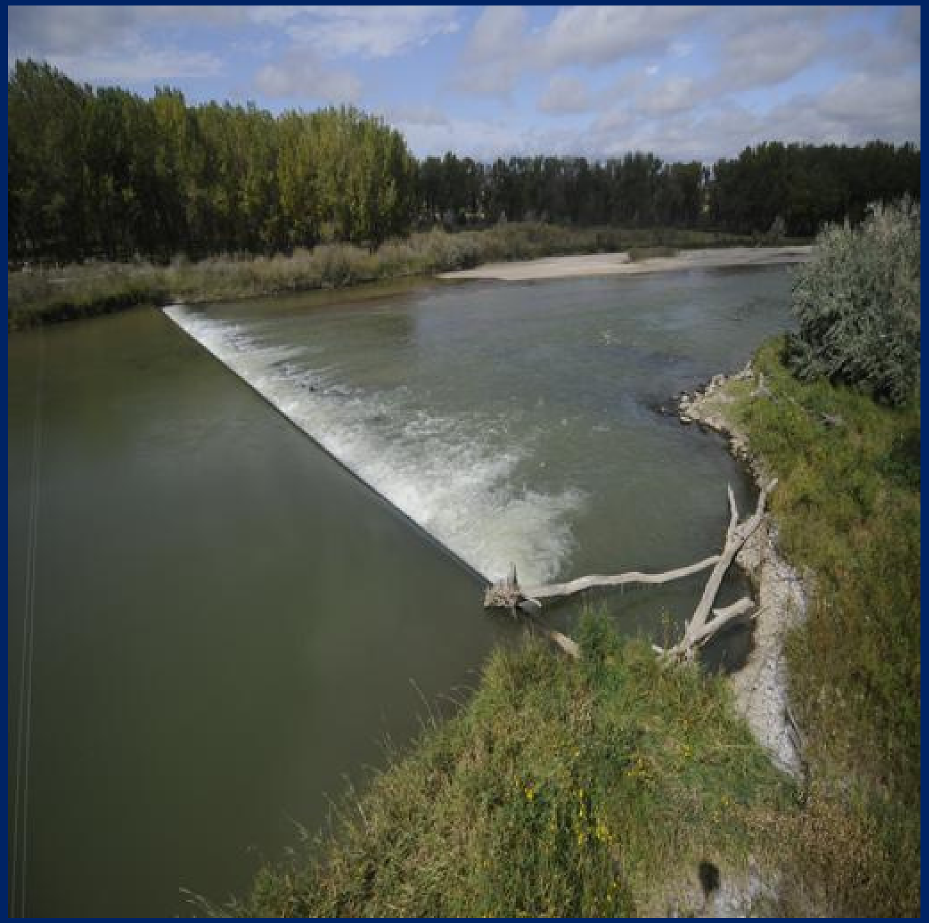
Camera-based Water Stage and Discharge Prediction with Machine Learning

TECNOLÓGICO DE MONTERREY

Objective

Obtain a model that predicts the accumulation and flow of water, this through the use of a database and images of photographs of a weir takes from the same angle over the years.



Database

The database was divided into two types of data, the generic data and hand-crafted (57 independent variables). The values that we want to predict (Stage and Discharge, dependent variables). The second database contains all the images that were used to get the information for the former dataset.

1

Understanding the problem

Studying the dataset and understanding the problem

7

Features don't add information to solve the problem at hand (Filename, Agency, SiteNumber, TimeZone, CalcTimestamp, Width, Height, "exposure", "fNumber", "isoSpeed", "shutterSpeed"). Were removed

2

Features that seem to have no correlation with the stage and discharge values or any other feature (areaFeatCount, WwCurveLineMin). Were removed

17

Features that are highly correlated with each other

Dataset Analysis

- For the data we only considered the values equal to zero to be outliers, because in the images it would seem that there is flow of water.
- Data was taken each hour, but there are gaps.

Stage range: 0 to 6.4
Stage without zeros: 1.34 to 6.4
Discharge range: 0 to 7920
Discharge without zeros: 6 to 7920

Time series analysis

- Data seems to have a correlation with time.
- Each year there seems to be a repeating pattern for the values of stage and discharge.
- KPSS test tells us the data is non-stationary, so it appears that the data probably has seasonality and a trend.
- Stage and discharge values can vary a lot throughout the years.

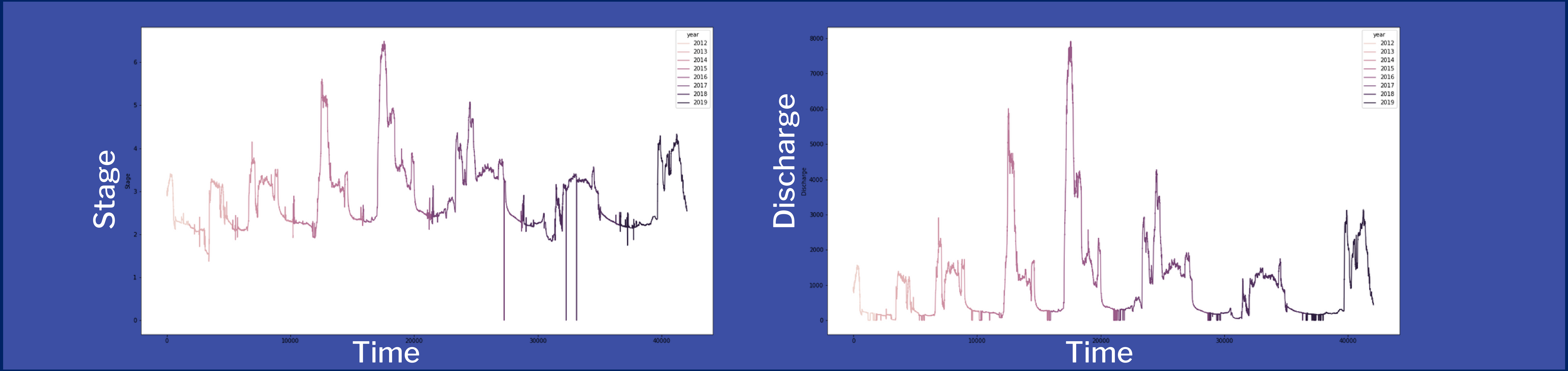


Fig. Time series of Stage and Discharge data through the years

CV(Coefficient of variation) = Std / mean
SD = Stadandard deviation

Dataset split

- Standard deviation is very small for stage and discharge, for all data and divided by months.
- There is a pattern repeated in the data each year.
- From this we decided instead of doing the common **dividing of data by a 70% 30%** randomly, to instead **divide by years** (ex. train 2012 to 2015 and 2016 to 2017 testing).
- if the data is grabbed randomly from all the years the train model would already contain that data for testing.

