TEAM ARGHYAM

Wave2Web Hack

Sustainability: Predicting Reservoir Water Levels

Introduction

Cities in the Global South face unreliable,inadequate and polluted supply of freshwater. About one billion people don't have access to safe and continuous water supply. Due to rapid increase in urban population,there is increased competition for water supply,which poses increasing risks for water supply. This has serious implications on health and wellbeing.

There is a need for transparent data to inform water supply risk management during periods of water insecurity. Data should include: access to real time water risk management and short term forecasting of reservoir water availability. This is a concept note towards developing a predictive model that can strengthen water resources planning and risk preparedness. The model to be developed should estimate reservoir water availability at least 1-3 months in advance using hydro-meteorological parameters and reservoir water level data.

Hydrological Model

Bangalore, located in the southeastern part of Karnataka, is the sixth largest city of India and one of the fastest growing cities of Asia, with a total population of close to 10 million. Below table shows land use dynamics in Bangalore.

Table 1: land use dynamics in Bengaluru

Area in sq.km								
Year/LU	**1973**	**1992**	**1999**	**2003**	**2008**	**2010**	**2016**	**Pred2020**
Built up	5448	18650	24163	25782	35301	37266	54807	66463
Vegetation	46639	31579	31272	26453	20090	16031	5364	2108
Water	2324	1790	1542	1263	613	617	696	696
Others	13903	16303	11346	14825	15256	14565	10394	2002
Area as %								
Year/LU	**1973**	**1992**	**1999**	**2003**	**2008**	**2010**	**2016**	**Pred2020**
Built up	8.0	27.3	35.4	37.7	49.5	54.4	76.9	93.3
Vegetation	68.3	46.2	45.8	38.7	28.2	23.4	7.5	3.0
Water	3.4	2.6	2.3	1.8	0.9	0.9	1.0	1.0
Others	20.4	23.9	16.6	21.7	21.4	21.3	14.6	2.8

Water Usage Dynamics

As evident from the above table, encroachment and unplanned urbanization over the past 4 decades has cost the city heavily in terms of its water bodies [1]. Today the number of water bodies in the city is just over 20% of what once existed. Bengaluru has two main sources of water: The River Kaveri (Cauvery) and Groundwater. Water from the Cauvery is collected in the Krishna Raja Sagar Dam in Mysore. From here, the BWSSB (Bangalore Water Supply and Sewerage Board) pumps water and transports it to over 660,355 BWSSB connections in Bengaluru.

Krishnarajasagara Reservoir

Krishnarajasagara Dam is situated at Latitude 12º 24’58” North and Longitude 76º34’26” East, in the village areas of Krishnarajasagara village, Srirangapatna taluk and Mandya district. See Figure 1 below [2]. The planning is as follows:

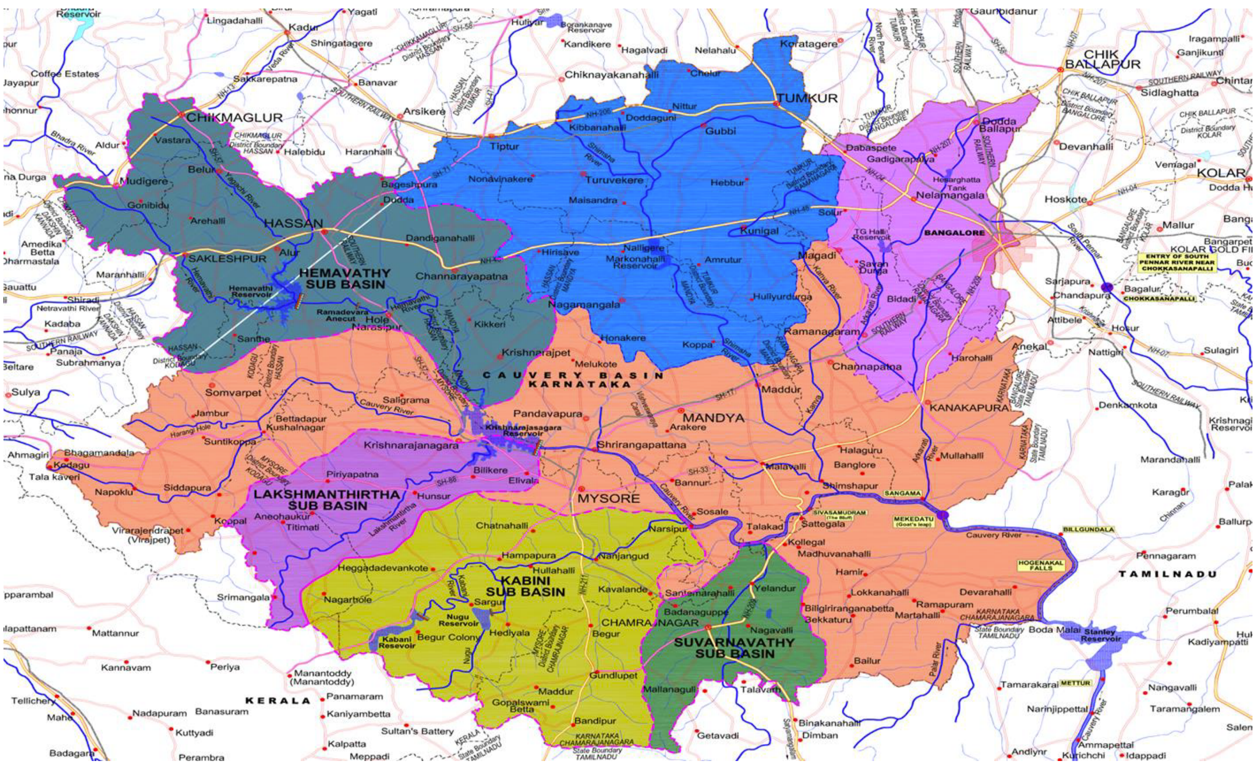


Figure 1- Cauvery Basin Map

(i) **Catchment**: The dam has a total catchment area of 10880.63 sq km consisting of 3229.58 sq km intercepted by reservoirs on major rivers of Harangi (419.58 sq km) and Hemavathy (2810 sq km) reservoirs and 7651.05 sq km independent.

(ii) **Yields**: The maximum and minimum annual virgin yields at Krishnarajasagara Dam site from derived run off data for the years 1901 to 1980 area 34870.5MCM (1232.17 TMC) and

13183MCM (465.83 TMC) respectively. The 75% dependable yield at this site is worked out to be 21057MCM (743.64 TMC). Average annual yield is 24071.17MCM (850.59 TMC).

