CALIFORNIA DEPARTMENT OF WATER RESOURCES

Utilizing TensorFlow for Training an ANN Model to Estimate Salinity and X2 in the Sacramento-San Joaquin Delta

Hamed Zamanisabzi

Malinda Wimalaratne

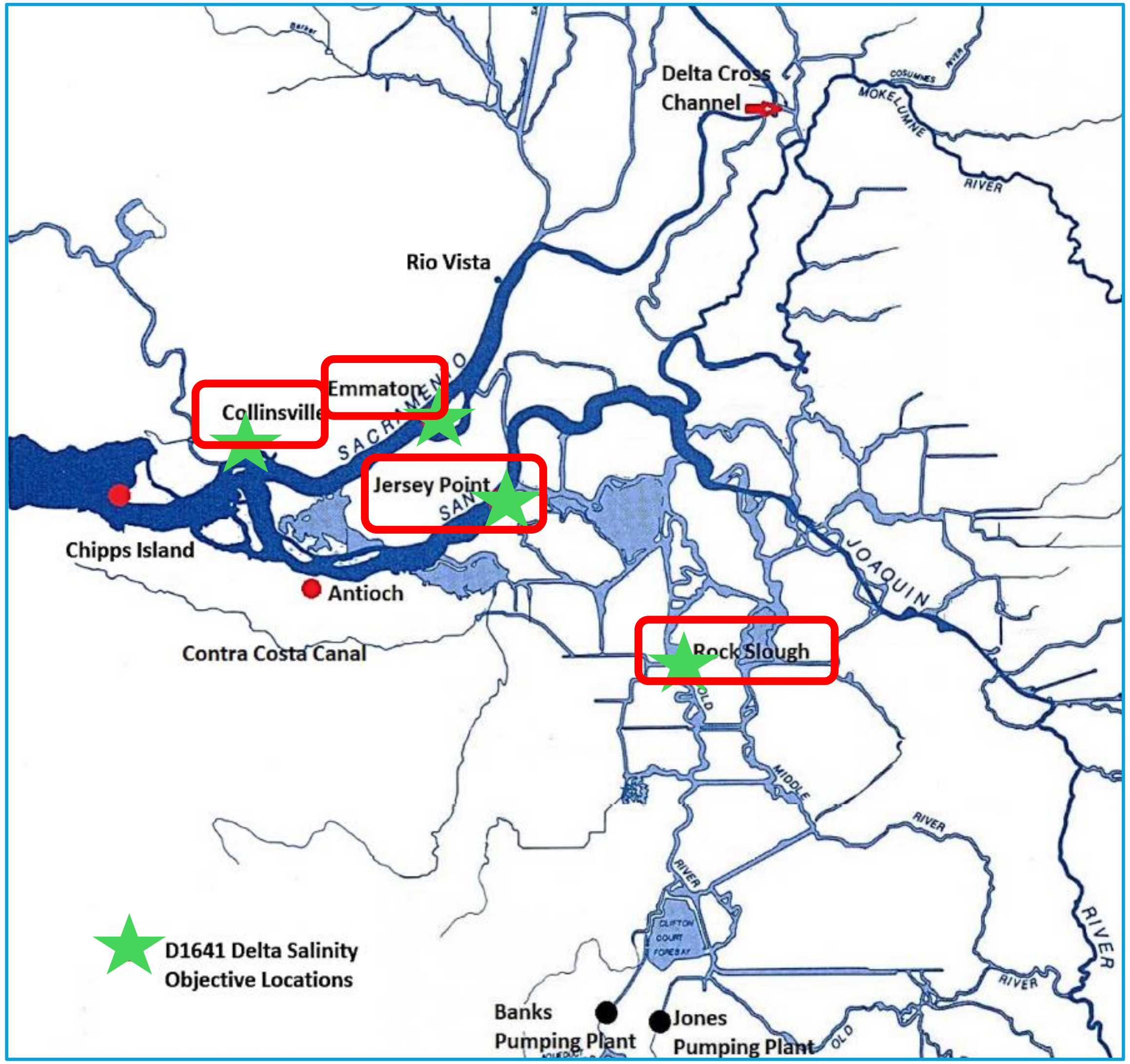
Objective

Update Delta Flow – Salinity relationship in CalSim

Why ANN in CalSim3?

- Maintain regulatory requirement such as, D1641 water quality standards

- Due to the limitations of hydrodynamic simulation within CS3, an artificial neural network (ANN) was integrated as a surrogate for the Delta Simulation Model 2 (DSM2) to estimate salinity and X2 parameters.

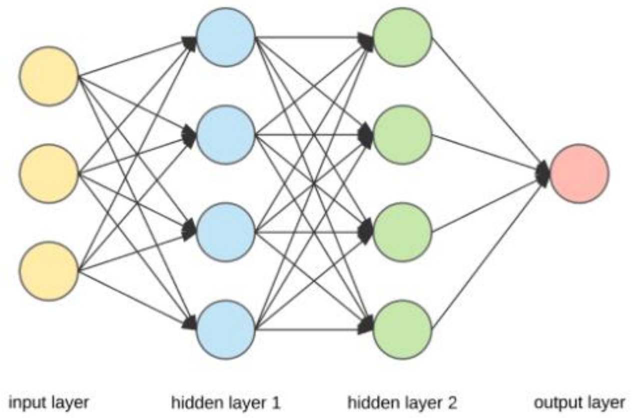


* Jayasundara et al. Journal of Water Resources Planning and Management 146.4(2000)

Introduction

What is an Artificial Neural Network (ANN)?