	Stage	Stage without zero	Discharge	Discharge without zero
SD	0.80	0.80	1,192.27	1,200.88
Variance	0.65	0.65	1,421,513.21	1,442,125.45
CV	0.28	0.28	1.23	1.18

Table. Stage and discharge sd, variance and CV for the whole dataset

	Stage	Discharge
Max SD	1.33	2,197.79
Mean SD	0.39	558.25
Max Variance	1.77	4,830,301.55
Mean Variance	0.29	729,272.88
Max CV	0.41	1.52
Mean CV	0.13	0.63

Table. Stage and discharge sd, variance and CV by month of all the years

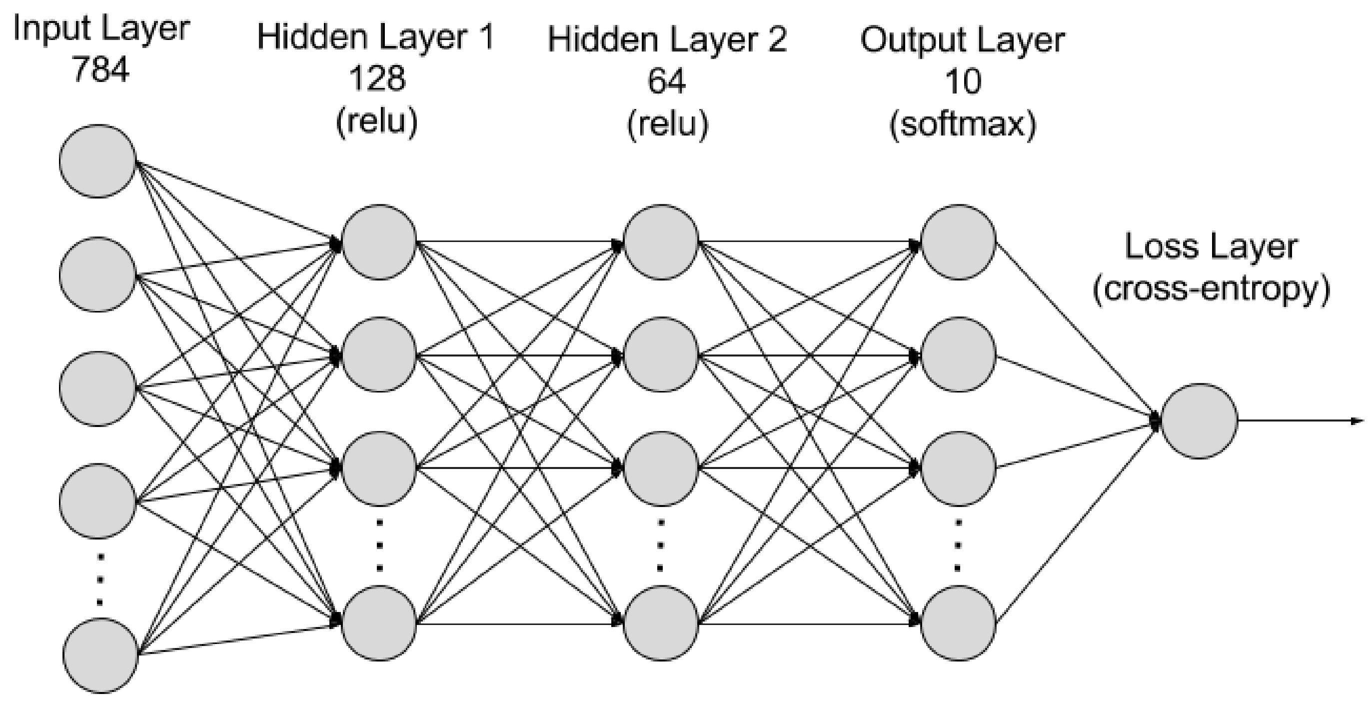
2

Solving the problem

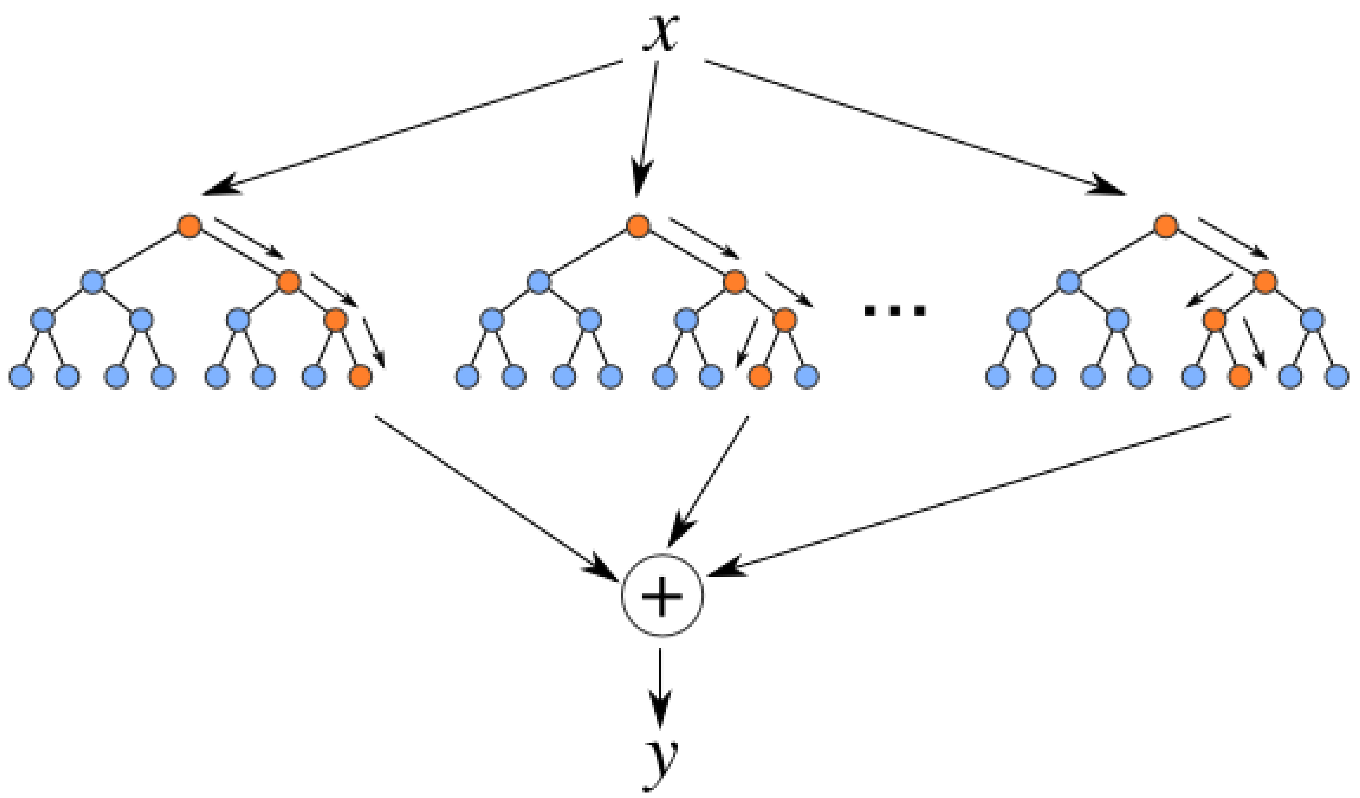
Solving the problem

Dataset split: train (2012 to 2017), test (2018 to 2019).