(iii) **Storage**: The reservoir formed by the dam has a gross storage capacity of 1400 MCM(49.452 TMC) and live storage capacity of 1276 MCM(45.051 TMC)

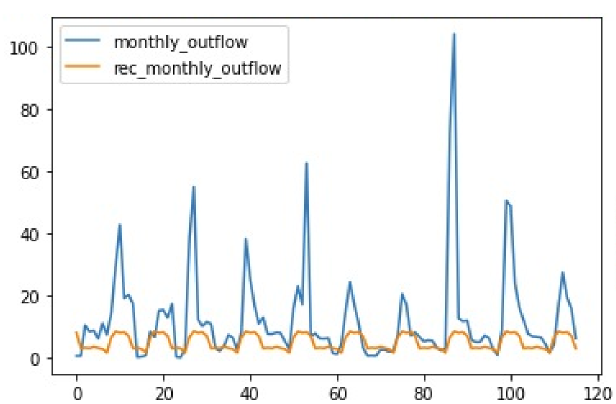
(iv) **Water Spread**: The reservoir water spread submerges an area of 49.90 sq. miles displacing a population of about 25 villages.

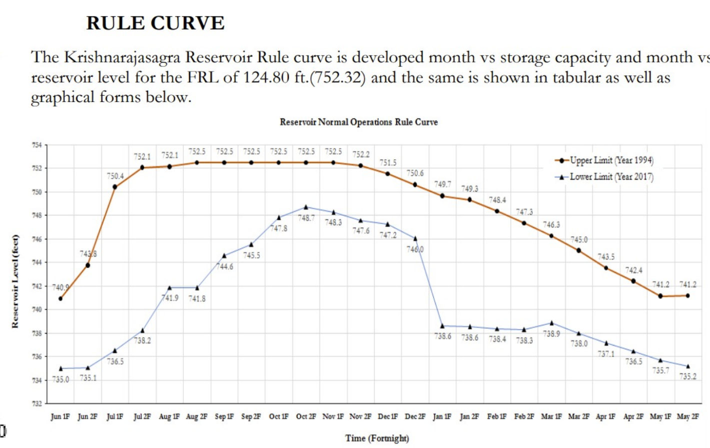
Outflow Dynamics

Table 2: KRS Outflow Analysis

Month	Recommended water releases in TMC	
	Irrigation	Drinking Water Supply
June	0.00	1.685
July	4.92	1.685
August	6.857	1.685
September	6.390	1.685
October	6.603	1.685
November	5.364	1.685
December	1.368	1.685
January	1.565	1.685
February	1.414	1.685
March	1.873	1.685
April	1.490	1.685
May	1.160	1.685

Table on the left shows the details of monthly recommended water releases to the irrigation and drinking water supply from KRS dam [2]. These recommendations are based on the rule curve shown below. The plot below left compares the actual recorded outflow to the monthly recommendations. It is evident how the actual usage is much more in the dry seasons of the year.





As shown in Table 1, the total of the built up area and the vegetation does not have much variance. This also confirms the outflow trend being flat over years with seasonal spikes as shown above.

Inflow Dynamics

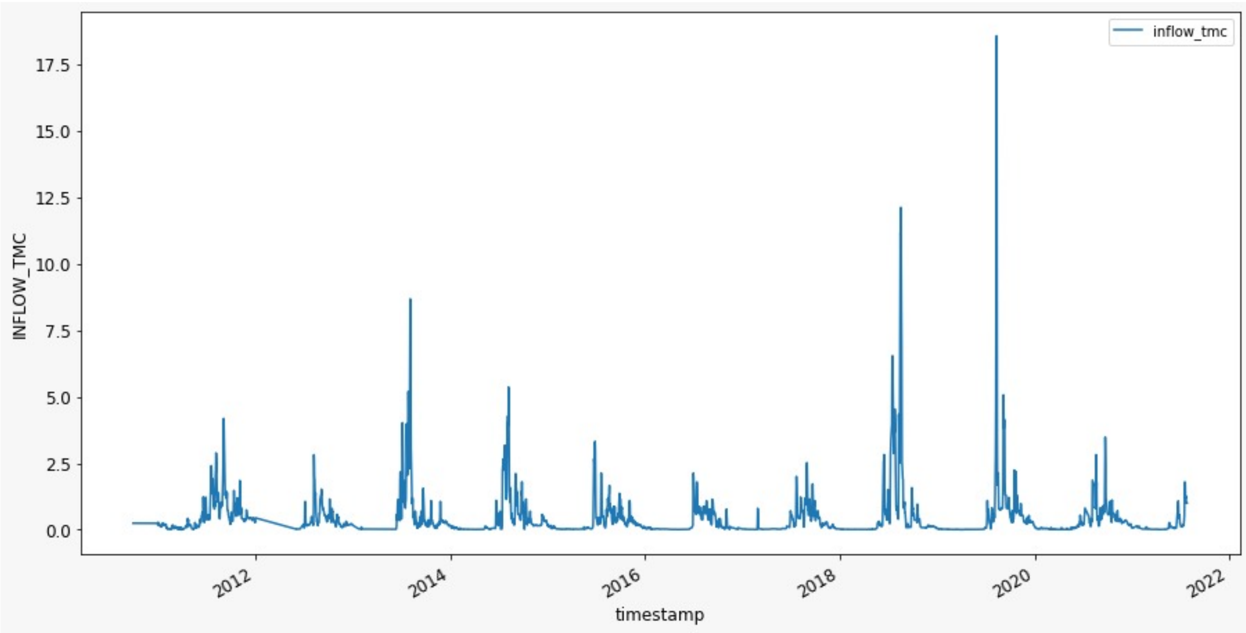


Figure 2 - Inflow in to KRS over past 10 years

The inflow depends upon the three streams coming in - Hemavathy, Cauvery and Harangi. Figure 3 shows the correlation heatmap of how the inflow, outflow and storage of the three reservoirs (KRS, Harangi and Hemavathy) are related to one another. The inflow of KRS has 0.72 correlation with outflow of Harangi and a 0.77 correlation with the outflow of Hemavathi. The inflow to inflow correlations are also high. Hemavathi being the larger reservoir, its storage has high correlation to storage of KRS. Owing to these high correlations, we decided to independently model KRS instead of making the modeling process complex following Occam's Razor.

Weather plays a major role in the amount of water that flows in. We analyzed weather across the major cities along the three inflow channels as well as the mean Karnataka weather, with data collected from visualcrossing weather data service. The free usage gives 1000 records a day per user. The collected data is in our GitHub Repository.

