Abstract
Hydrologic forecasting is normally based on physically-based (process) models based on equations describing the water flow. Data-driven models use methods of machine learning (e.g., neural networks, support vector machines, etc.) and built using the historical data describing flows and hydrological loads. Modular models include sub-models each of which is responsible for a particular hydrological condition, and they are typically more accurate than the overall global models. Many algorithms for allocating such regions to local models typically do this in automatic fashion, and this is not well-accepted by experts. In forecasting natural processes domain experts want to bring in more knowledge into such allocation, and have to control certain over the choice of models. Another problem is that modular models often use so-called greedy algorithms which are sub-optimal and can be improved. The research investigates an approach to make modular models more "expert-friendly". An issue of including a domain expert into the modeling process is

\[ Qt+6 = (REt, Qt) \]

Predictive model: \( Qt+1 = f(\theta) \)

The research investigates an approach to make modular models more "expert-friendly". An issue of including a domain expert into the modeling process is

\[ \begin{align*}
\text{AN} & : 5.84 \\
\text{M5' } & : 5.05 \\
\text{M5flex} & : 4.75 \\
\text{M5opt} & : 4.29 \\
\text{ANN} & : 4.08 \\
\text{HiMM} & : 4.13 \\
\text{Mixed} & : 4.04 \\
\text{MFS} & : 4.29 \\
\text{Regression} & : 4.30 \\
\text{Total} & : 4.34
\end{align*} \]

**Optimization of hierarchical local models: M5opt algorithm**

**M5opt algorithm**
- Semi-non-greedy model tree algorithm
- To search the optimal tree structure and split rules
- User enables to define level of tree until which non-greedy algorithm is employed
- It is a trade-off between speed and optimality
- Exhaustive search or random search (e.g. GA) is applied for global optimization

**Additional features of M5opt**
- Left-out splitting
- Range of instances where the value split is searched (in %)
- Initial approximation
- Average values or linear models Level investigation
- Exhaustive search
- Semi-non-greedy algorithm
- Optimizes tree structure and split rules
- Compacting tree
- Prunes the tree until user-defined level of the tree
- Works together with standard pruning technique

**Case study:**
- Bagmati river basin: Central Nepal, area of about 3700 km².
- Sieve catchment: Central Italian Apenines, closed to the section of Fornacina, area of 841.96 km².

**Data preprocessing**
- Slieve Catchment
  - Hourly data resolution
  - Predictive models: \( Q(t+1) = f(REt, REt-1, REt-2, REt-3, REt-4, REt-5, Qt, Q(t-1, Q(t-2)) \)
  - \( Q(t+3) = f(REt, REt-1, REt-2, REt-3, Qt, Q(t-1) \)
  - \( Q(t+6) = f(REt, Q(t) \)

**Bagmati Catchment**
- Daily data resolution with threshold 300 m³/s for separating high and low flows
- Predictive model: \( Q(t+1) = f(REt, REt-1, REt-2, Qt, Q(t-1) \)

**Prediction results**

\[
\begin{align*}
\text{ANN} & : 5.94 \\
\text{M5' } & : 6.19 \\
\text{M5flex} & : 6.36 \\
\text{M5opt} & : 6.13 \\
\text{ANN} & : 6.12 \\
\text{HiMM} & : 6.10 \\
\text{Mixed} & : 6.06 \\
\text{MFS} & : 6.16 \\
\text{Regression} & : 6.11 \\
\text{Total} & : 6.13
\end{align*}
\]

**Predicting discharge Q(t+1) for Bagmati-All**

**Conclusions**
- The modular models allow for building accurate specialized models that can capture the details of the processes characteristic to certain regions of the input space.
- Optimal algorithm M5opt is an improvement of the greedy M5 algorithm and allows for flexible trade-off between speed and optimality.
- Since the local models are often simpler than the global ones, they can be made more transparent for decision makers.
- The incorporation of domain knowledge into modular modeling, like M5flex method, allows for a more accurate account for the details of the behaviour of the modelled system. Such models are typically trusted more than the purely machine-learning predictors.

**Acknowledgment**
This work is partly supported by the Delft Cluster Research Programme (project safety from flooding).

**References**