- The models used were an MLP Regressor, KNN Regressor, Random Forest Regressor.
- All the generic and hand-crafted variables were used (46 variables).
- Standard scalling was used for the input data.
- 1 output variable (stage).



Architecture of a MLP



Architecture of a Random Forest

Solving the problem

CNN

- Dataset split: train (2012 to 2016), validation (2017), test (2018 to 2019).
- Image size: 512x512x3

- The CNN used was a pretrained Resnet50 for transfer learning.
- The top part of the images was removed with black to reduce noise for the model.

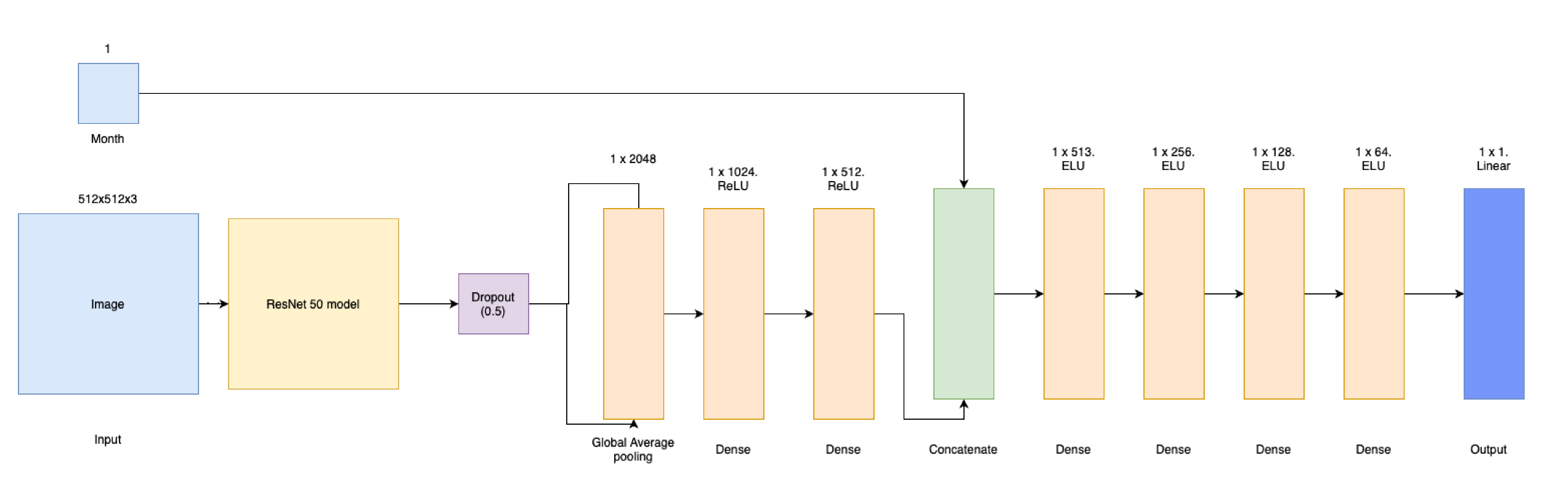


Fig. CNN model used

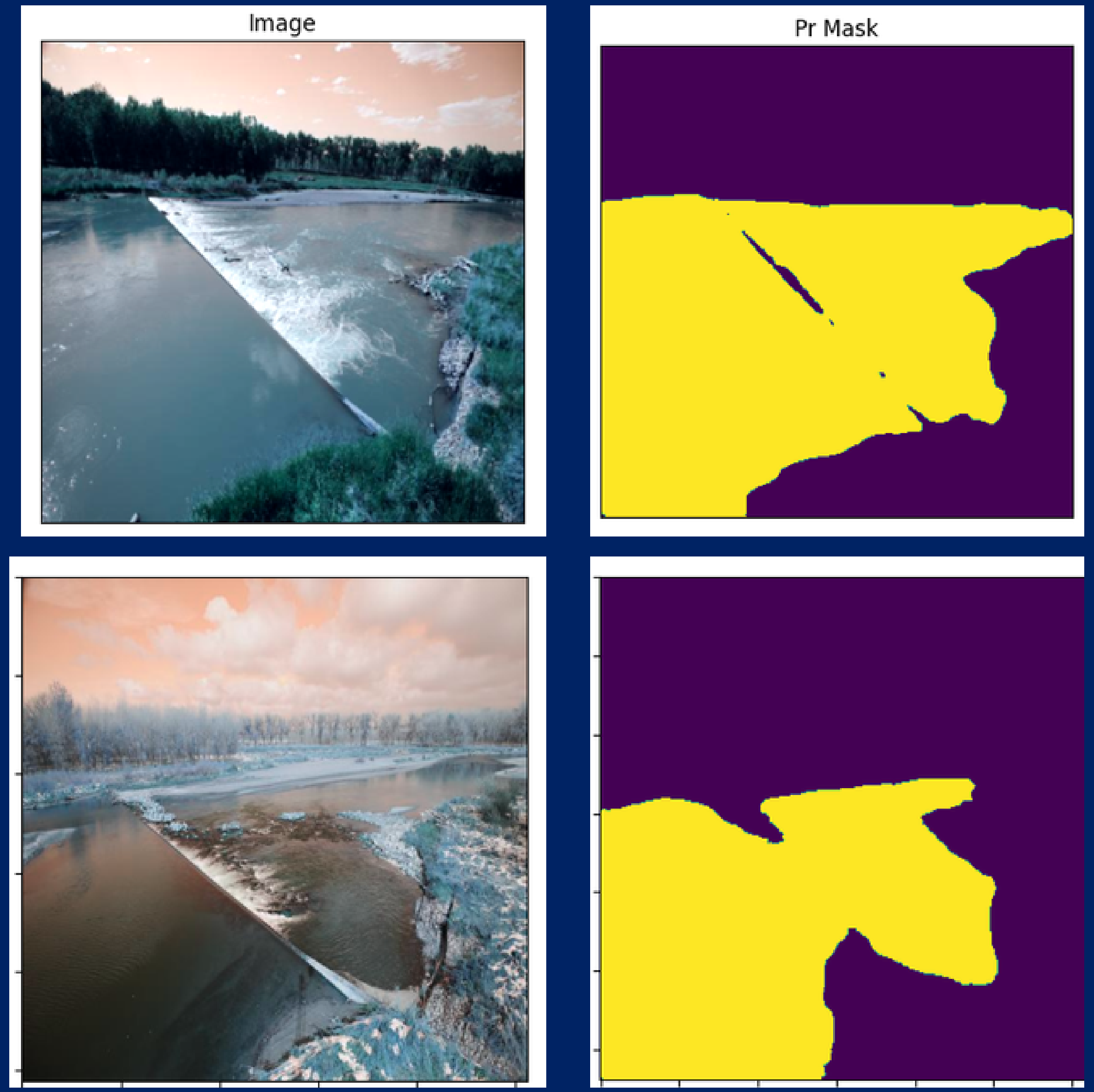
Fig. Image with top part removed

Solving the problem

Segmentation-MLP

- Dataset split: train (2012 to 2016), validation (2017), test (2018 to 2019).
- Image size: 320x320x3

- Trained U-Net with seresnext101 backbone.
- Data augmentation was done to the images .
- From the binary mask, river width and river area is calculated.
- There is a paper that has already calculated the stage based on the width of the river and bathymetry.



