1. Who are you (mini-bio) and what do you do professionally? If you are on a team, please complete this block for each member of the team.

Roman Chernenko: Machine Learning Engineer in Agreena ApS company with 4 years of experience in ML for geospatial applications, satellite imaging and precision agriculture. Also, I have 10 years of experience with C++ cross-platform development, especially in the medical imaging domain, and for embedded solutions.

Vitaly Bondar: ML Team lead in theMind (formerly Neuromation) company with 6 years of experience in ML/AI and almost 20 years of experience in the industry. Specialises in the CV and generative AI.

2. What motivated you to compete in this challenge?

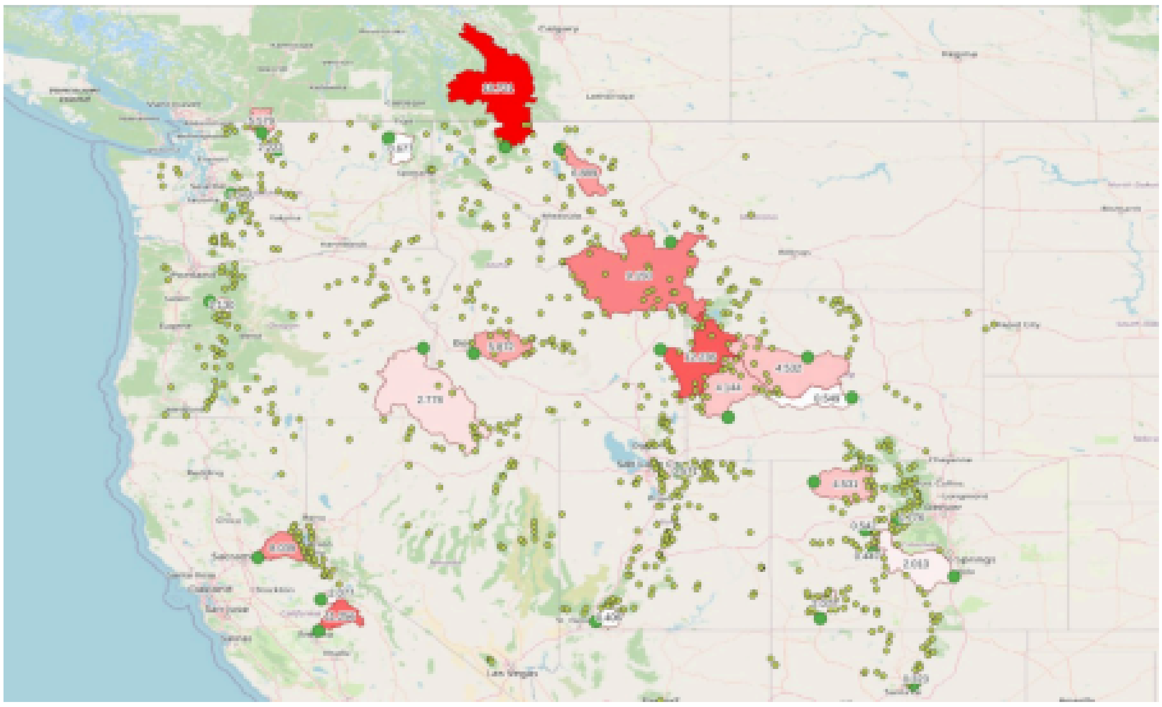
The main motivation is a complicated interesting challenge with a significant prize amount. Also, we participated in a similar challenge "Snowcast Showdown" at DrivenData 2 years ago and finished 6th.

