Postprocessing of Numerical Weather Forecasts Using Online Sequential Extreme Learning Machines

Aranildo R. Lima[1] Alex J. Cannon[2] William W. Hsieh[1]

[1]Department of Earth, Ocean and Atmospheric Sciences
University of British Columbia

[2]Pacific Climate Impacts Consortium
University of Victoria

AMS 95th Annual Meeting

Machine Learning methods have been successful used in Environmental Sciences

- downscaling precipitation and temperature.
- prediction of flow, water level, salinity, etc.
- forecasting wind power.

However, the availability of real-time data can also play an important role in forecast problems (improves the forecasting).

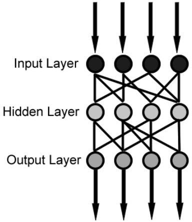
- batch learning: whenever a new data is received it uses the past data together with the new data to retrain the model.
- online sequential learning: discard the data for which the training has already been done (do not require the past data to retrain the model).
  - Computationally much faster and more space efficient.
- Traditionally: linear models (OS-MLR, Moving average, etc), UMOS, Kalman Filtering.

Extreme Learning Machine (ELM)

- Hidden layer weights and biases are randomly assigned
- ANN without iterative tuning
- Solution as a linear system
- Faster than gradient-based training algorithm
- Two parameters:
  - Number of hidden neurons
  - Range of initial uniformly-distributed random weights and bias

OS-ELM (extension of ELM)

- It can learn data one by one or chunk by chunk with fixed or varying chunk size.
- Totally automated procedure



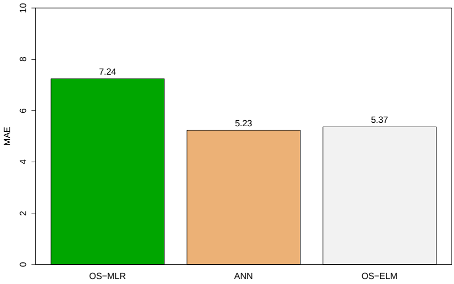
Forecast daily streamflow for the Stave River above Stave Lake, BC , Canada.

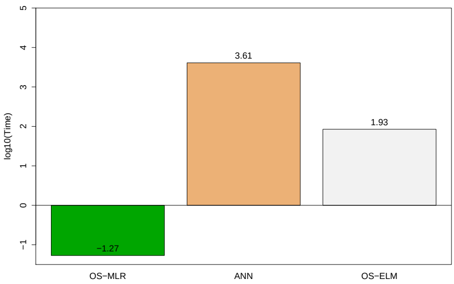
1 day ahead - 25 predictors (GFS model and past observations).

Training and Testing

- To emulate a real scenario of real-time data, we used three years (from 1995-1997) for the first model development (including parameters choice).
- From 1998 to 2001 we used the sliding retroactive validation as an emulation of real-data availability.
- We retrain the models chunk-by-chunk (daily, weekly, monthly) doing the verification (i.e. testing) of the model subsequently.

Streamflow - Monthly update





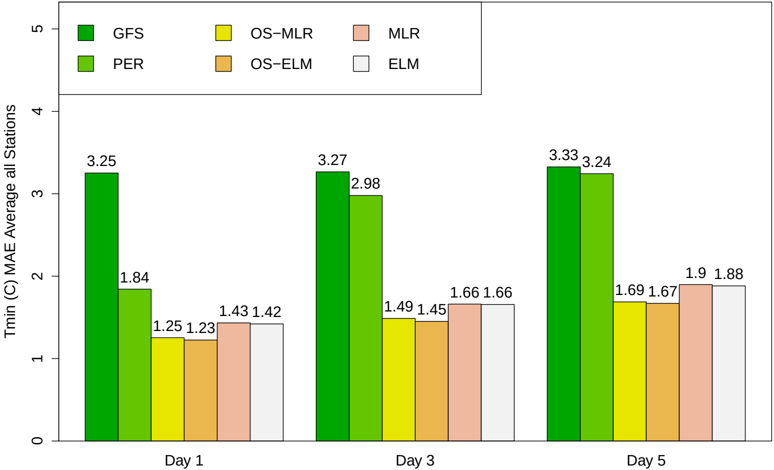
Postprocess NWP model forecasts of the daily maximum and minimum temperatures and quantitative precipitation (days 1-5).

- Observations and forecasts from 15 locations in southwestern British Columbia.
- NOAA second-generation global medium-range ensemble reforecast dataset.