Good segmentation

bad segmentation

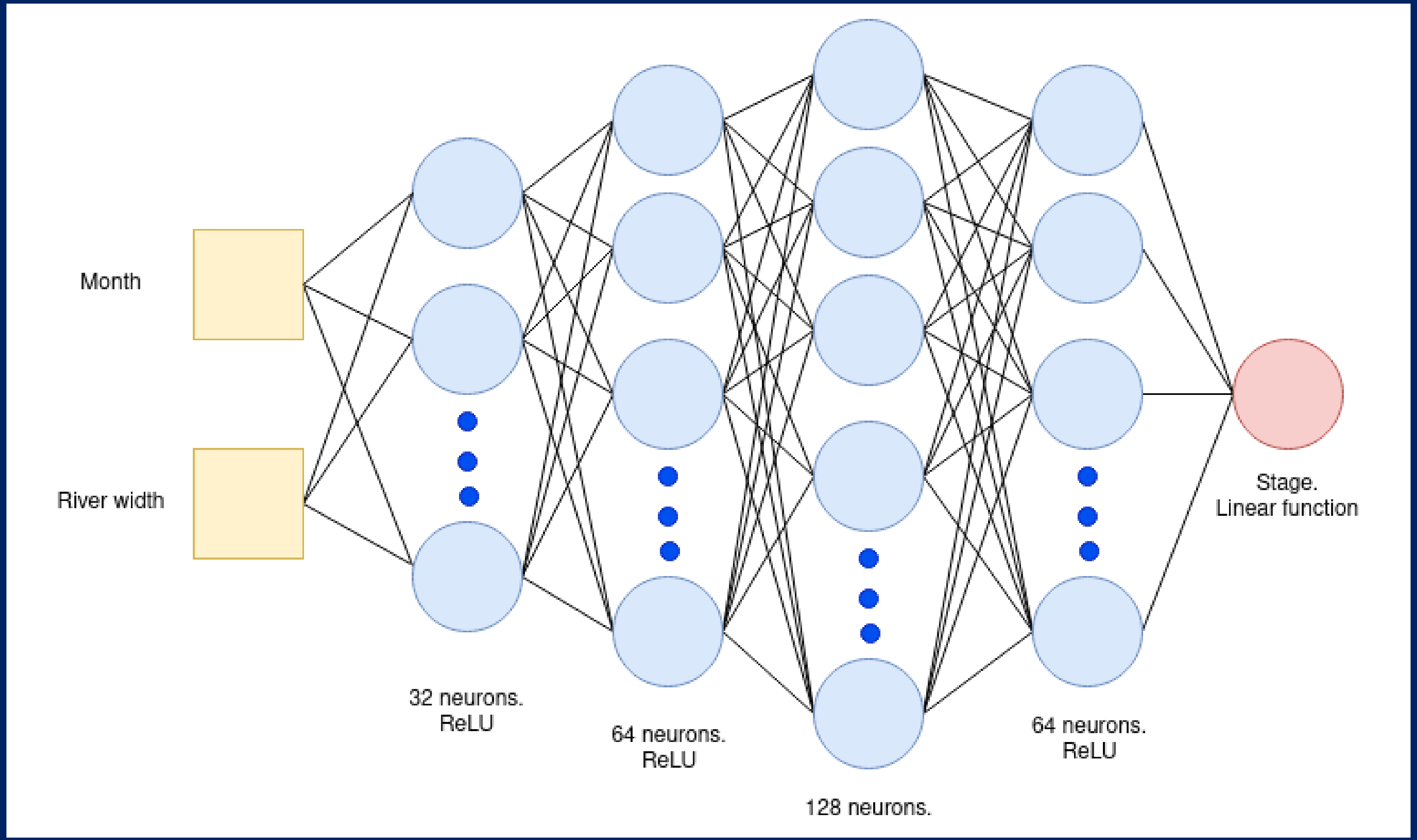


Fig. MLP model for the segmentation data

Solving the problem