3. High level summary of your approach: what did you do and why?

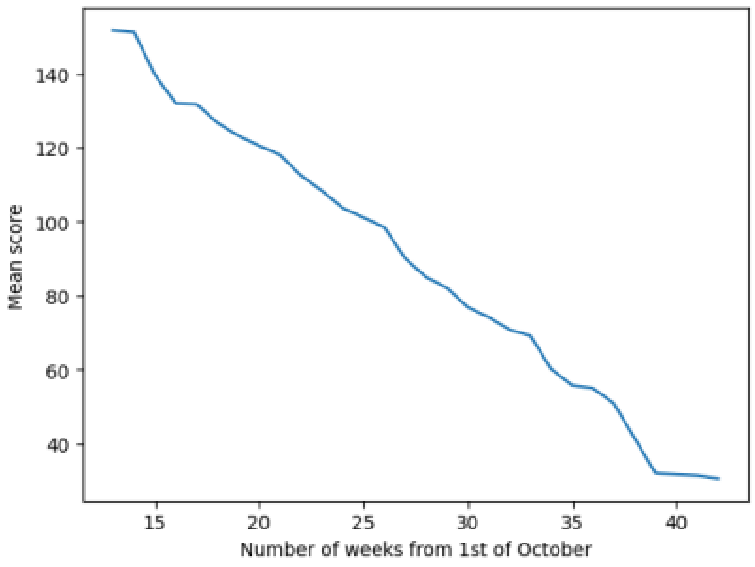
We recommend to refer our "Water Supply Final Report" competition report (report.pdf), where we made quite detailed explanations of data handling and ML decisions.

4. Do you have any useful charts, graphs, or visualizations from the process?

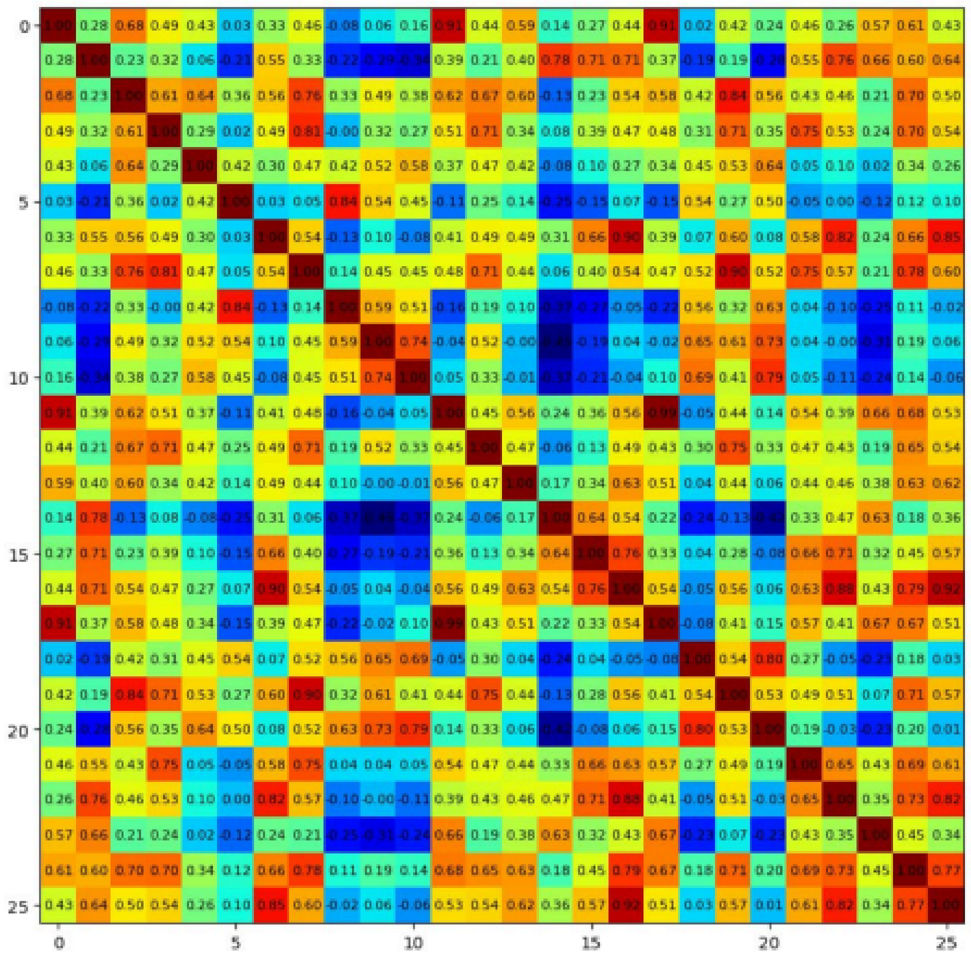
The individual addition to the score of each basin into the full score on the validation set. Basins are labeled with red polygons. Small circles - SNOTEL stations. Big green circles - sites in the training dataset.