Computational Model inspired by the structure and functioning of the human brain



Training ANN for decision/estimation

Training [Inputs(CalSim3)> target(DSM2- Salinity in Electrical conductivity (EC)/**X2***)]

*** X2:** Location of the 2640 umhos/cm (2PPT) EC isohaline measured in km from Golden Gate

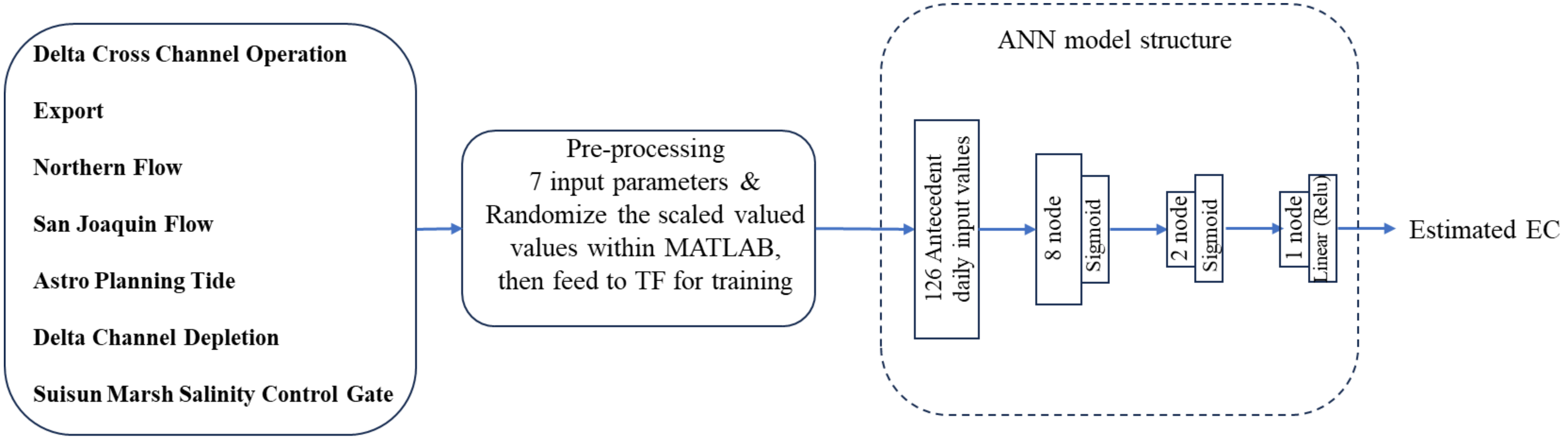
Motivation

Transferring from existing MATLAB to TensorFlow

1. TensorFlow (TF) is open-source
2. Large Python community among CS3 users
3. Public availability of TF on GitHub, and continual developments
4. Version upgrade challenges for MATLAB (ML) tools with other tools such as HEC-DSS

Key training parameters/assumptions utilized by MATLAB and TensorFlow for training ANN models to estimate the Salinity (EC) and X2

Parameter	MATLAB	TensorFlow
Antecedent conditions	118 days including current day	Same as MATLAB
Predictors/ Training Inputs	EC - 7 inputs X2 – 3 inputs	Same as MATLAB
Target - Delta Salinity Objective Locations	4 Salinity objective locations X2 km	Same as MATLAB
Number of the layers	Three layers	Same as MATLAB
Activation functions on the layers	Logsig, logsig, purelin	Same as MATLAB
Training and Validation Data Selection	80% Training, 20% Validation (Randomized)	Same as MATLAB



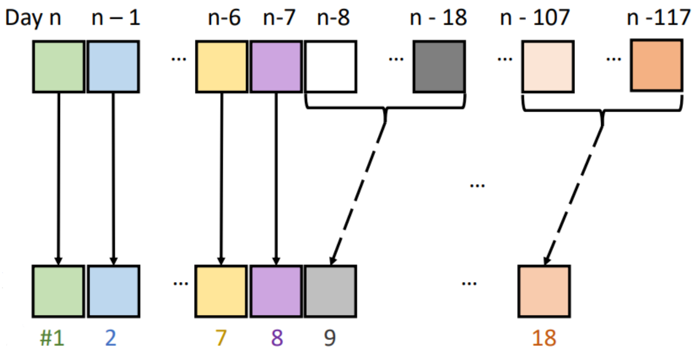
Schematic of developed TF ANN model to estimate the EC (sources: Adapted from Jayasundara et al., 2020, Qi et al., 2021)

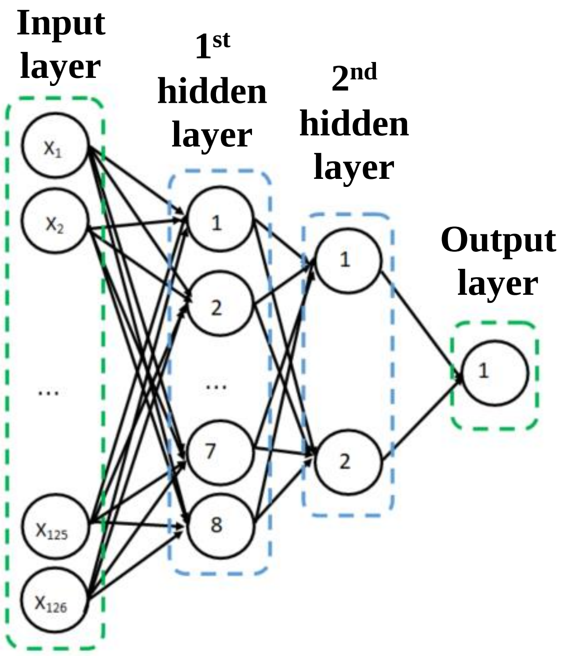
Considering historical observation for Pre = Processing

(10*11+7+1 = 118days)