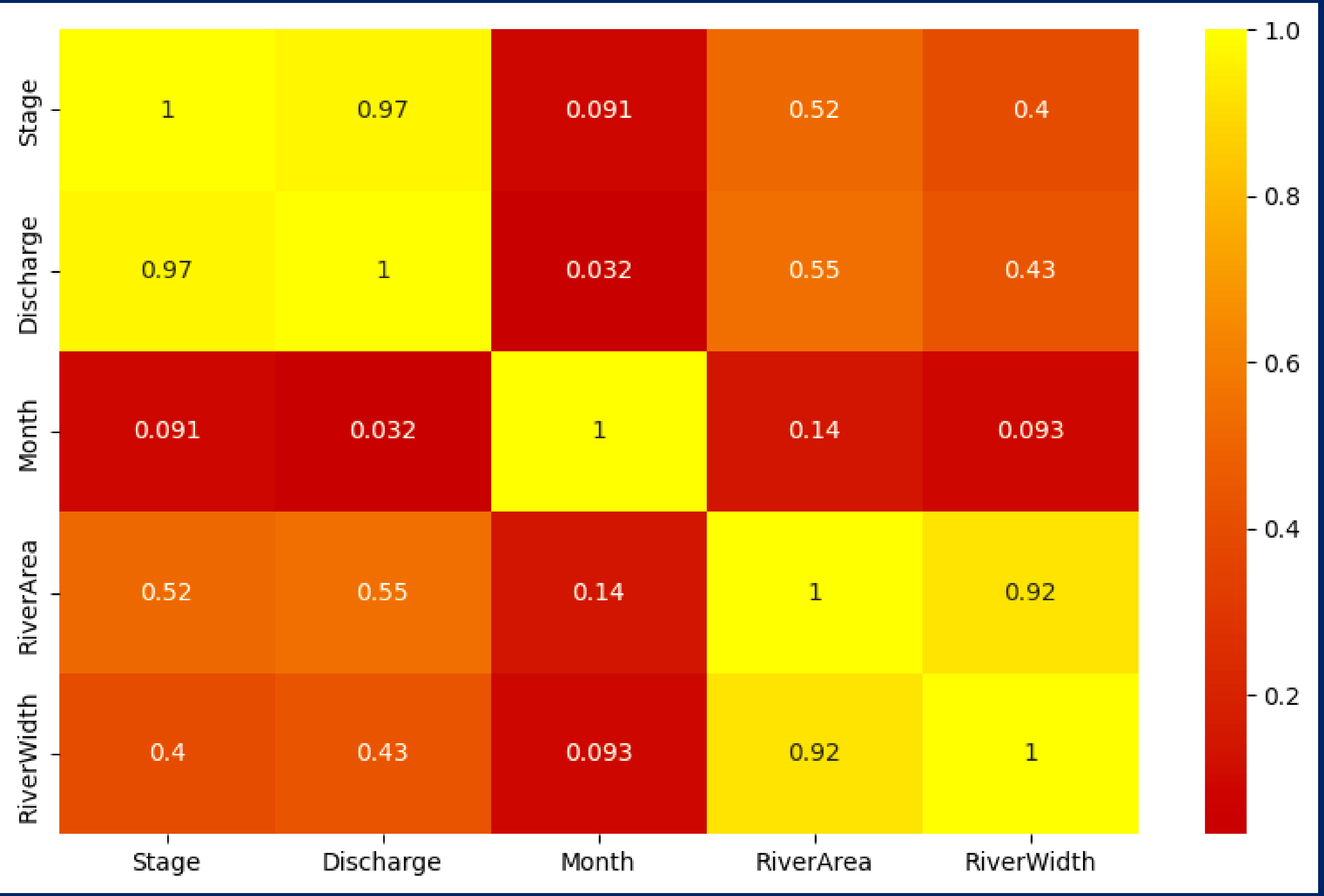


Fig. Pearson correlation graph of the segmentation results

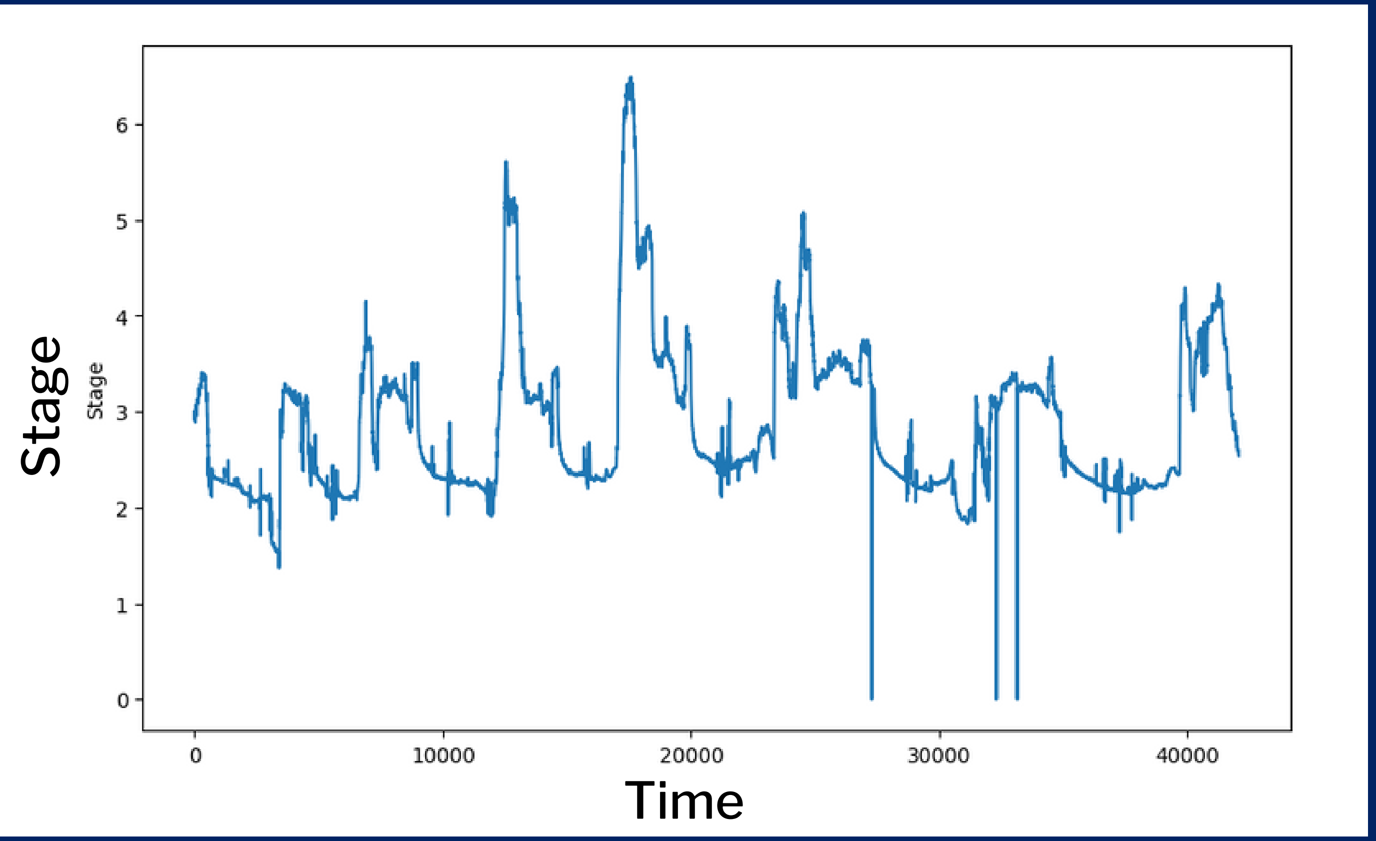


Fig. Time series of stage.

