DETECTION OF ARCTIC RIVERS STREAMFLOW DRIVERS THROUGH AUTOMATIC FEATURE SELECTION

M. Zeno, M. Sangiorgio and A. Castelletti

POLITECNICO MILANO 1863

(1) ABSTRACT

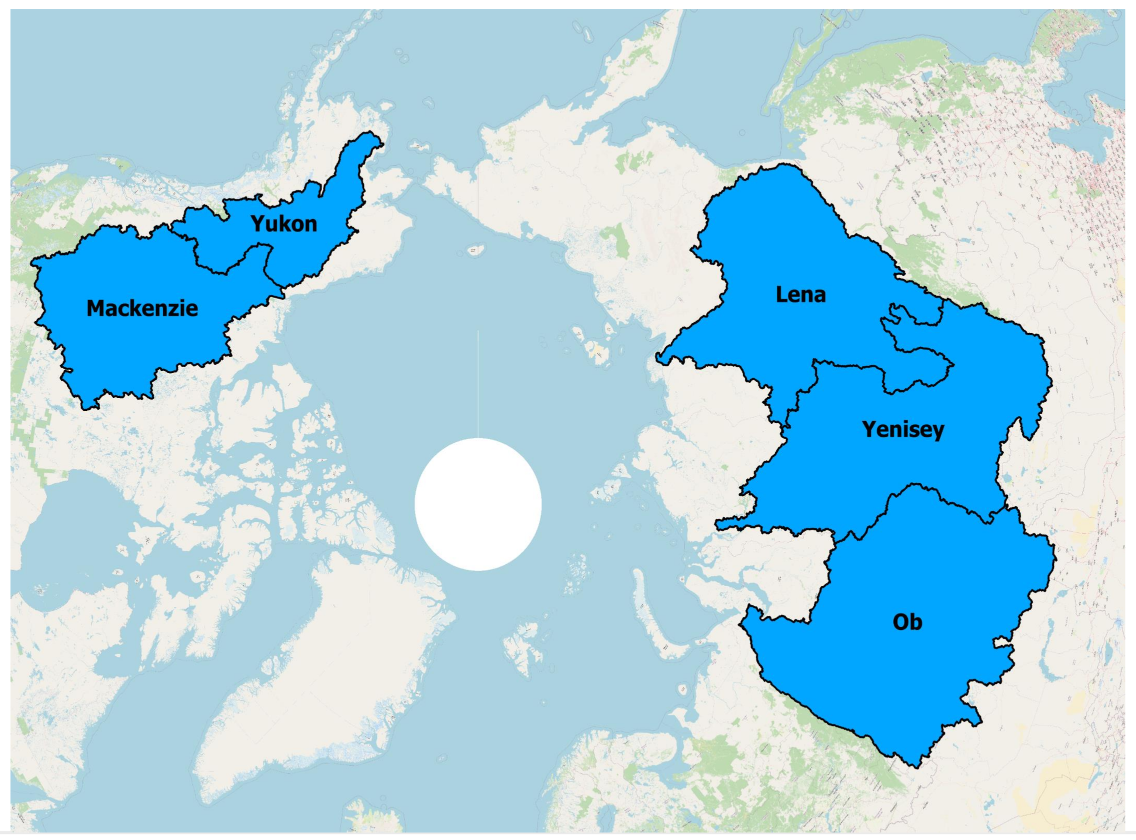
The Arctic system is undergoing profound and rapid changes due to climate change. The hydrological dynamics that characterize the region are affected as well, with rivers and streamflows over the whole pan-Arctic territory showing increasing trends in the discharge. Nonetheless, the drivers and the mechanism that characterize these dynamics are still not completely understood and therefore necessitate further investigations.

Here, we implement an automatic input variable selection technique to determine which are the most relevant drivers of streamflow dynamics for the five main Arctic rivers (Ob, Yenisei, Lena, Yukon and Mackenzie), at different time scales (daily, 10-day, monthly).