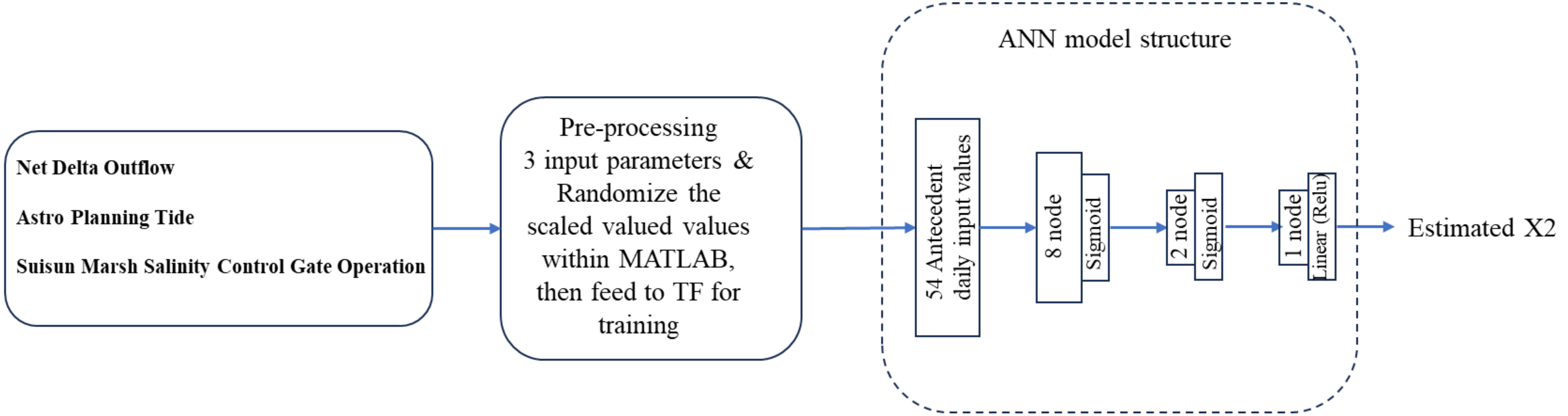
Array size per variable (1+7+10 = 18)

For EC: Total array size for 7 variables 7*18=126 ;



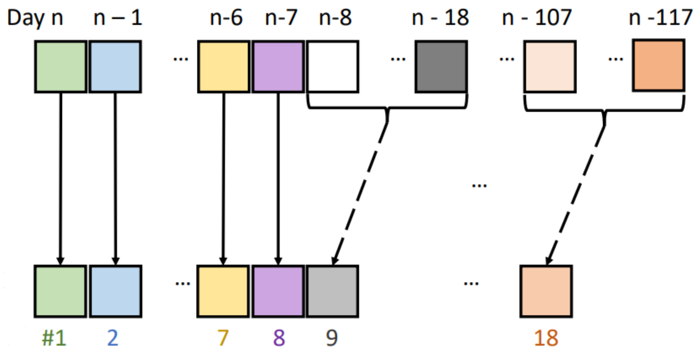


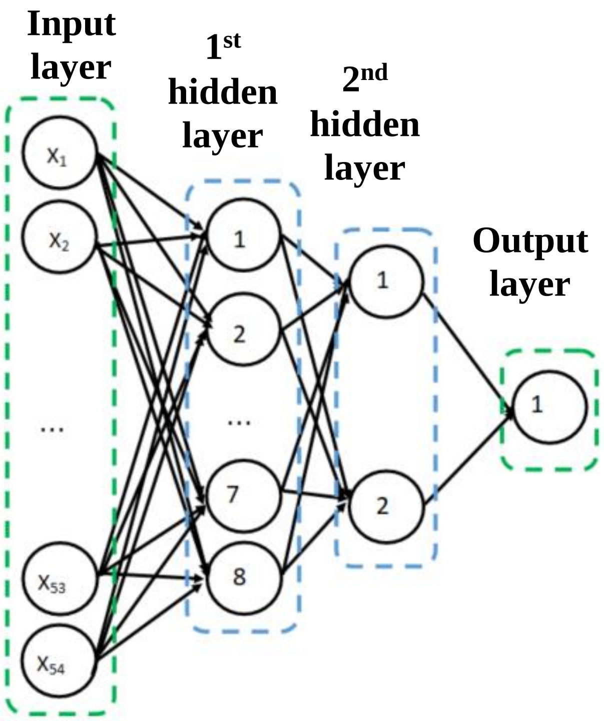
Schematic of ANN Structure for EC Prediction



Schematic of developed TensorFlow ANN model to estimate the X2.

Considering historical observation for Pre = Processing
(10*11+7+1 = 118days)
Array size per variable (1+7+10 = 18)
For X2: Total array size for 3 variables 3*18=54





Schematic of ANN Structure for X2 Prediction

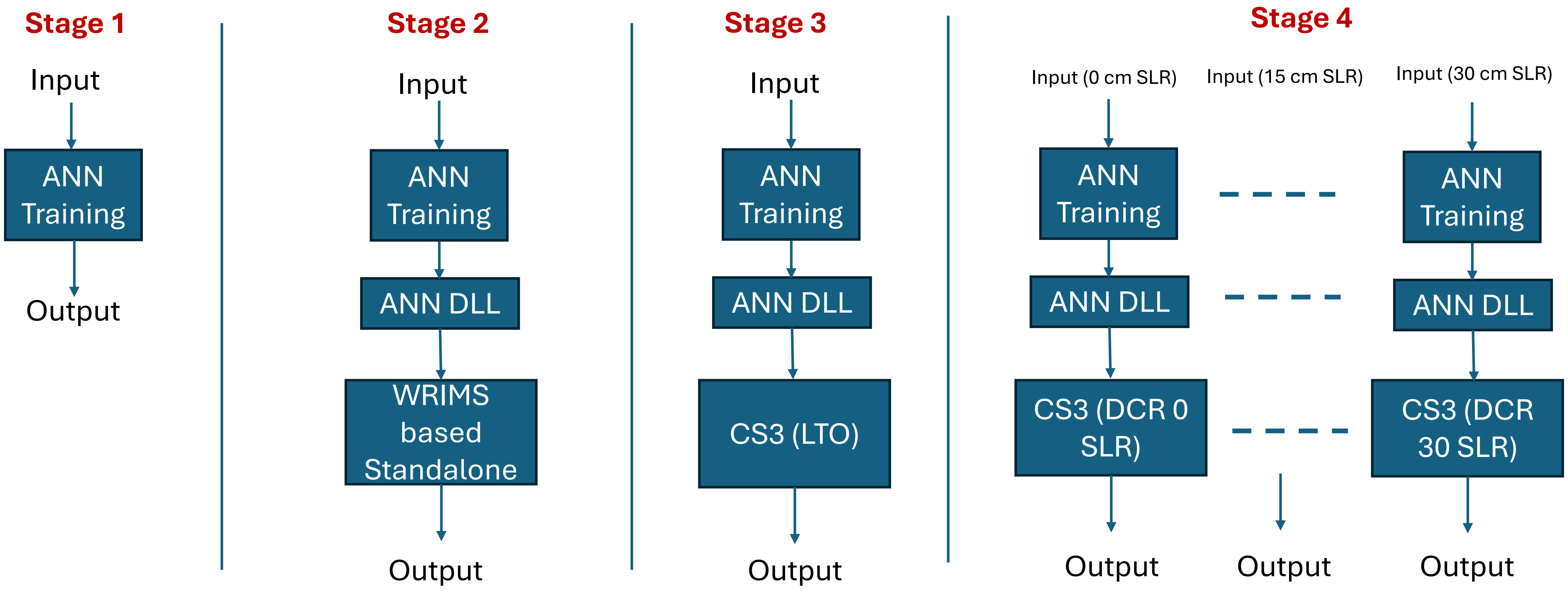
Performance Comparison Methodology, Statistical Performance Criteria are RMSE and R-Squared.