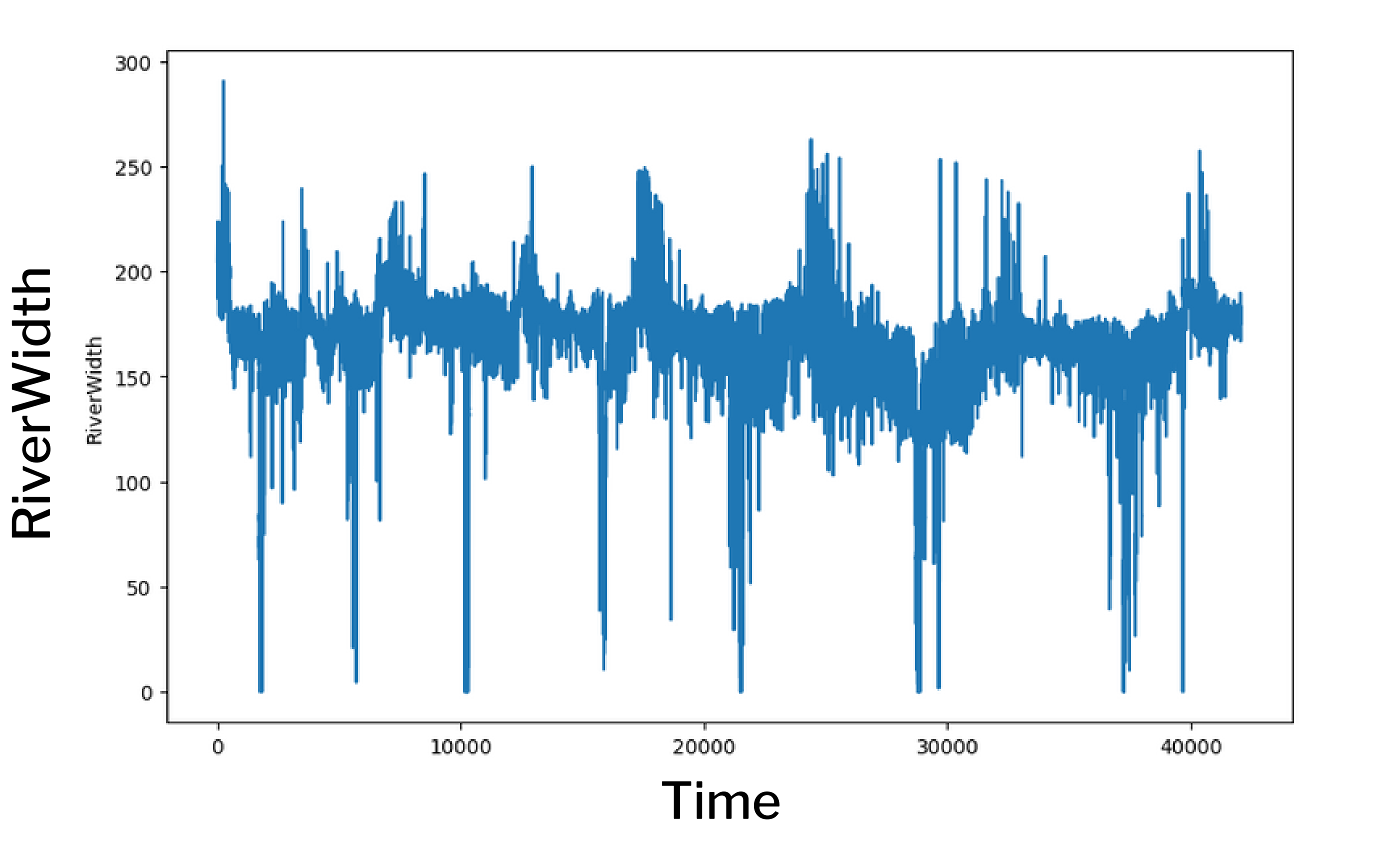


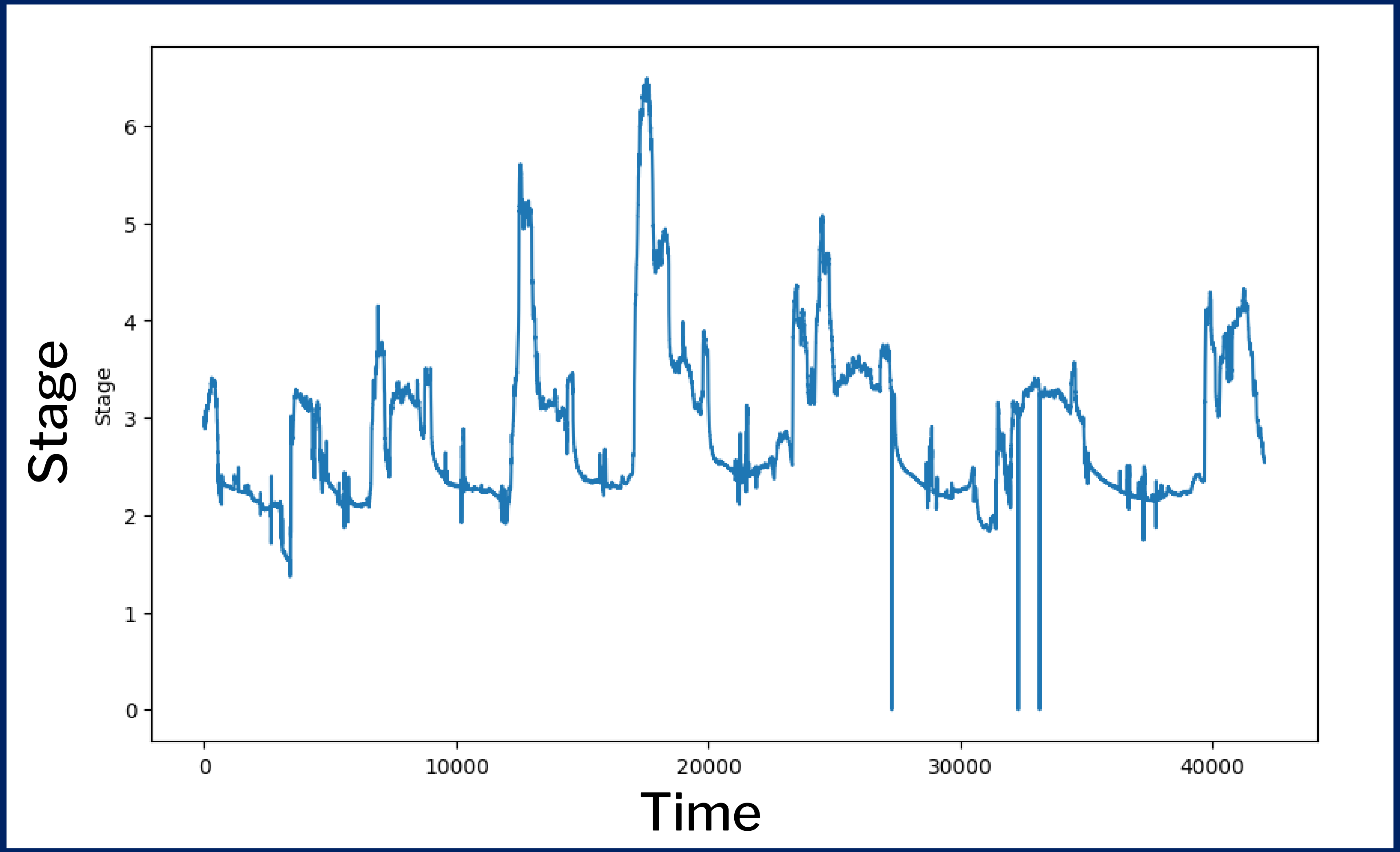
Fig. Time series of River width.