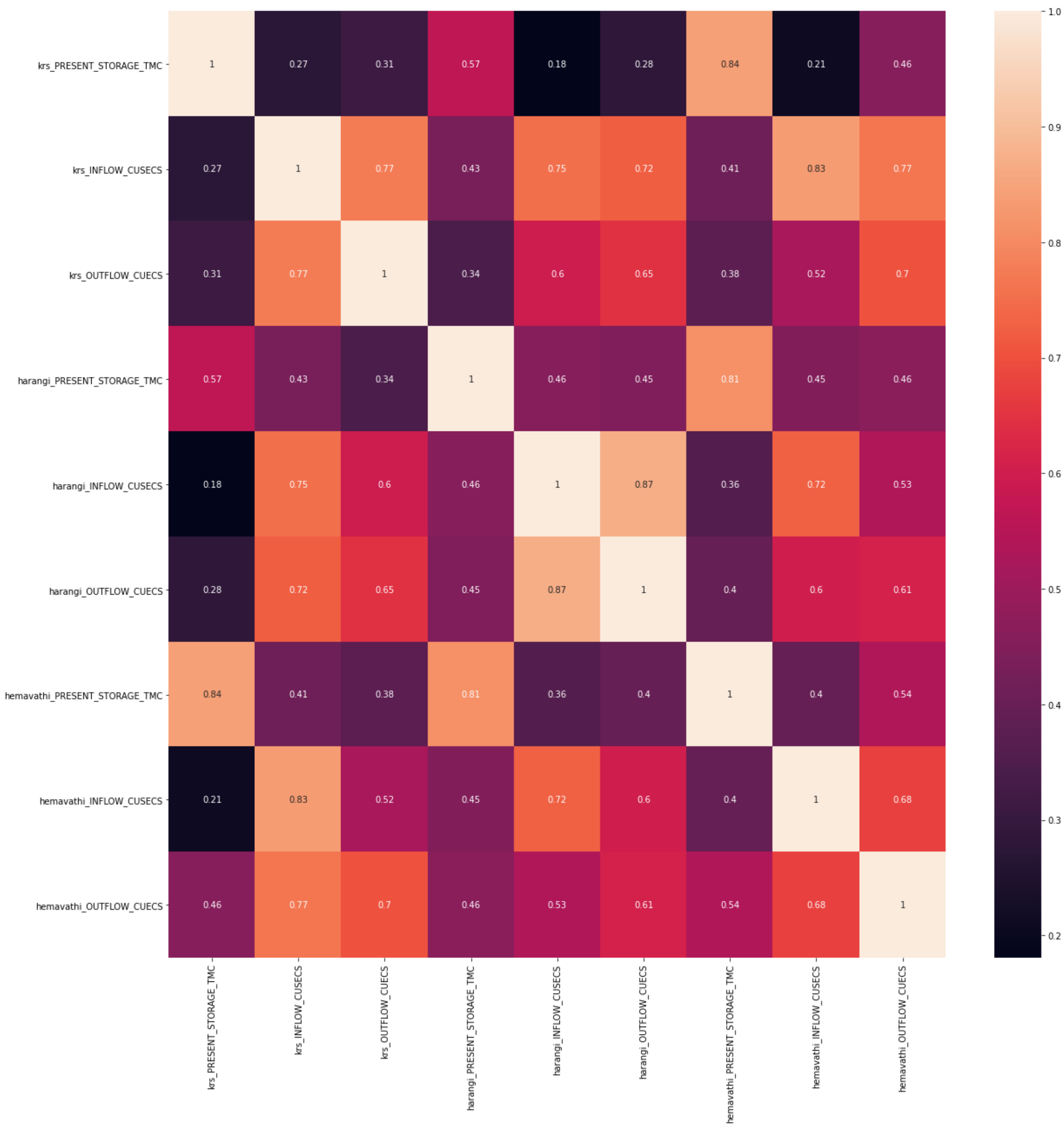


Figure 3 - Correlation Heatmap between inflow, outflow and storage of three reservoirs KRS, Harangi and Hemavathi.

Figure 4 below shows 1 year of mean Karanataka weather and the KRS inflow plotted on the same time scale for 2018. Notice how the inflow is more when temperature is lower and humidity and cloud cover is more and vice versa.