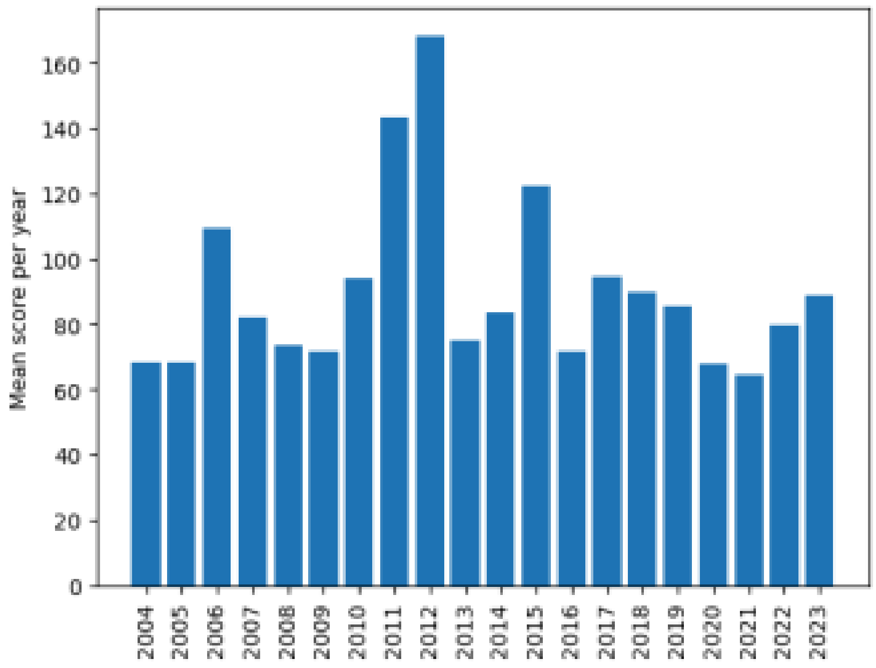
The relationship between the number of weeks from starting of the water season and the mean score on the cross-validation dataset.



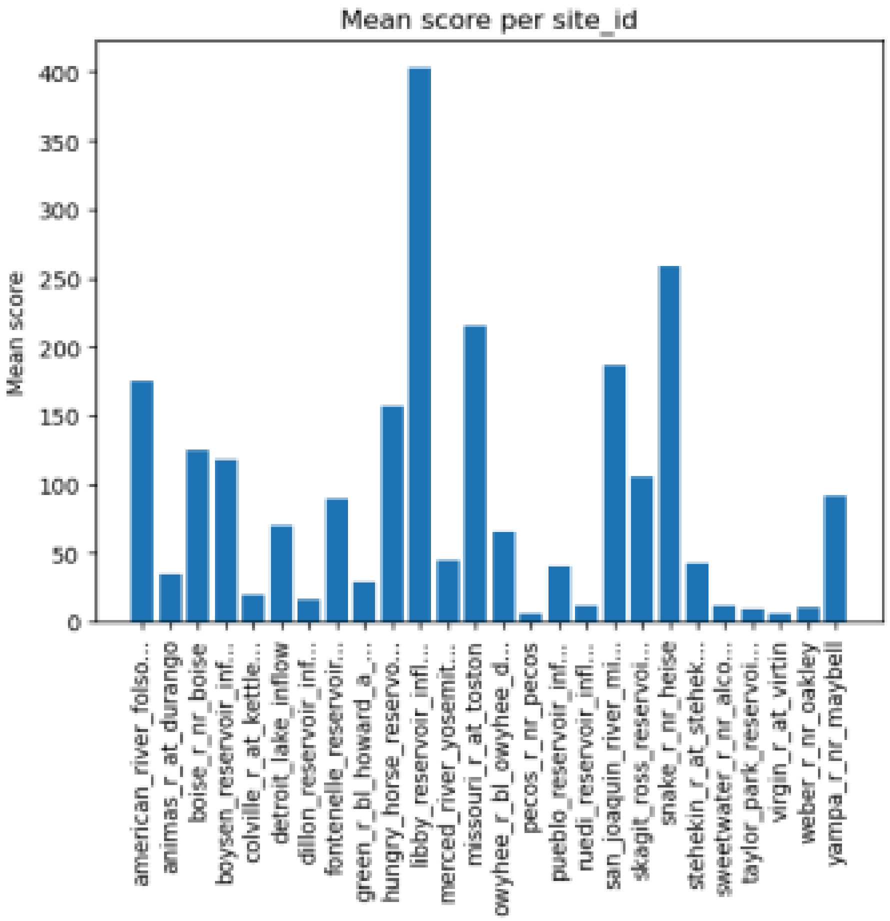
The correlation matrix of target values between all 26 sites in the training dataset.