3

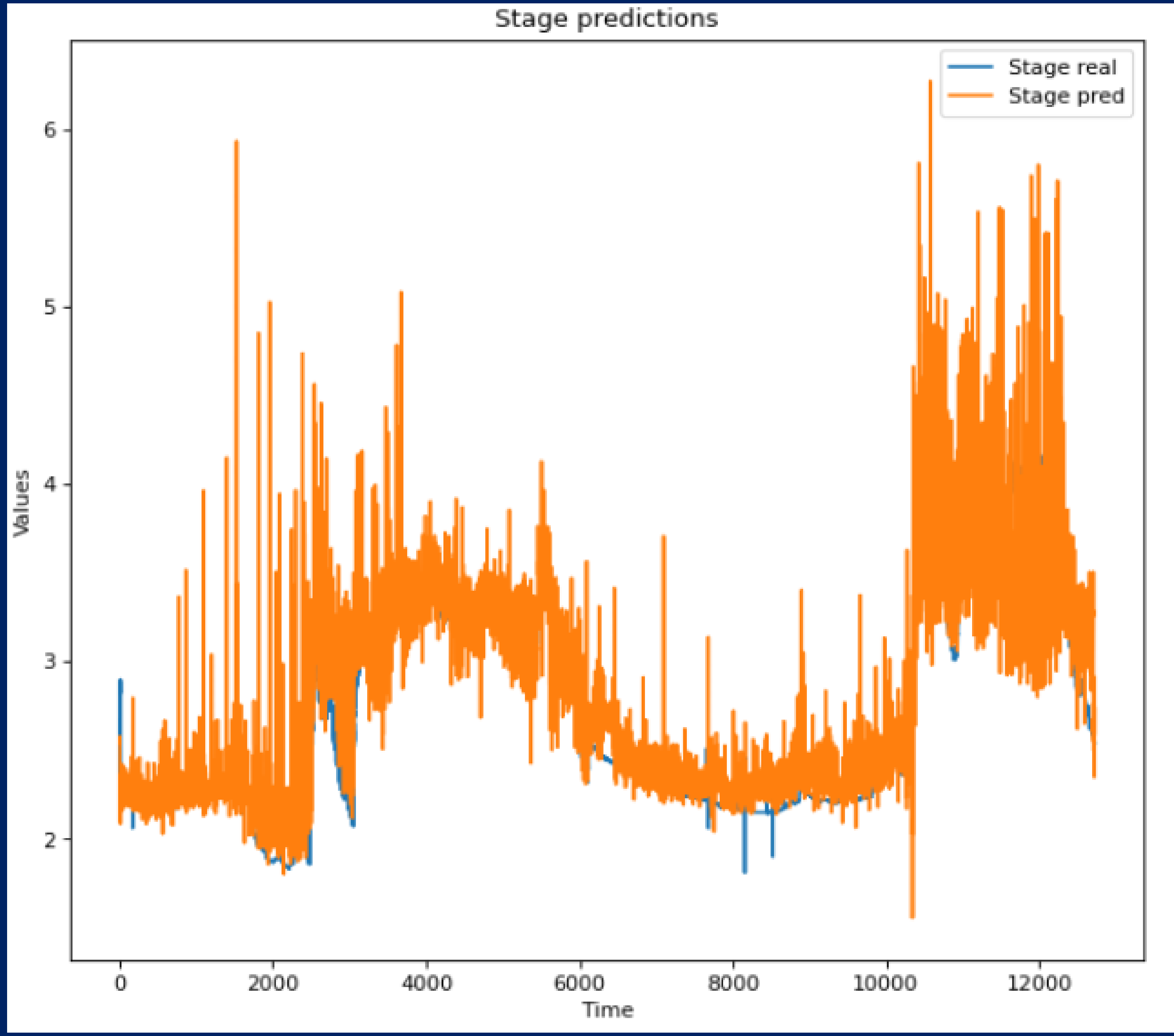
Results

Results

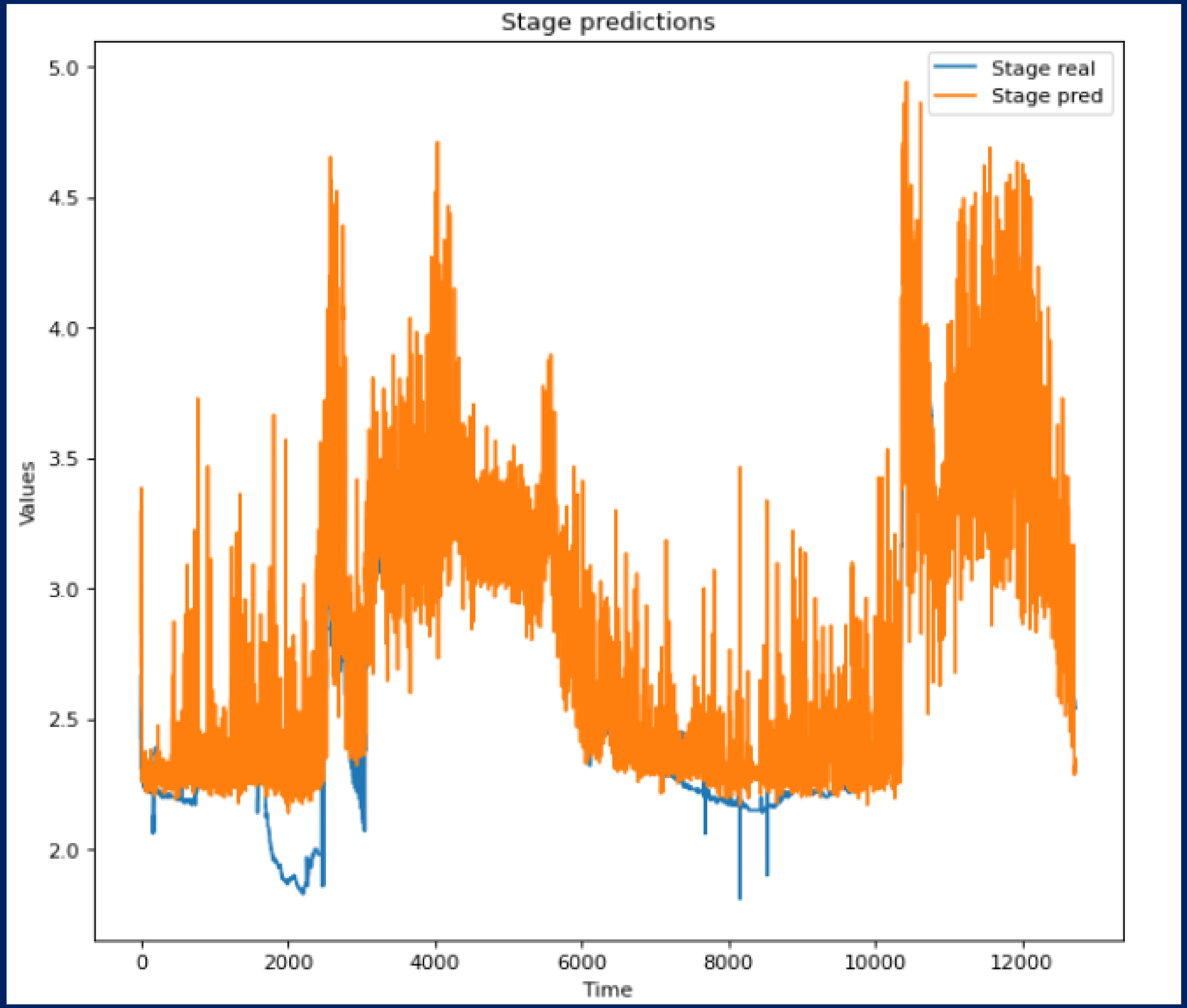
	CNN	MLP Regressor	MLP Regressor with Segmentation	KNN Regressor	Random Forest Regressor
R^2	0.8043055231	0.6911826503	0.4220925667	0.54566232	0.4919470624
MSE	0.07642631698	0.1206052058	0.2258010654	0.1774365638	0.1984144645
MAE	0.1919919813	0.2223266992	0.2790626533	0.2821863754	0.3127794728
MAPE	7.221717524	7.80759087	0.09730067528	0.1064906306	0.1214667585