Stage 1: An Excel-based tool was developed to simulate the computational processes of models across both platforms.

Stage 2: A standalone tool was utilized to input datasets and employ the trained ANN DLLs from both platforms to estimate EC and X2 values for comparison.

Stage 3: A complete CS3 model run was conducted

Stage 4: Complete Process Flow Performance Comparison based on three SLR CS3 studies



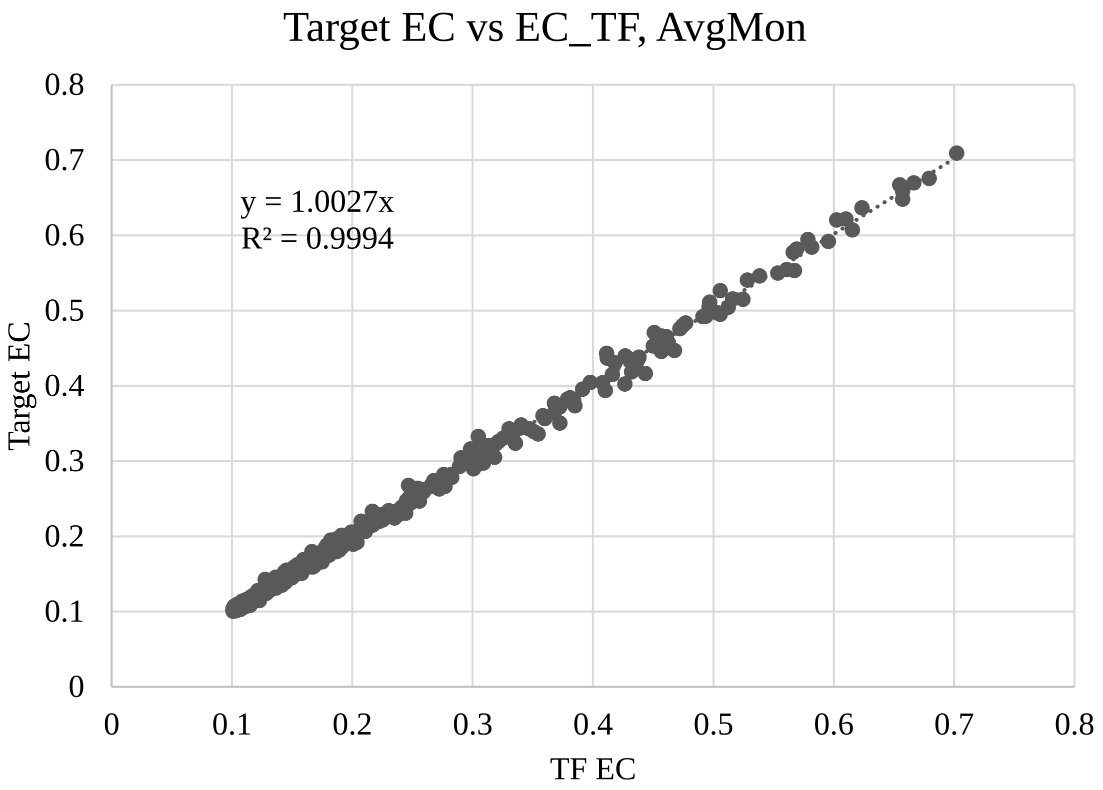
Performance Comparison Methodology

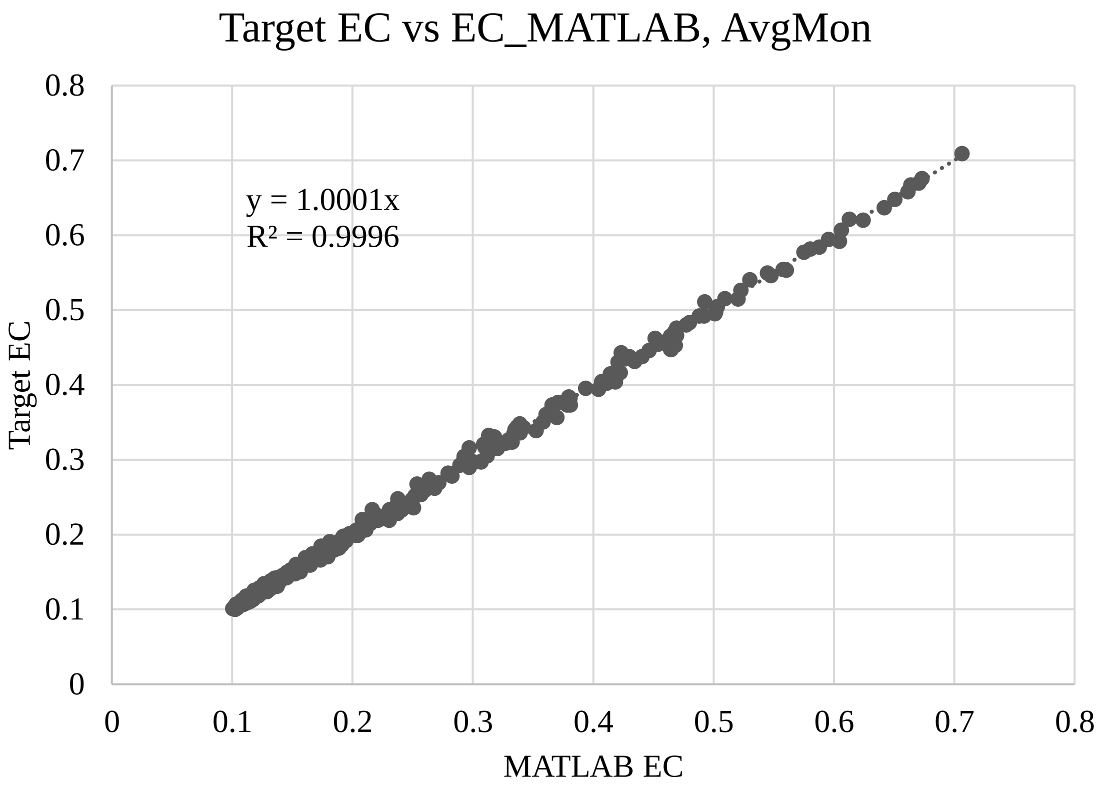
Stage 1: An Excel-based tool was developed to simulate the computational processes of models across both platforms.

