CASE STUDY:

Development of a rainfall and stream water level quality control subsystem for Belize Climate Data Management Systems (CDMS)

WMO HydroHub Second Innovation Call

Implementation dates:	November 2020 to May 2021
Innovation supplier:	Elligence Solucoes em Tecnologia
Project implementation country:	Belize

Challenge

Standard Quality Control (QC) tests that check if measurements are included within valid thresholds usually do not detect issues related to the reliability of rainfall and water level measurements. Accumulated dirt, dead leaves and insects in sensors, which lead to false zero rainfall measurements, create a problem for hydrometeorological observation networks. Early detection of false zero rainfall measurements, for example, has the potential to improve the availability and reliability of rainfall data, and to improve indirectly hydrological forecasts.

Over many years, the use of artificial intelligence (AI), or more specifically machine learning, to resolve this problem has been studied with different degrees of success. However, recent scientific work, based on deep learning and ensemble algorithms, has demonstrated that these methods have reached a level of accuracy that is promising for the implementation of fully automated AI based QC tests. These AI solutions could even alert hydromet services about the type(s) of problem with the rain gauges and water level sensors, which standard tests cannot do.

The Software

Surface – a modern web-based CDMS – ingests data from manual and automatic stations. It has a real-time QC module that uses standard tests (range, persistence and step) and flags (good, bad, suspect, inconsistent) based on the Oklahoma Mesonet network QC system.

The software proposed for the project, a new system to be integrated with Surface, was developed to run automated QC tests based on machine learning algorithms. It would fully-automate the CDMS QC subsystem, based on supervised classification methods and on learning patterns from manually flagged measurements by the operators of the National Meteorological Service (NMS) of Belize. In turn, this would improve the reliability of hydrometeorological measurement data and thereby build trust in this new source of information that the NMS of Belize is starting to provide to external stakeholders.

Approach

The project was implemented in five phases:

1 In order to enable the training of machine learning models based on deep learning algorithms, a Graphics Processing Unit board was purchased and installed in the NMS.

2 A rainfall dataset covering a period of three years was extracted from Surface CDMS for a set of stations located in the Belize watershed. Known problems with the rain gauges in the selected stations over the three years were manually labelled and introduced as artificial fault events in the original dataset.

3 Different modern machine learning algorithms networks were evaluated. The prepared dataset was used to train and evaluate machine learning models based on metrics such accuracy, precision and recall. The models with the best results were selected for use in the new QC module.

4 A module for the automation of the machine learning based models was developed in Python and integrated with Surface CDMS. This module was developed based on a plugin architecture that allows new ML-based models to be added to the system. A web-based user interface for managing and evaluating the execution of the QC models was also developed. This module was installed in the server that currently hosts the Surface CDMS.

5 An online workshop introduced the NMS to the fundamentals of machine learning concepts and application in hydrology, reported the results obtained in the modelling and experimentation phase, and trained staff on the use and maintenance of the new QC module.

Results

- The project enhanced the reliability of rainfall and water level measurements by detecting issues that usually go unnoticed in standard tests.

- A new open-source QC system (HydroML) with a web-based user interface was installed in the NMS's servers. Its source code, released under an open source license, is hosted in GitHub (https://github.com/HydrohubChallenge/hydroml).

- Three of the NMS's staff members were trained to understand the core concepts of the new QC system and taught how to maintain and operate it.

Way Forward

1 Continue collaborating with Belize to improve HydroML by implementing automatic AI-based quality control procedures to other hydrological stations managed by Surface

2 Implement new algorithms/models for other meteorological parameters (temperature, humidity, wind, solar radiation)

3 Continue exploring external datasets such as weather radar products and predictors for rainfall models.

Partners

Elligence Solucoes em Tecnologia

National Meteorological Service in Belize.

> **To contact the team please send an email to: hydrohub@wmo.int**