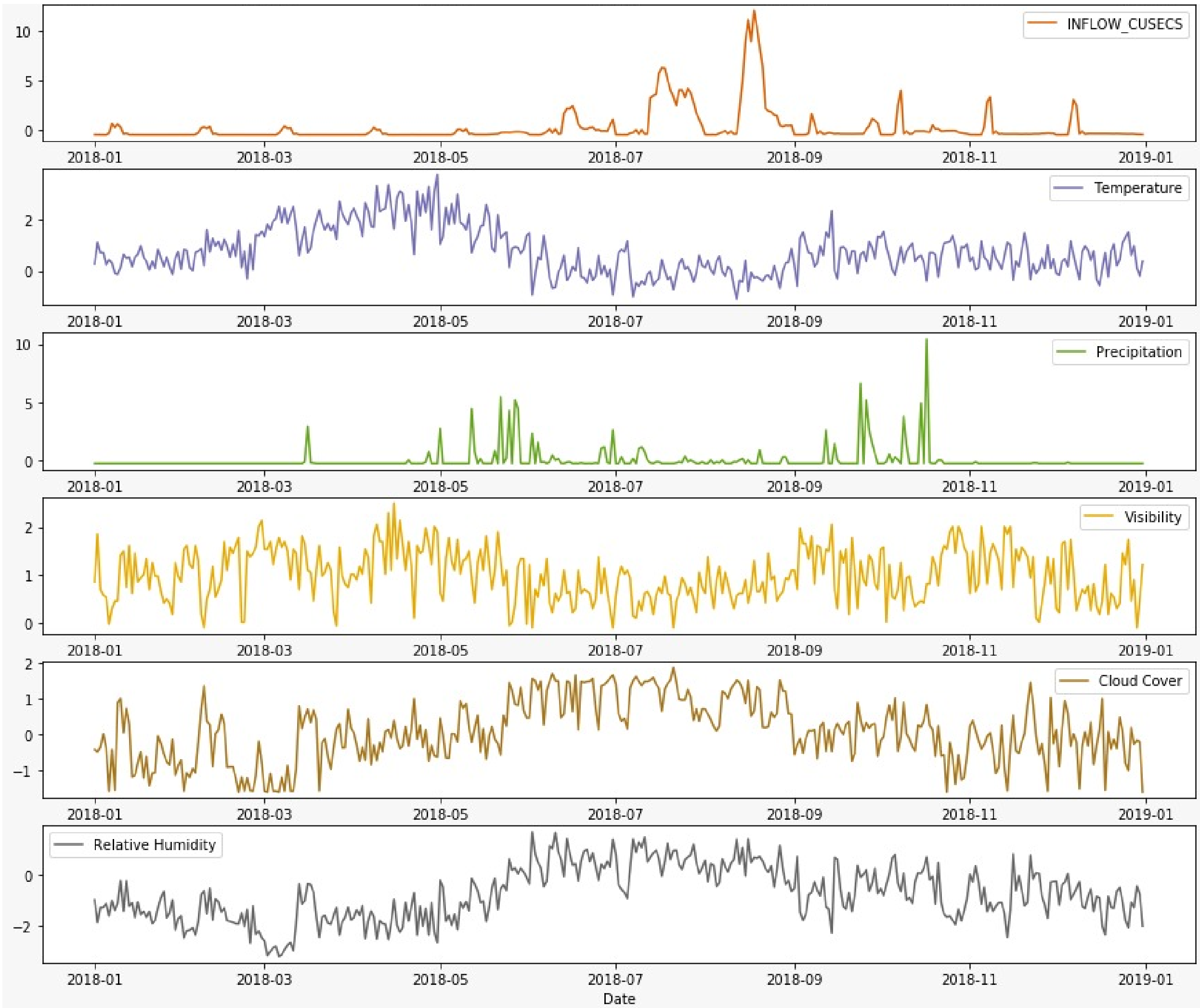


Figure 4: 2018 variations in Karnataka weather and KRS inflow.

We analyzed the correlation of various weather indicators to the KRS water characteristics. The cities we analyzed for weather are: Haranhalli, Heggadikpplu, Akkihebbal, Chikmanglur, Keralapura, Kodugu, Krishnarajanagara, Mandya, Mysore and Saligrama. All these regions are topographically at a higher elevation. Bangalore and further east regions are at lower topography. While Bangalore rains may reduce outflow from KRS it has no impact on the inflow.

We found that each location had some correlation with the reservoir however the average Karnataka weather had more of a relationship. Figure 5 shows the correlation of daily humidity and max temperature at various locations to daily inflow and storage.

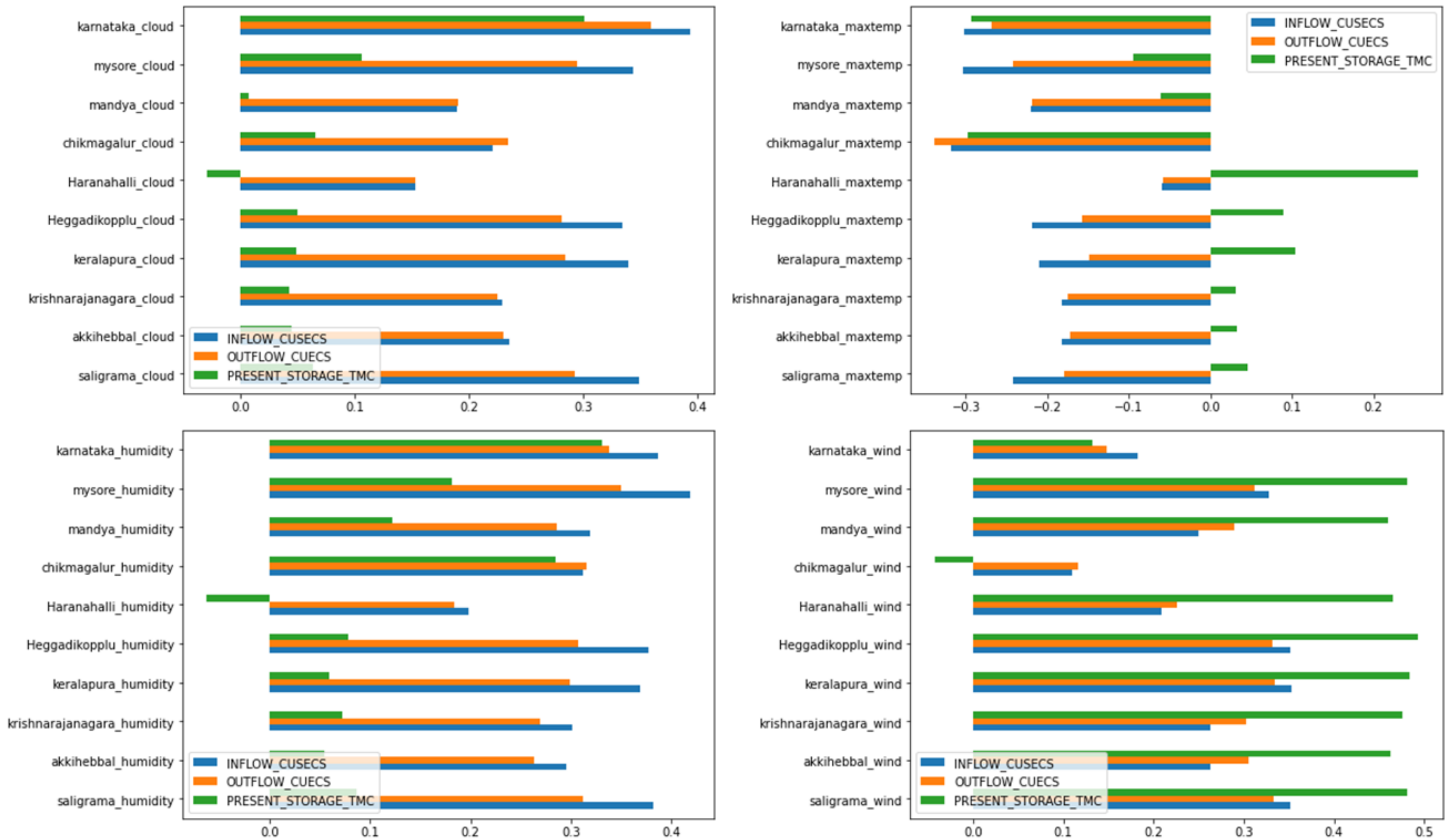


Figure 5: Correlation of cloudcover (top left), maximum temperature (top right), humidity (bottom left), and wind (bottom right) with inflow, outflow and storage for KRS across various locations.

The correlations of Figure 5 indicate that overall there is a positive correlation of storage to wind, humidity and cloud cover, but a negative correlation with maximum temperature which seems logical as hotter days would mean more evaporation, less inflow and more water use. However certain locations closer to KRS, show a reverse correlation, suggesting that we need to look at a wider region than only the neighborhood. Overall the average Karnataka weather shows similar and consistent trends and hence we chose to inluse that only in our predictive model.

Cumulative rains would have a greater impact and hence we tried to analyze how cumulative weather will correlate with KRS features. Figure 6 shows a characteristic curve suggesting that average over an optimal number of days gives a highest correlation to the inflow for KRS reservoir.