Stage 1

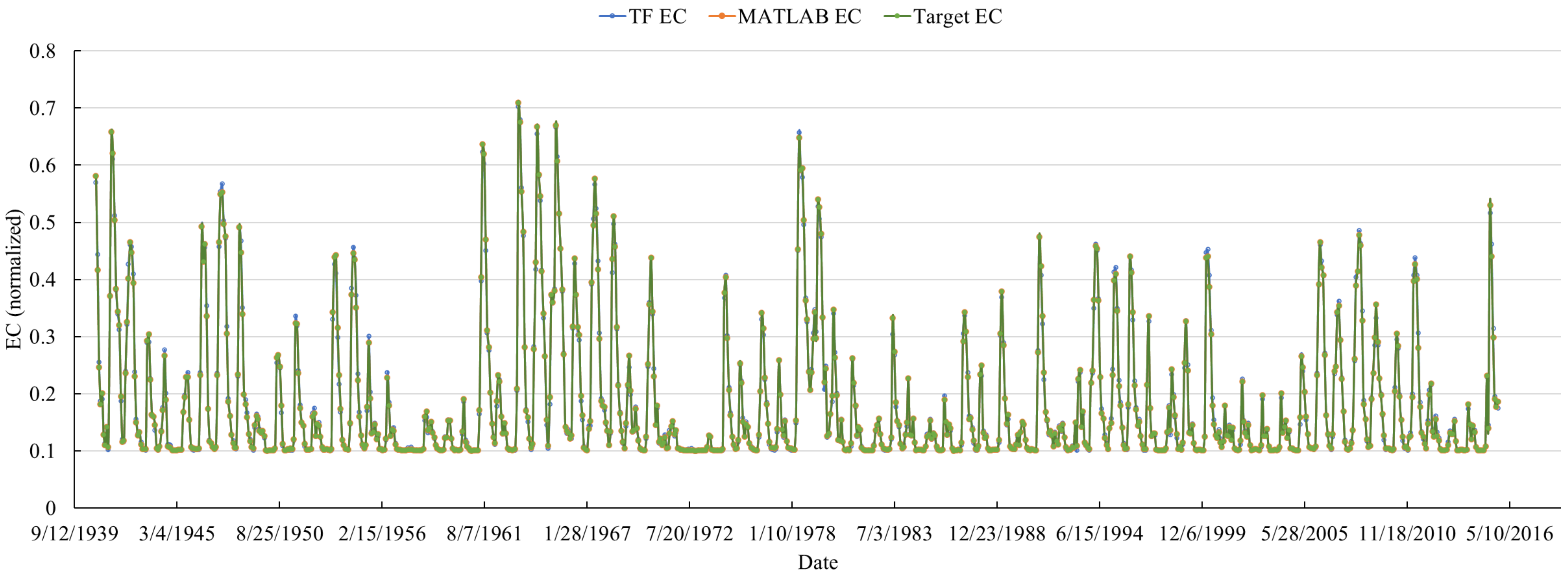
Stage 1 Results: Comparison out of ANN training and validation

Monthly EC comparisons from ANN training and validation (for the period of Oct1921 – Sep2021)

