Development of Machine Learning-based rainfall outlook models in Madagascar

Victor Muñoz S.², Lorenzo Bianco¹, Mark Sumka², Andrew Magee³

¹ Independent Consultant, Finale Ligure, Italy
² SRK Consulting, Vancouver, Canada
³ Rio Tinto Energy & Climate, Brisbane, Australia
Rainfall Outlook at QMM, Madagascar

Surface water management at the site is impacted and dependent by climate variability and event intensity.

The climate at the site can be classified as Tropical. The area is affected by tropical cyclones and, rainfall is significant at the site with mean annual precipitation (MAP) of 1700 mm per year.

Climate change and climate oscillations are expected to impact the amount of water available for mine operations and for water storage.

Due by the high rainfall variability in the region, operations often require more specific insights to assist with decision-making.
Climate at QMM

Climatic gridded models were used to develop a comprehensive record of changes to weather and climate over time. Historical observation and numerical models are combined to historical climate conditions through bias correction.

Overall, due to climate change effects:
- annual precipitation is expected to decrease and,
- potential evaporation is expected to increase.

A decrease in annual precipitation and an increase in potential evaporation may shift the site water balance into a water deficit.
Climatic Influence

**MERRA-2** was selected as the best source for gap-filling rainfall records at the site.

For local meteorology, monthly climate parameters were also obtained real-time from MERRA-2 (air temperature, relative humidity, specific humidity, and atmospheric surface pressure).

The understanding of *relevant climatic influences that impact rainfall variability* at QMM site is crucial to forecast rainfall amount.

Variable Importance was studied, the amount of **rainfall** accumulation at the site results **depended** by:

- the **historical weather conditions** for the site (MERRA-2) and,
- the **Climate Indices** (Long-Term Climate Oscillations) such as:
  - El Niño 3.4 Weekly Sea Surface Temperature (SST)
  - Cold & Warm Episodes by Season (ONI)
  - Outgoing Long Wave Radiation (OLR) and,
  - North Pacific Gyre Oscillation data.
Objectives

- Study of the **relevant climatic influences** that impact rainfall variability.
- Implementation a machine learning model to **forecast short to medium-range rainfall**.
- Delivery of a user-friendly web application/dashboard to be used as a **tool on the site to assist with water management decision-making**.

![Diagram showing the process of model inputs, machine learning algorithm, model evaluation, and usage with relevant long term climate oscillations, historical climate parameters, and historical rainfall (1981-2018) on the left, and predictive model, rainfall forecast, and web application on the right.](image-url)
Machine Learning

In this work, machine learning (ML) techniques were implemented to study the relevant climatic influences that impact monthly rainfall variability at the site and to predict rainfall up to 6 months in the future. **Multiple machine learning models (MLMs) were built to estimate the cumulative precipitation at the site for periods from 1-month up to 6-months in duration.**

Monthly rainfall records from January 1981 to January 2023 were used for model implementation, where model inputs were divided by periods into **calibration** (80%), **validation** (20%), and **testing**. The information was resampled with the R library **caret**: data were resampled with a 10-fold cross validation, with this process the machine learning models were tuned.
All the ML models had mathematically comparable performances, but because the primary objective was to evaluate higher values, the review and model selection went beyond the usual model metrics (MAE, RMSE, $R^2$).

Cubist technique does not perform outstanding beyond the other models; however, it tends to have a more consistent performance for small and large precipitation quantities, based on the calibration and validation dataset.
Machine Learning - model performance

To account for the unavailability of reliable day-to-day meteorological records, four different sources of data were chosen to represent the site's precipitation, including MERRA-2 reanalysis data that was bias-corrected at the site.

The six cumulative rainfall outlook models were compared with these sources using scatter plots.

The comparison between the different representative site records and machine learning models achieved an $R^2$, and KGE higher than 0.4 which considered as reasonable results.
Monthly Rainfall Outlook Results

The six cumulative rainfall MLMs implemented, were used simultaneously to extract the results in terms of monthly rainfall predictions up to six months ahead the present date.

<table>
<thead>
<tr>
<th>Outlook model</th>
<th>Aug 2023</th>
<th>Sep 2023</th>
<th>Oct 2023</th>
<th>Nov 2023</th>
<th>Dec 2023</th>
<th>Jan 2024</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-month</td>
<td>147.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2-month</td>
<td>-</td>
<td>23.2</td>
<td>89</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3-month</td>
<td>-</td>
<td>49.2</td>
<td>153.6</td>
<td>65</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4-month</td>
<td>-</td>
<td>105.3</td>
<td>131.8</td>
<td>147.2</td>
<td>-999.9</td>
<td>-</td>
</tr>
<tr>
<td>5-month</td>
<td>-</td>
<td>46.1</td>
<td>140.5</td>
<td>41.6</td>
<td>-999.9</td>
<td>159.5</td>
</tr>
<tr>
<td>6-month</td>
<td>554.8</td>
<td>226.4</td>
<td>38</td>
<td>75.9</td>
<td>197</td>
<td>327</td>
</tr>
<tr>
<td>Median</td>
<td>351</td>
<td>49</td>
<td>132</td>
<td>70</td>
<td>197</td>
<td>243</td>
</tr>
<tr>
<td>Average</td>
<td>351</td>
<td>90</td>
<td>111</td>
<td>82</td>
<td>197</td>
<td>243</td>
</tr>
<tr>
<td>Maximum</td>
<td>555</td>
<td>226</td>
<td>154</td>
<td>147</td>
<td>197</td>
<td>327</td>
</tr>
<tr>
<td>Minimum</td>
<td>147</td>
<td>23</td>
<td>38</td>
<td>42</td>
<td>197</td>
<td>160</td>
</tr>
</tbody>
</table>

Each model were built independently from the others and the models were calibrated and validated as monthly cumulative models.

The extrapolated monthly values are displayed as differences between models, those differences may result in negative values. In some circumstances, no-values are displayed.
Rainfall Outlook Web Application

These results are intended to be used in an operational water management context with ease of use, enabling more informed decision-making processes for water management projects.

A R-Shiny web application was developed to use as a output the models results.

The aim of the application is to make simple for users from the site to visualise and control the analyses from the MLMs and increase the variability of future monthly rainfall.
Vielen Dank für Ihre Aufmerksamkeit!!!
Monthly Rainfall Correlation
Rainfall time series