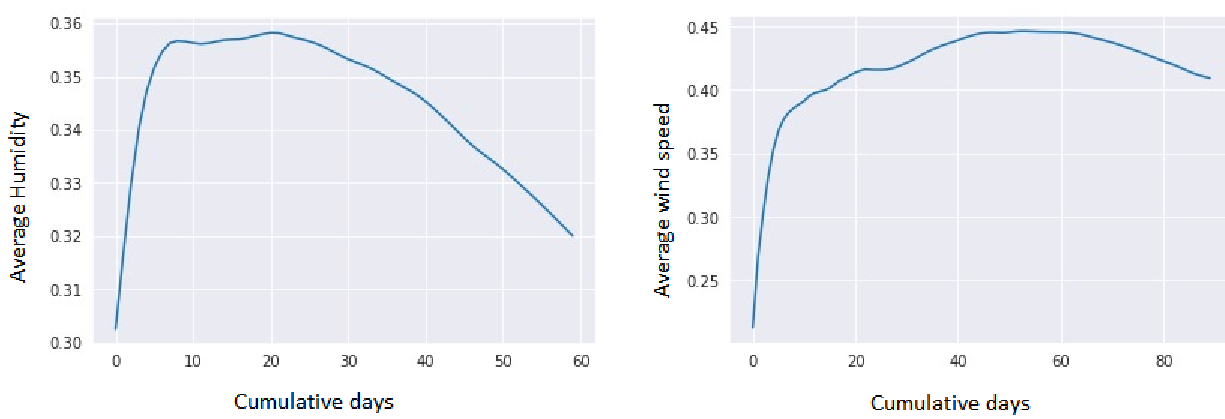


Figure 6: Correlation between average humidity and wind speed over cumulative days and KRS inflow.

Figure 7 shows the correlation matrix over 21 days as an example. If we need to predict the storage TMC, the hydrological model can clearly benefit from moderate to highly correlated facts. In this case we can use Max Temperature (correlation coefficient = -0.48), Visibility (correlation coefficient = -0.52), Humidity (correlation coefficient = 0.31), Cloud cover (correlation coefficient = 0.47). Since humidity and cloud cover have high correlation (0.83) we can use either one of them. Wind is weakly correlated to storage but shows a 0.41 correlation with wind, so that makes it a good candidate too.

Figure 8 shows the same correlation, but this time comparing 21 days cumulative KRS data to 21 days cumulative weather data. This does boost inflow correlations while reducing slightly the storage ones as storage has also a relationship to the outflow that also shows more correlation now.

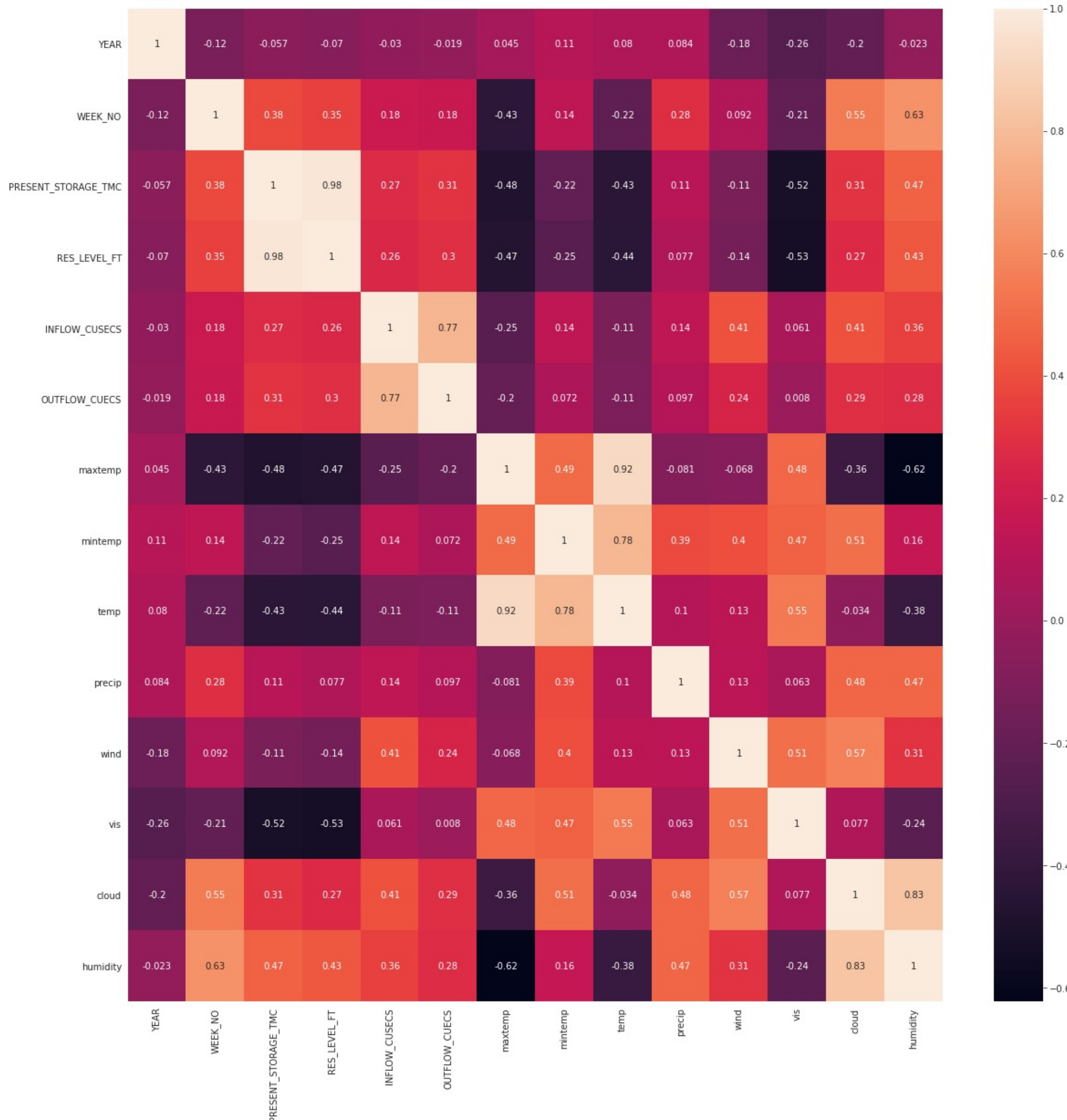


Figure 7: Correlation between KRS data and mean Karnataka Weather with 21 days cumulative.