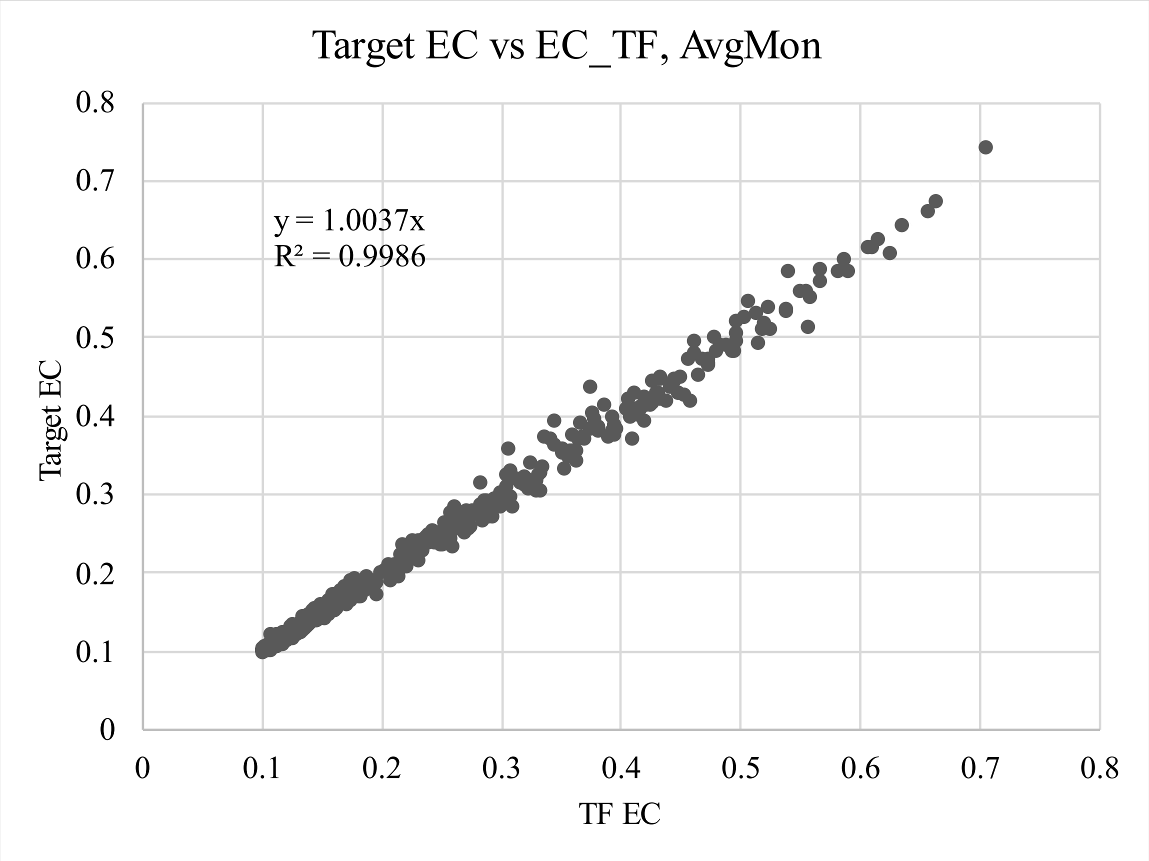


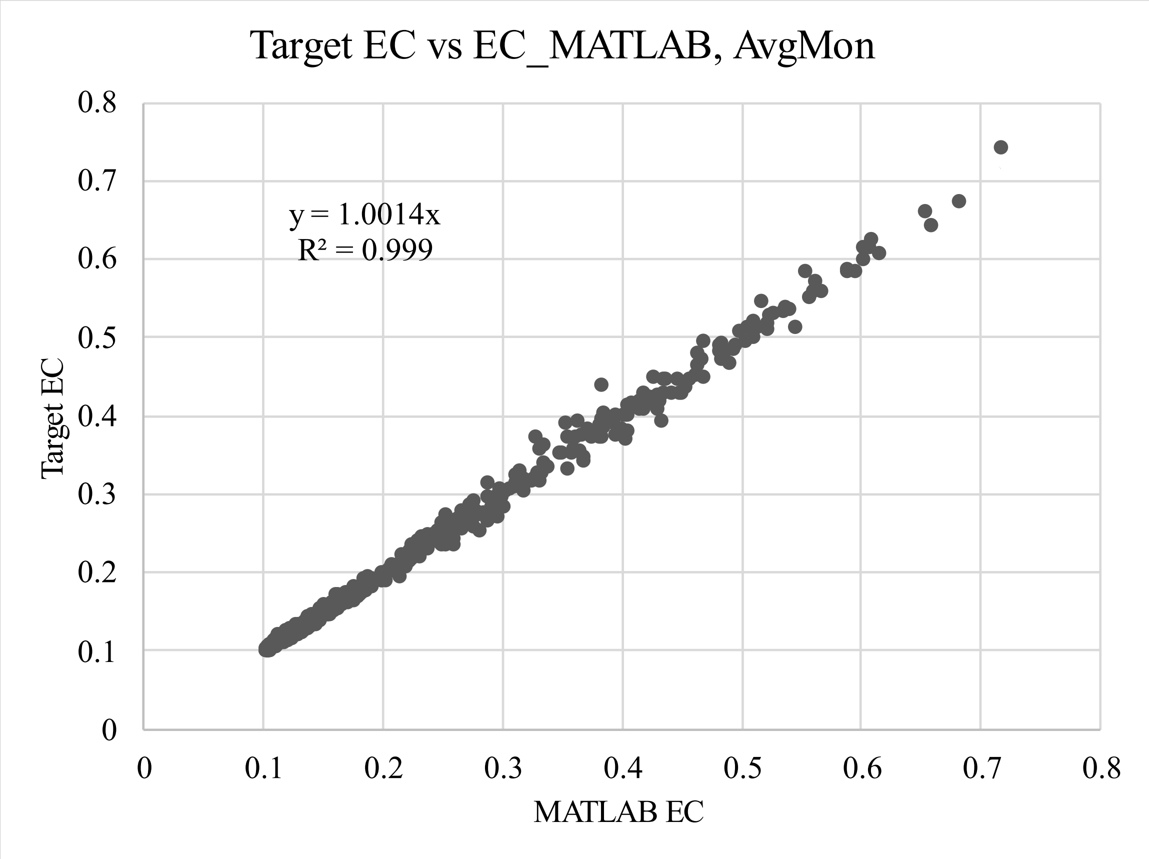


- Comparing the training results based on the TensorFlow, Emmaton (EMM), Training dataset, (80% of data selected from period 1940-2015, dataset were scaled and randomized.)