The average score for each year during the cross-validation period.



Mean score per site for cross-validation years



5. Copy and paste the 3 most impactful parts of your code and explain what each does and how it helped your model.

The weird units transformations between USGS streamflow values to daily volume:
```
CUBIC_FOOT_m3 = 0.028316846592
ACREFOOT_m3 = 1233.48183754752
CFs_to_KAFd = CUBIC_FOOT_m3 * 3600 * 24 / ACREFOOT_m3 / 1000
ds['volume'] = ds['00060_Mean']*CFs_to_KAFd
```

Simple adding of one skipped connection from the input feature (current season water level) to the last layer significantly improves the score and training stability:
```
class MLPSumRes(MLP):
   def forward(self, x: torch.Tensor) -> Dict[str, torch.Tensor]:
       y = super().forward(x)["out"]
       y = y + x[:, 28].unsqueeze(-1)
       return {"out": y}
```

PyTorch implementation of mean quantile loss. We can train the model to predict the 10th percentile, mean value, and 90th percentile of the forecasted value simultaneously using the same model with this custom loss-function.
```
class PercentileLoss(nn.Module):
   def _percentile(self, predicted: torch.Tensor, target: torch.Tensor,
percentile: float):
       return 2*(percentile*torch.clamp(target-predicted, min=0.0) +
 (1-percentile)*torch.clamp(predicted-target, min=0.0)).mean()

   def forward(self, predicted: torch.Tensor, target: torch.Tensor):
       return (
           self._percentile(predicted[:, 0], target, 0.5) +
           self._percentile(predicted[:, 1], target, 0.1) +
           self._percentile(predicted[:, 2], target, 0.9)
       )/3.
```

6. Please provide the machine specs and time you used to run your model.

- CPU (model): `AMD Ryzen 7 5700G`
- GPU (model or N/A): N/A
- Memory (GB): `32 GB`
- OS: `Ubuntu 23.10`
- Train duration: 1h 35m
- Inference duration: 1m 20s

7. Anything we should watch out for or be aware of in using your model (e.g. code quirks, memory requirements, numerical stability issues, etc.)?

USBR reservoir inflow data has some API access limitations. To fix this, we used a cloud instance in the USA to download the data instead.

So we recommend running a data download script from the USA-located machine or cloud instance.

8. Did you use any tools for data preparation or exploratory data analysis that aren't listed in your code submission?
We used the QGIS software for geospatial data exploration.

9. How did you evaluate performance of the model other than the provided metric, if at all?

We implemented quantile loss as it was the main metric for the task. And then we use it as the loss during training, and as the metric during validation.
We did not use any additional metric.

During the testing of the hypothesis we run all tests on the several cross-validation splits to keep other splits as the hidden test-set. And only for the submission we made a full cross-validation run for all the leave-one-out splits. We believe that this helped us not to overfit the approach for the exact testset and make it useful in future years.

10. What are some other things you tried that didn't necessarily make it into the final workflow (quick overview)?

- Incorporating dropout regularization into our solution always reduced the final score.
- Additionally, our team explored some Bayesian-based approaches for the estimation of the full probability density function of the target value. But we faced many obstacles in the estimation of the posterior probability and we refused this direction.
- Although we initially planned to explore the feasibility of small transformer models, but this direction remained unexplored due to time constraints.

11. If you were to continue working on this problem for the next year, what methods or techniques might you try in order to build on your work so far? Are there other fields or features you felt would have been very helpful to have?

Possible improvements we see:

- add timeseries-like features and handle it with the convolutions or transformer based architecture
- usage of additional data from SNODAS model.
- including different kind of weather data
- including elevation data
- adding SNOTEL-like stations data from Canada

12. What simplifications could be made to run your solution faster without sacrificing significant accuracy?

An important aspect of our approach is fast inference. The model may handle all the cross-validation prediction in 1m20s on CPU, including features preparation.