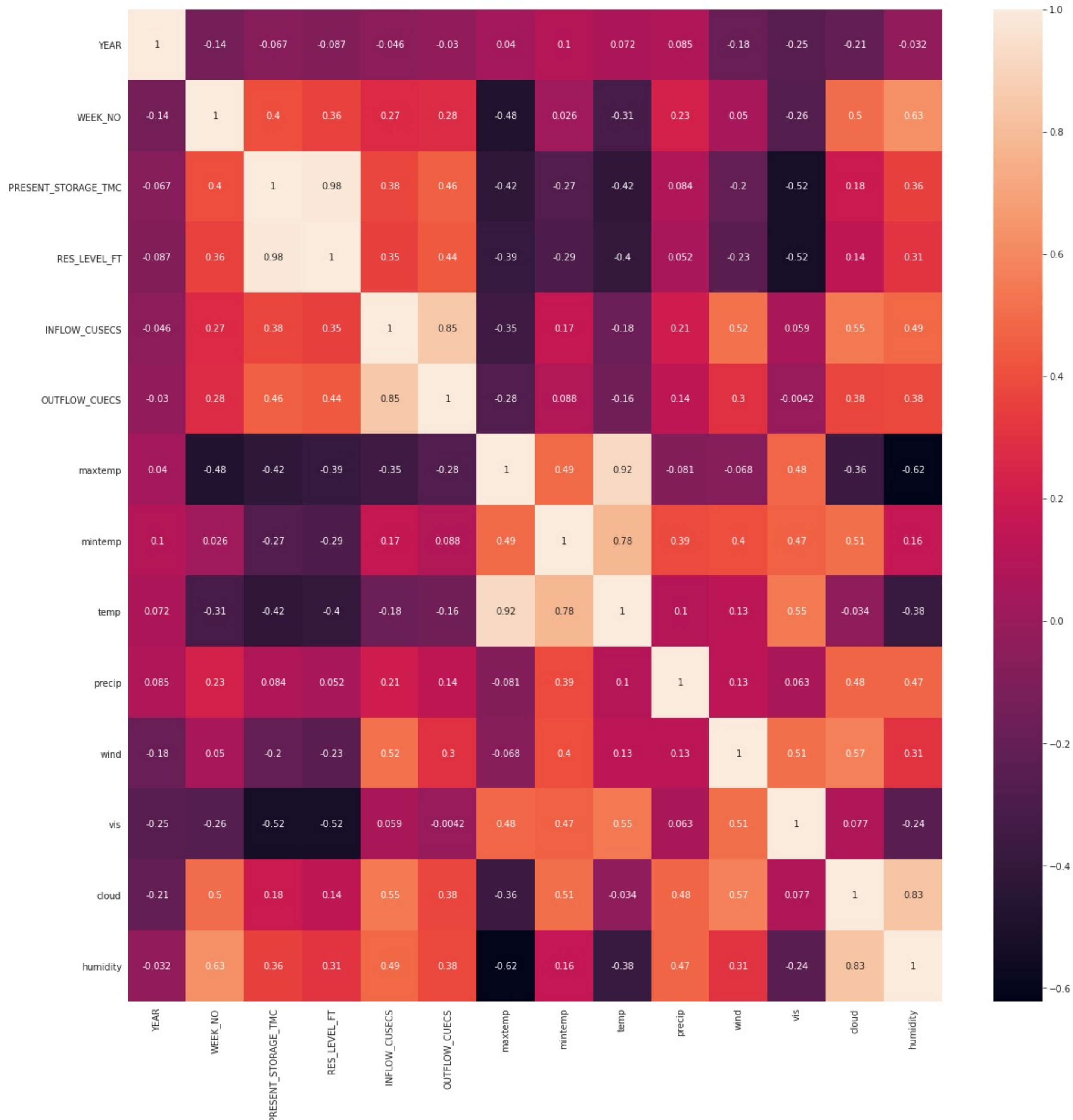


Figure 8: Correlation heatmap between 21 day cumulative KRS reservoir and Karnataka weather data.

Predictive Model

We have three sources from which we are getting data to populate the database for the future prediction

1) **Observed data at the four reservoirs provided by CWC**. We not only are using the given data, but also collecting live data from available services [here](). Our system daily runs updates to newly incoming data and updates the predictions accordingly.

2) **Simulated data using hydrological modeling**. Land usage and climate change can impact water levels, inflow and outflow. In our analysis as presented in the section for outflow dynamics, the overall outflow has shown consistency and while the usage distribution has changed, the total usage has not shown much variation.

3) **Spatially aggregated data from remotely sensed data products**. We collect weather (both historical and future predictions up to 90 days) from available weather service and store it in a sqlite database. There can be various parameters to measure, but many are correlated. For example evaporation or evapotranspiration will have close relation to temperature and humidity. So the weather data in itself is complete.

Modeling Strategies

We have experimented with various modeling strategies that we shall discuss in this subsection.

Standard Time Series Analysis

A standard self regressive time series analysis will not weigh in multiple factors. However this is the simplest place to start.

Deep-learning models can deal with time series in a scalable way and provide accurate forecasts.

Baseline Model

This base model is a simple RNN based model to perform univariate time series forecasting on water availability. The Latent Dimension is the number of units in the RNN and Horizon is the number of values in prediction. The Horizon is set to 1 as this model is intended to forecast only one value. The model architecture shown in Figure 9 and it has a latent size of

5. Taking this model as inspiration, we have built a multivariate and multi-horizon stacked RNN based model to forecast 30 days, 60 days, and 90 days.

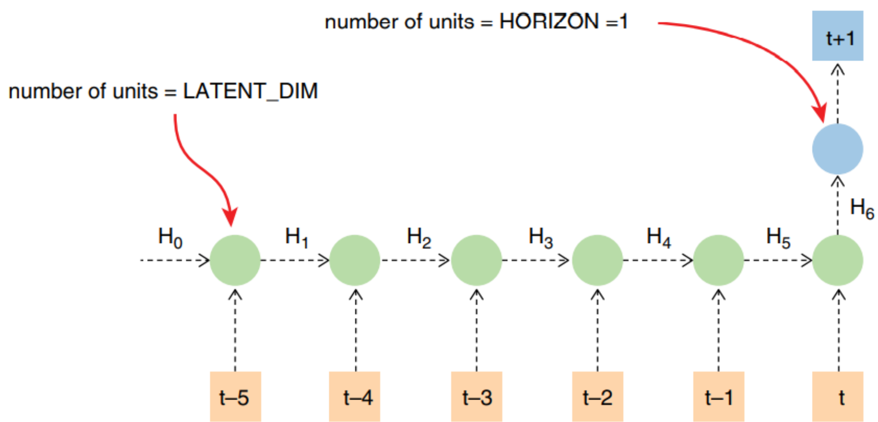


Figure 9: Structure of Simple RNN model to forecast water availability of the next day

Multivariate Time Series Model

As more observations are showing the impact on the present water availability of the reservoir, we have trained a multivariate time-series model. The considered features are present_stroage_tmc, inflow, outflow, max_temperature, visibility, humidity, and wind. The steps followed in model development are

1. prepare time series data for training an RNN based Deep Learning model.
2. Get data into the required shape for the Keras API.
3. Implement an RNN model in Keras to predict the next step (time t + 1) in the time series. This model uses recent values of inflow, outflow, max_temperatue, visibility, humidity, and wind as the model input.
4. Enable early stopping to reduce the likelihood of model overfitting.
5. Evaluate the model on the test dataset.