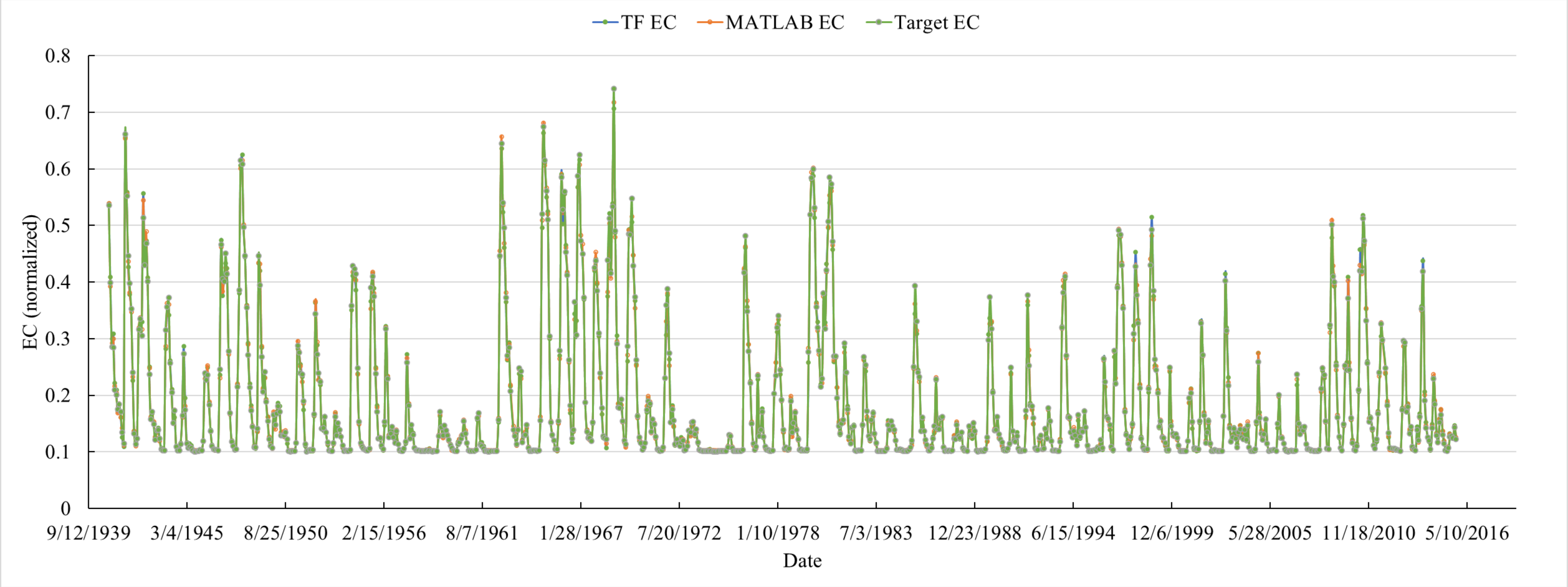


- Comparing the time series of training results based on the TensorFlow, Emmaton (EMM), Training dataset, (80% of data selected from period 1940-2015)



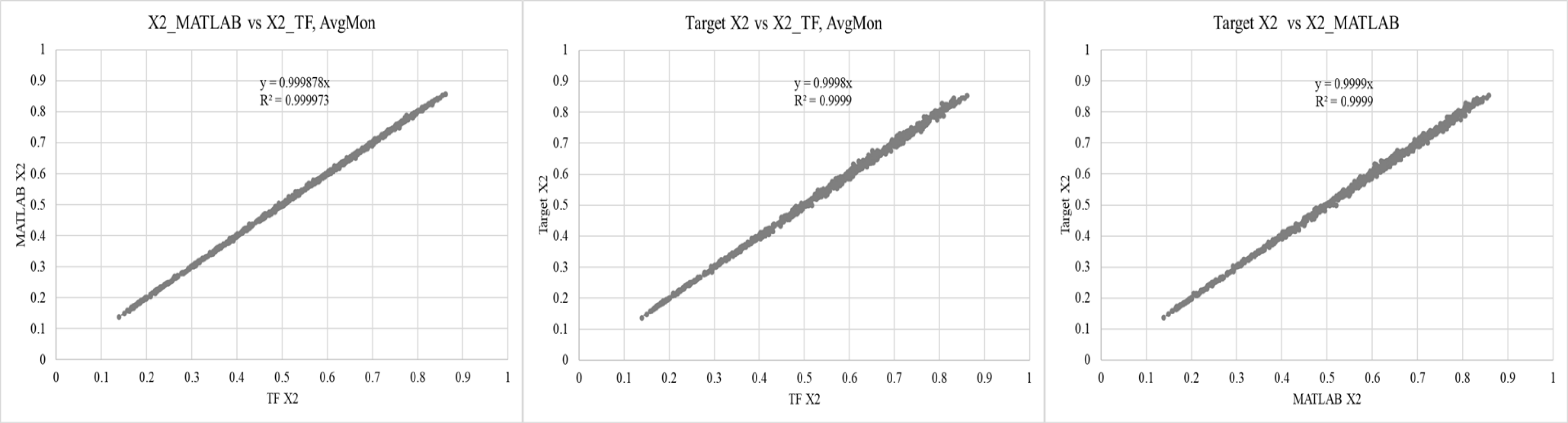


- Comparing the training results based on the TensorFlow, Emmaton (EMM), Validation dataset, (20% of data selected from period 1940-2015, dataset were scaled and randomized.)



- Comparing the training results based on the TensorFlow, Emmaton (EMM), Validation dataset, (20% of data selected from period 1940-2015)

Stage 1 Results: Comparison out of ANN training and validation (X2)

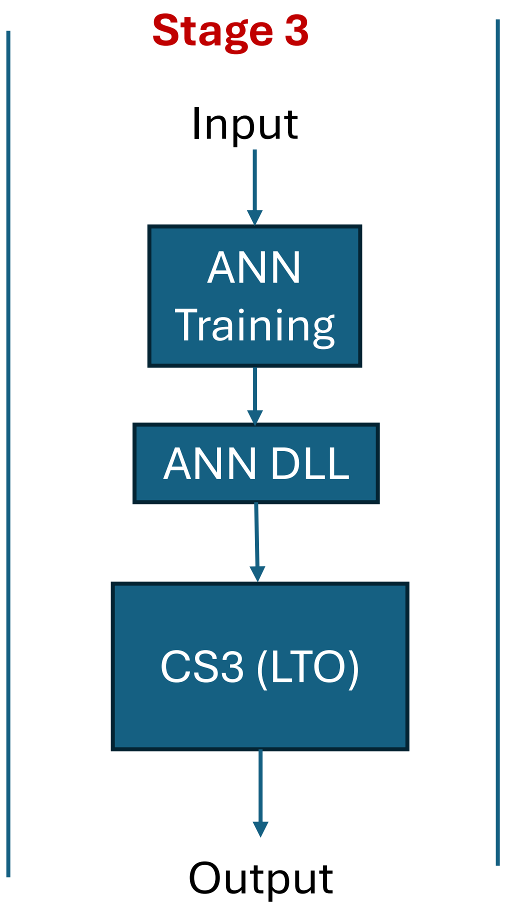


Comparing the training results based on the TensorFlow, X2, km, Training dataset, (80% of data selected from period 1940-2015).

