaviour of the river, if some climatic and human conservation factors are considered. In the study, we learnt that small and cheap sensor systems can be used to quantify various natural phenomena. The initial cost of setting up the networks to handle data transfer is high but long term benefits of data collections and analysis can be very crucial in the making of various environment related decisions. Apart from the IoT network blackouts and very little vandalism cases, our systems were able to collect data for a long stretch of time. This fact means our system is capable of large scale deployment.

In the near future, we aim to incorporate multiple data sources such as weather data from TAHMO to build water level - machine learning prediction models and use other machine learning algorithms in anomaly detection. Also, we plan to expand our sensor network by deploying more water level monitoring devices, turbidity monitoring devices and flow rate monitoring devices.

The long term vision of this project is to collect enough data that can be used in the development of inundation models. Inundation models are used in flood forecasting in a river basin. The main input of inundations models is river water-level data collected over a long period of time. The other vital input is high resolution elevation maps (depth maps) of the river basin terrain. This means that work in water-level data collection is the first key step in the development of accurate inundation models for the Ewaso Nyiro River basin.

Acknowledgement

We would like to thank the management at Ol-Pejeta Conservancy (Kenya) for allowing us to test our systems at the conservancy and also use the Ol-Pejeta Tech Lab to prepare for deployments and field tests. We also thank Mr Stephen Mathenge from Dedan Kimathi University of Technology for his technical support during system development and deployment. We also thank TAHMO (The Trans-African Hydro-Meteorological Observatory) for giving us access to their weather station data from stations in the Ewaso Nyiro basin. Finally, we thank the NRF (National Research Fund) for funding this work through a multidisciplinary research award.

References

[1] UNESCO, "Facts and Figures from the United Nations World Water Development Report 4," p. 16, 2012.

[2] J. Aeschbacher, H. Liniger, and R. Weingartner, "River Water Shortage in a Highland–Lowland System," *Mountain Research and Development*, vol. 25, no. 2. pp. 155–162, 2005, doi: 10.1659/0276-4741(2005)025[0155: rwsiah] 2.0.co; 2.

[3] B. Notter, L. MacMillan, D. Viviroli, R. Weingartner, and H. P. Linger, "Impacts of environmental change on water resources in the Mt. Kenya region," J. Hydrol., vol. 343, no. 3–4, pp. 266–278, 2007.

[4] R. Hirji SR, Ministry of Environment and Natural Resources, Ministry of Agriculture, National irrigation board, Ewaso Nyiro North and Ewaso Nyiro South Basin Authorities (2000) "Water Resources Management Issues, Challenges and Options: À Framework for a Strategy" Draft Issues Paper (Kenya).

[5] Giannoni F, Mancini M, Marinelli F. (2018) Anomaly Detection Models for IoT Time Series Data. ArXiv preprint arXiv: 1812.00890.

[6] O. Chipara, G. Hackmann, C. Lu, W. D. Smart, and G.-C. Roman, "Practical modelling and prediction of radio coverage of indoor sensor networks," *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks - IPSN '10*. 2010, doi: 10.1145/1791212.1791252.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2016.

[8] G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije, *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*. Cambridge University Press, 2013.

[9] T. Wagener, M. Sivapalan, P. Troch, and R. Woods, "Catchment Classification and Hydrologic Similarity," *Geography Compass*, vol. 1, no. 4. pp. 901–931, 2007, doi: 10.1111/j.1749-8198.2007.00039.x.

---

7 https://kabi23.github.io/anomalydetection/
8 https://www.olpejetaconservancy.org
9 https://lora-alliance.org
10 https://www.influxdata.com
11 https://tahmo.org/weather-stations/