The Architecture of the final model is shown in Figure 10.

In Figure 10, T is the latent dimension specifying the number of lags and H is Horizon specifies how many days in the future we want to predict. Input to the model is a set of T values and outcome is a set of H values.

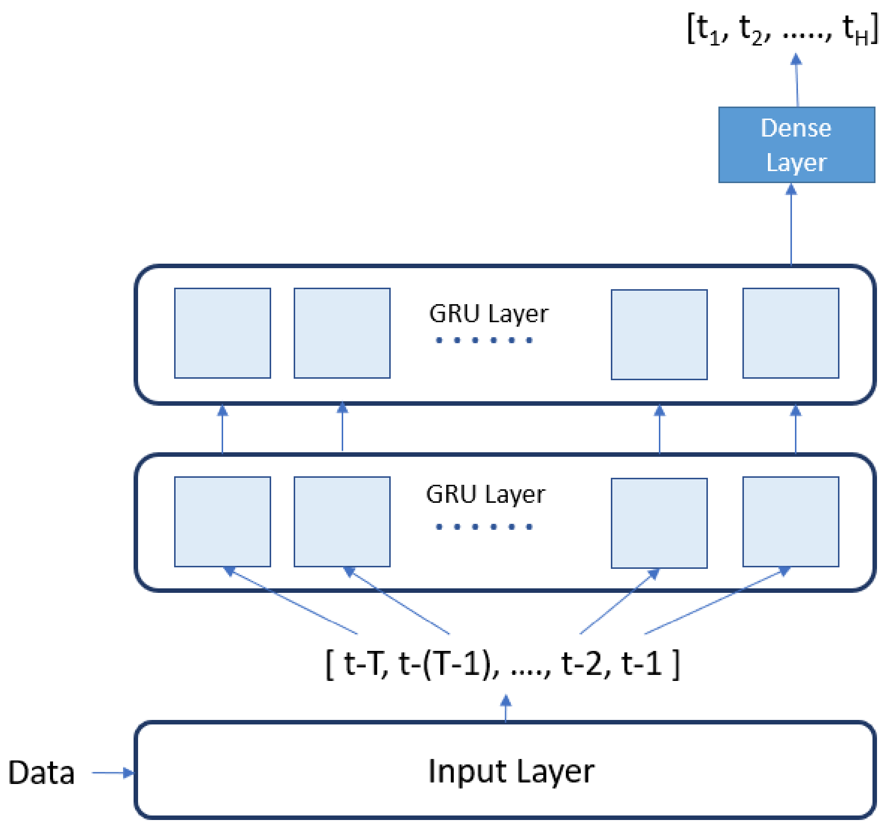


Figure 10: Adopted LSTM based architecture for 30 days forecast

For training the model, we have divided the data into train, validation, and test datasets, and their distribution is shown in Figure 11. We have created a convenience class with the name TimeSeriesTensor and used it as an input layer to prepare the input data. It performs the following steps.

1. Shift the values of the time series to create a pandas DataFrame containing all the data for a single training example.
2. Discard any samples with missing values
3. Transform this pandas DataFrame into a Numpy array of shape (samples, time steps and features) for input into Keras

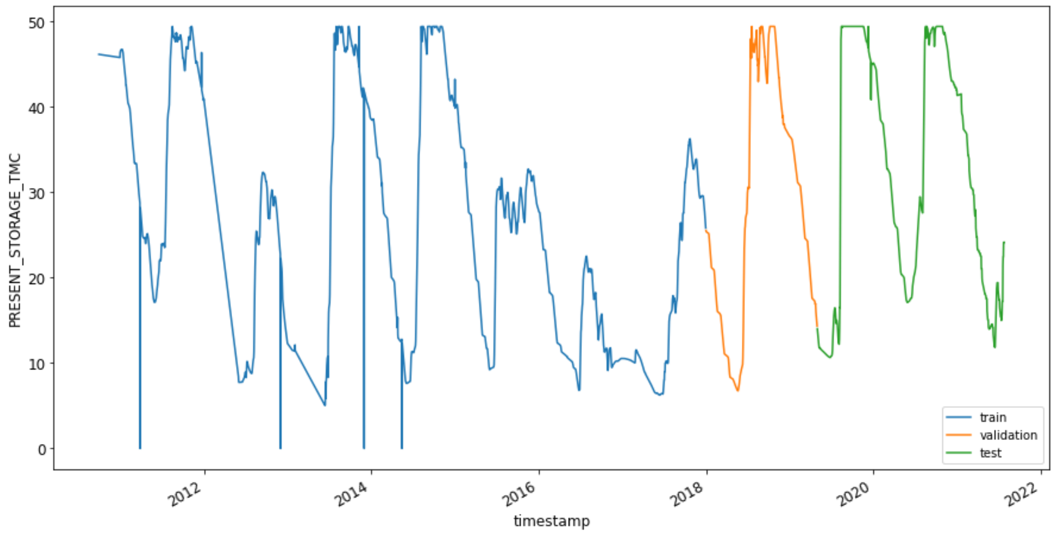


Figure 11: Train, validation and test data

We have experimented with different latent dimensions, horizons, batch_sizes, optimizers. The capacity of the model varies with latent dimensions, and horizon parameters. The accuracies of the different models (notebooks) trained shown in Table x.

Table 3. Performance of various Models with different Latent Dimension, Horizon

Model No.	Model Capacity	Latent Dimension	Horizon	MAPE	Accuracy
1	9,121	30	1	0.01	99
2	9,555	30	15	0.05	95
3	10,020	30	30	0.08	92
4	36,210	60	30	0.1	90
5	78,600	90	30	0.096	90.4
6	38,040	60	60	0.1	90
7	84,060	90	90	0.18	82

As the predictive model should forecast water availability of the future 1 to 3 months, we decided to use Model No. 3. For forecasting more than 30 days, an ensemble of Models 3, 4 and 5 can be used. The sample workflow of model 3 is , if a date is given then an input is prepared with all the necessary features for the previous 30 days and forecasts the water storage for the future 30 days. The model performance is illustrated in figure 12.