HIGHLIGHTS:

- **The autoregressive term (i.e., the past value of discharge) and the snow cover are identified as the most relevant drivers compared to the other variables considered (precipitation, temperature and teleconnections)**
- **Results seem to confirm, under a data-driven framework, that teleconnections patterns (in particular AO and NAO indexes) have an effect in determining some streamflow dynamics, such as the freshet timing (i.e the peak in river discharge due to snowmelt)**

(2) THE ARCTIC REGION AND THE FIVE RIVER BASINS



River	Length (Km)	Basin area (MKm²)	Mean discharge (m³/s)	Permafrost cover (%)
Ob	3700	2.9	13540	26
Yenisei	3487	2.5	21341	88
Lena	4294	2.4	18424	99
Yukon	3190	0.8	6596	99
Mackenzie	1740	1.8	10020	82

Table 1.
General information statistics of the 5 main Arctic river basins.

Figure 1.
Map of the 5 river basins.

(3) W-QEISS ALGORITHM RESULTS (Ob river, monthly time step)

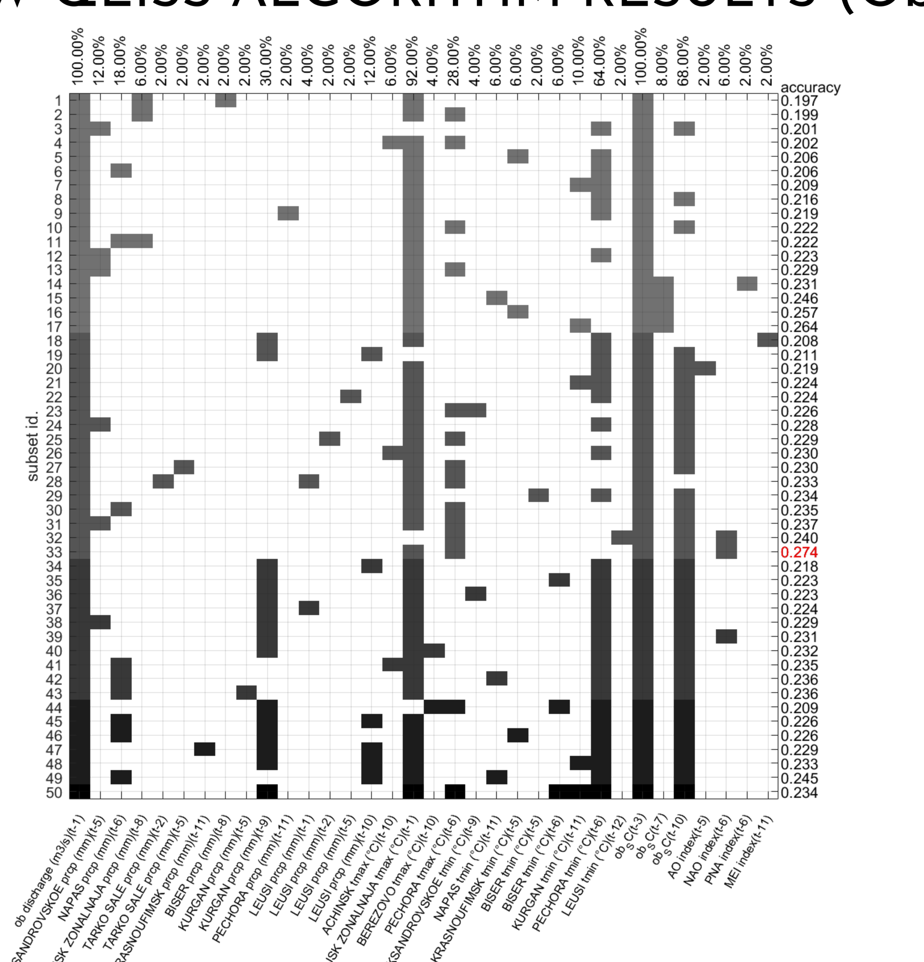


Figure 2.
Best performing alternative subsets for the Ob river basin at the monthly time step. The W-QEISS algorithm solves an optimization problem by evaluating 4 metrics: Accuracy, Relevancy, Redundancy and Cardinality