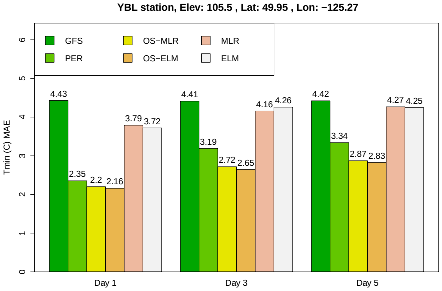
Training and Testing:

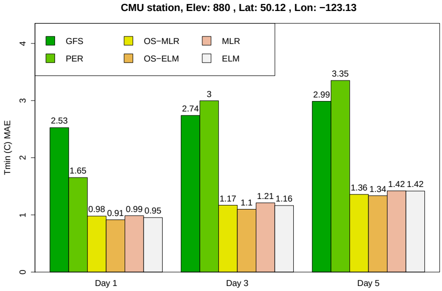
- We used two years (2004-2006) for the first model development.
- "Testing" from 2007 to 2010:
  - online approach: sliding retroactive validation as an emulation of real-data availability.
  - batch approach: only forecasting, without the emulation of real-data availability.

Tmin

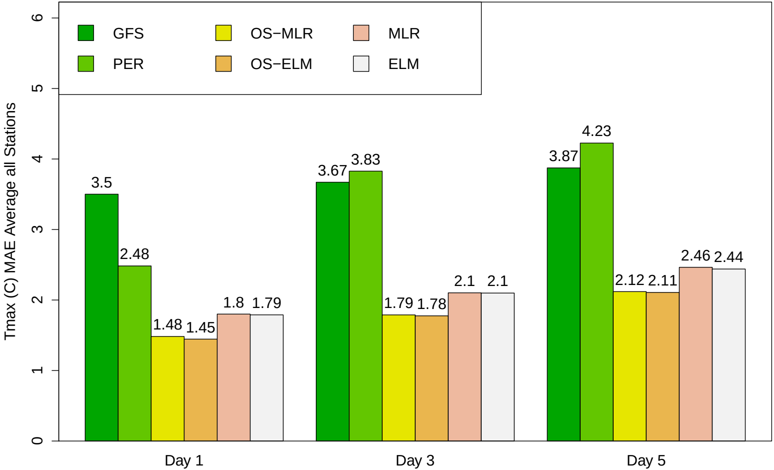

GFS
PER
OS−MLR
OS−ELM
MLR
ELM
Tmin (C) MAE Average all Stations
0
1
2
3
4
5
3.25
1.84
1.25
1.23
1.43
1.42
3.27
2.98
1.49
1.45
1.66
1.66
3.33
3.24
1.69
1.67
1.9
1.88
Day 1
Day 3
Day 5

Tmin - Online vs Batch

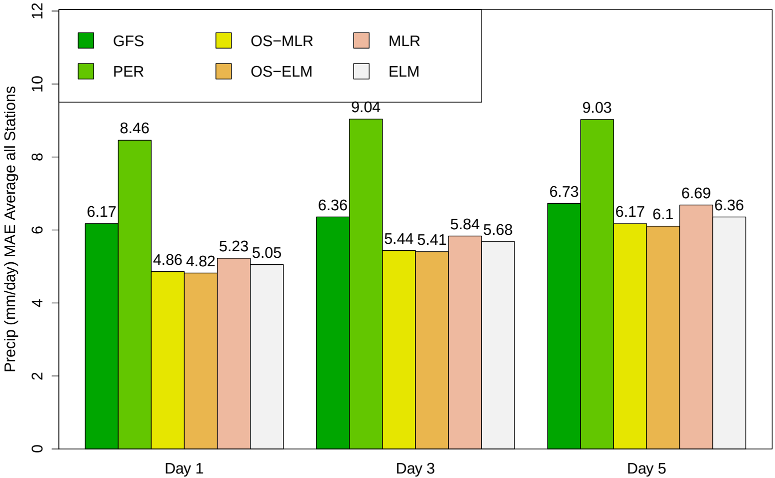




Tmax


GFS
PER
OS−MLR
OS−ELM
MLR
ELM
Tmax (C) MAE Average all Stations
0
1
2
3
4
5
6
3.5
2.48
1.48
1.45
1.8
1.79
3.67
3.83
1.79
1.78
2.1
2.1
3.87
4.23
2.12
2.11
2.46
2.44
Day 1
Day 3
Day 5

Precipitation



Conclusions

1. Indeed OS-ELM is faster (training time) than the others tested nonlinear methods
2. It can be fully automated
3. As expected there is no universal best learning method

Future work

- To test in short lead forecast problems
- Change number of hidden neurons

R package open to contributors

Forecasting with Artificial Intelligence (ForAI)
`r-forge.r-project.org/projects/forai/`