The above models used GRU as it takes less training parameters, thereby uses less memory, executes faster and trains faster than LSTM. But the model performance is quite low and they are overfitting.

Better performance Models

Accuracy is more critical, if the input sequence is large. The GRU is replaced with LSTM as LSTM is more accurate on dataset using longer sequences. To reduce overfitting, the model capacity is reduced, regularization and dropout are introduced.

S.No.	Model Summary		Parameters	Window Size	Prediction	Training Loss	Validation Loss
	Hidden Units	Output Units		T (in days)	H (in days)		
1	30	30	5490	30	30	0.098	0.0011
2	60	60	19980	60	60	0.037	0.001
3	90	90	43470	90	90	0.098	0.0011
4	40	30	8910	180	30	0.0028	0.0015
5	70	60	26100	180	60	0.0048	0.0015
6	100	90	52290	180	90	0.059	0.0016
7	30	30	4351	90	1	0.0008	0.00038

Models 1 to 6 were trained on both weather and reservoir data. The features considered are Storage TMC, Inflow, Outflow, Max Temperature, Visibility, Wind, Humidity. Model 7 is trained only on weather data.

The source code is available at https://github.com/Arghyam-Team/KRSPrediction.

Prediction Dashboard

The prediction dashboard is deployed in a free tier AWS EC2 ubuntu instance using Streamlit. Figure 13 shows the simplified architecture of the dashboarding process. A daily updater process collects updates from weather service and reservoir daily update service to update/append data to a pre-created SQLite database. A Predictive model repository stores the various models we have trained. Each model has different predictive power and works with different hydrological parameters. The predictions from each model are also stored in the database. The weather data, reservoir history and forecast data is then visualized in an interactive Streamlit dashboard. Figure 14 shows the schema of the database.

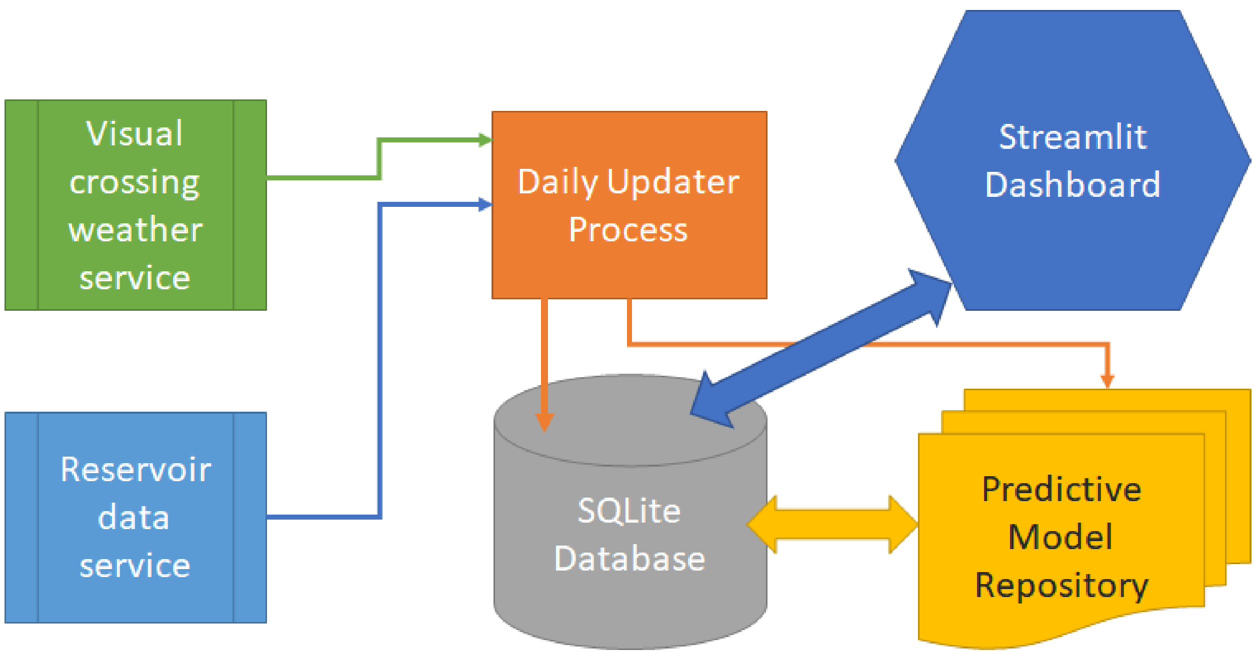


Figure 13 - Dashboard architecture

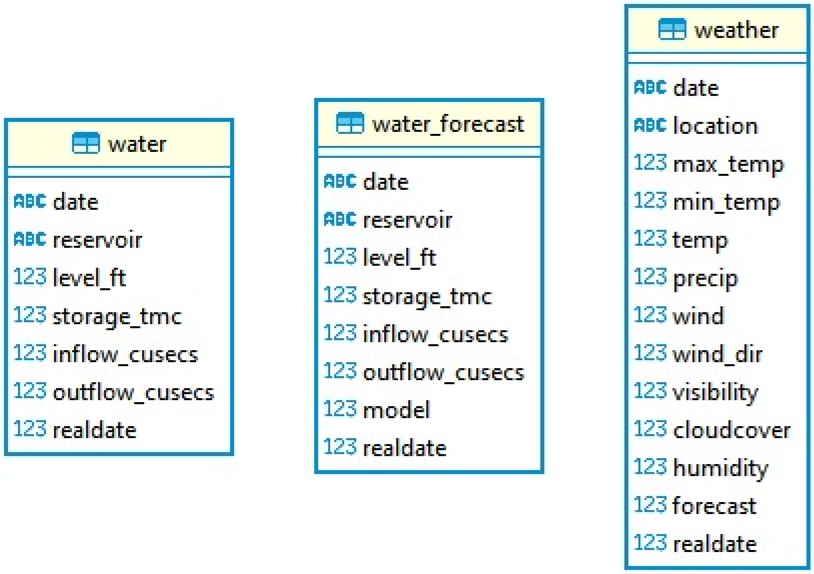


Figure 14 - SQLite Database Schema

Generalize Ability

The analysis and dashboard is amenable to generalization. The same process can be used to train multiple models. The model metadata contains all information like what reservoir it is for, how many days, the parameters it uses etc. So one can train multiple models and visualize on the dashboard. We have only done the KRS reservoir prediction in this case.

References

1. https://www.karnataka.com/bangalore/what-is-the-source-of-drinking-water-in-bangalore/
2. http://waterresources.kar.nic.in/KRS_OM_KaWRD.pdf