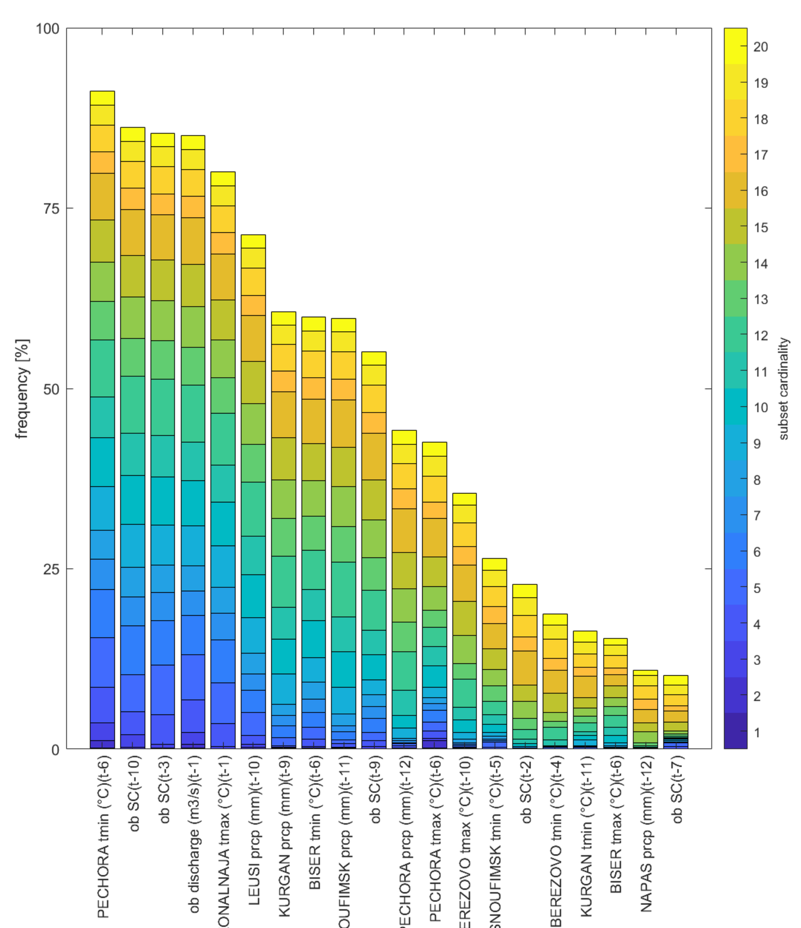


Figure 3.
Frequency bar plot for the Ob river basin at the monthly time step. The graph shows the top 20 most represented features, i.e. the features that are most frequently included within the group of subsets selected by the W-QEISS algorithm

(4) PERFORMANCES OF THE NEURAL ENSEMBLE MODEL

River	Time Step	R^2 Score	Optimal input set
Ob	Daily	0.99	Dis
	10-day	0.75	Dis + Snow
	Monthly	0.54	Dis + Snow + Meteo+ Telec
Yenisei	Daily	0.96	Dis
	10-day	0.46	Dis + Snow + Meteo+ Telec
	Monthly	0.60	Dis + Snow
Lena	Daily	0.93	Dis
	10-day	0.60	Dis + Snow
	Monthly	0.79	Dis + Snow
Yukon	Daily	0.98	Dis
	10-day	0.51	Dis + Snow
	Monthly	0.41	Dis + Snow + Meteo+ Telec
Mackenzie	Daily	0.97	Dis
	10-day	0.52	Dis + Snow
	Monthly	0.50	Dis + Snow + Meteo+ Telec

Figure 4.
Summary table of the performances of an ensemble of ELM models and the associated optimal input set used in the training process. The R^2 score decreases as the time scale passes from daily to monthly, this is due both to the increasing difficulty of the forecasting task when longer time steps are considered, and also to the 5 river dynamics. Additionally, at 10-day and monthly time scales, it can be seen that the persistency of the autoregressive tem reduces, and more exogenous variables have to be included in the optimal training set in order to obtain the best performances.