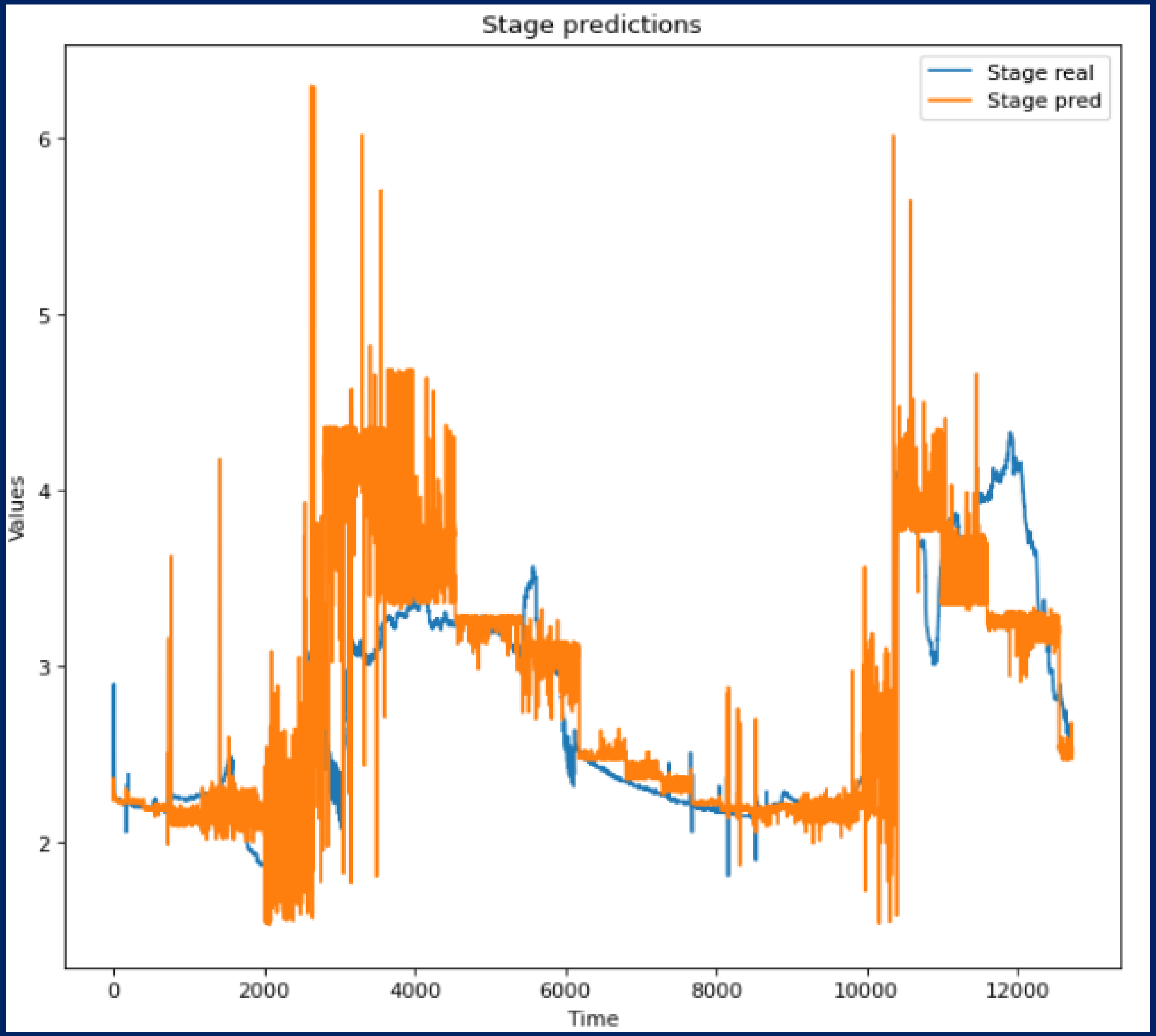


Time series of stage of the ground truth values.



MLP Regressor





CNN

MLP Regressor with data from Segmentation

4

Conclusion

Conclusion & Future work

- Create 100 masks from the river at hand to prove hypothesis.
- Calibrate the cameras to get the conversion from pixels to cm, ft, etc... to get a more precise model.
- Put the cameras at a higher position, to reduce reflection and get a better view of the river.
- Find a place near the river where is flat to get better segmentation.
- Seasons (winter).
- Artificial river in The University's civil engineering department (Université de Sherbrooke). Video of the river, This river can be switched on and off (Tom Scott)

Contributions

- **Andres Eduardo Nowak de Anda:** Dataset analysis, dataset models (Random Forest Regressor, KNN Regressor), image analysis, image models (CNN, MLP Regressor with Segmentation), Report.
- **Isaac Emanuel García González:** Dataset analysis, dataset models (MLPRegressor, KNN Regressor), image models (CNN), Report.
- **Samuel Alejandro Diaz del Guante Ochoa:** Dataset analysis, image analysis, image models (CNN, MLP Regressor with Segmentation), Report.
- **Ernesto Lopez Villareal:** Dataset analysis, image analysis, image models (CNN), Report.

References

- Eltner, A., Bressan, P. O., Akiyama, T., Gonçalves, W. N., & Marcato Junior, J. (2021). Using deep learning for automatic water stage measurements. Water Resources Research, 57, e2020WR027608. https://doi.org/10.1029/2020WR027608
- Chapman, K. W., Gilmore, T. E., Chapman, C. D., Mehrubeoglu, M., & Mittelstet, A. R. (2020). Camera-based Water Stage and Discharge Prediction with Machine Learning. Hydrology and Earth System Sciences Discussions, 2020, 1–28. doi:10.5194/hess-2020-575