(5) PERFORMANCES FOR THE OB RIVER BASIN

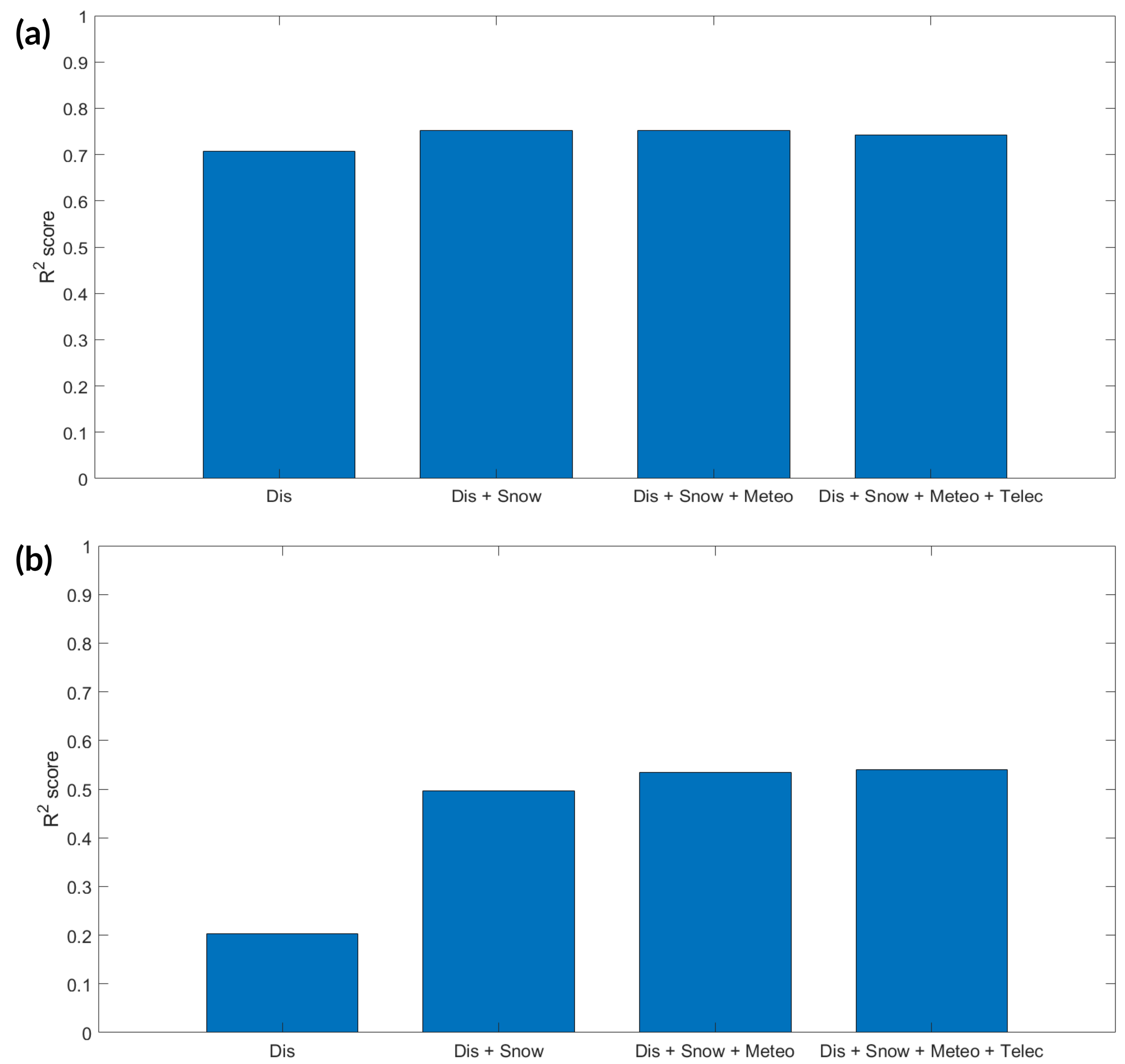


Figure 5.
Performances of the ELM ensemble for the Ob river at 10-day (a) and monthly (b) time steps. At the 10-day time step the autoregressive term has still a significant persistence. Instead, a significant increase in the model performances can be observed at the monthly time step when exogenous information, in particular snow cover, is added to the input feature set.

(6) MODEL OUTPUT TRAJECTORIES COMPARISON FOR THE OB RIVER BASIN

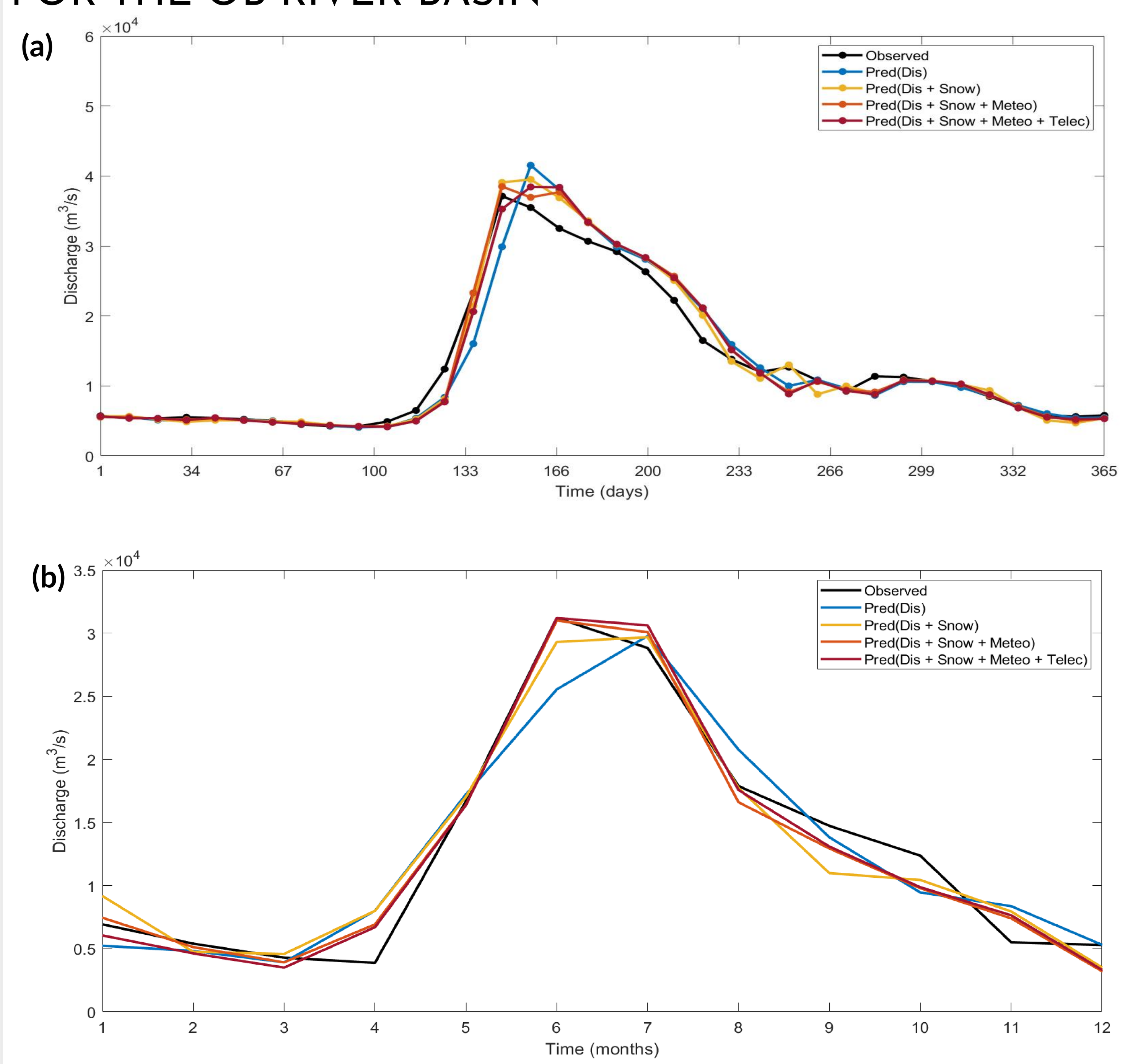


Figure 6.
Model output trajectories comparison for the Ob river basin at 10-day (a) and monthly (b) time steps. A single year has been considered as representative of the time series (2011 and 1980 respectively)

CONTACT

Mattia Zeno
mattia1.zeno@mail.polimi.it

Online presentation page

ENVIRONMENTAL INTELLIGENCE LAB

SCAN ME