- Overall, results compared through Excel-based tool are showing that TF-trained ANNs provide results that are visually and statistically comparable to MATLAB-trained ANNs.

Performance Comparison Methodology

Stage 3: A complete CS3 model run was conducted



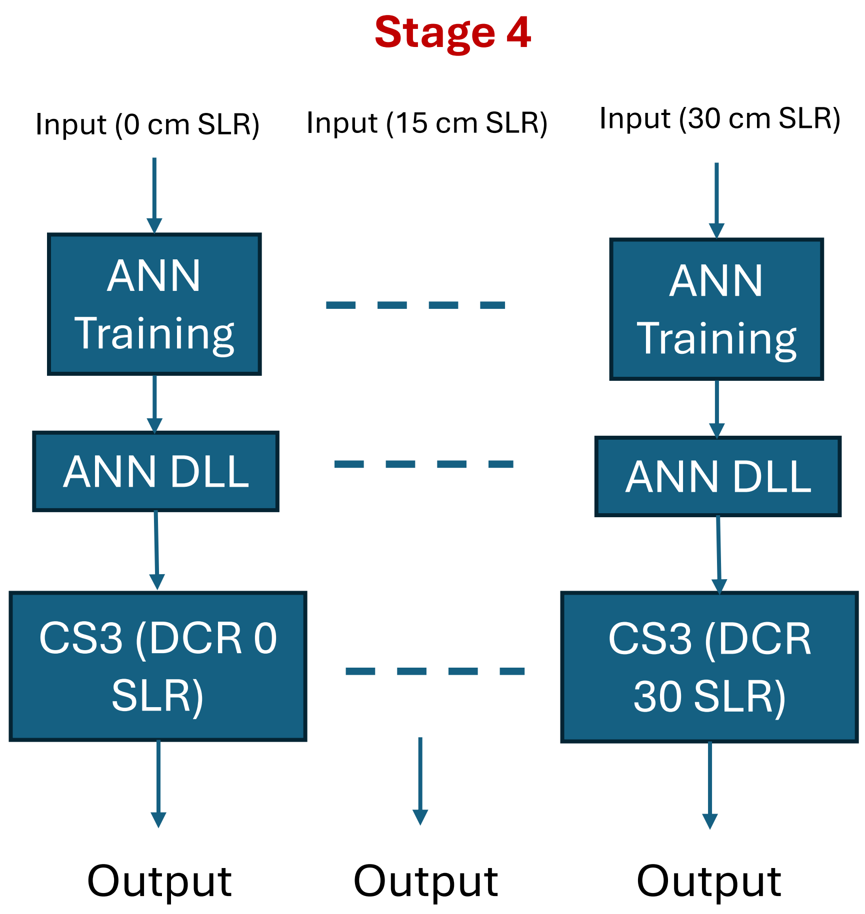
Stage 3 Results: Comparison of CS3 Key System Variables

	1922-2021			
	ANN_LTO _TF	ANN_LTO _ML	Diff	% Diff
River Flows				
Trinity R blw Lewiston	736	736	0	0
Trinity Export	502	502	0	0
Clear Cr blw Whiskeytown	147	147	0	0
Sacramento R @ Keswick	6133	6133	0	0
Sacramento R @ Wilkins Slough	6085	6085	0	0
Feather R blw Thermalito	2992	2992	0	0
Feather R at Sac R confluence	5246	5246	0	0
Yuba R @ Marysville	1497	1497	0	0
Sacramento R @ Verona	12753	12753	0	0
American R blw Nimbus	2482	2482	0	0
American R at Sac R confluence	2423	2423	0	0
GW Pumping Sac Total	3096	3097	-1	0
Delta Inflow	**21537**	**21535**	**1**	**0**
Sacramento R @ Hood	15407	15407	0	0
Yolo Bypass	2481	2480	1	0
Mokelumne R	873	873	0	0
Calaveras R	109	109	0	0
San Joaquin R d/s Vernalis	2667	2666	1	0
NDOI	**15256**	**15247**	**9**	**0**
Min Outflow	5219	5154	65	1
Additional Outflow	10037	10093	-55	-1
Delta Outflow - ANN	438	466	-27	-6
Delta Outflow - CVP	5184	5201	-17	0
Delta Outflow - SWP	4264	4276	-11	0
Delta Outflow - SJRR	121	121	0	0
Delta Outflow - VA	0	0	0	-19
Delta Outflow - WHLCV	13	13	0	3
Delta Outflow - WHLJP	0	0	0	
Delta Outflow - WTS	17	17	0	1
Delta Exports	**5031**	**5040**	**-9**	**0**
Banks	2546	2556	-10	0
Banks SWP	2449	2457	-9	0
Banks CVP	65	65	-1	-1
Banks WTS	33	33	0	-1
Jones	2484	2484	0	0
Jones CVP	2484	2484	0	0
Jones WTS	0	0	0	42
Delta Recapture	0	0	0	-37
SRSC LF Transfer	0	0	0	69
SWP Delivery: TA+CO	**2384**	**2391**	**-7**	**0**
Table A	2143	2151	-7	0
Article 21	96	98	-2	-2
Article 56	241	240	0	0

Performance Comparison Methodology

Stage 4: Complete Process Flow Performance Comparison based on three SLR CS3 studies

- Compare the performance of the whole ANN process flow in ML and TF (including their respective preprocessing modules).
- DCR studies under 0 cm, 15 cm, and 30 cm sea level rise (SLR) scenarios.
- Similar randomization process of ML was applied before training the model within TF platform.



Stage 4 Results: Complete Process Flow Performance Comparison

Comparing training performances of ML and TF ANNs in salinity (EC) estimation under 0 cm SLR scenario.

	RMSE				
	EMM	JP	RS	CO	X2
MATLAB	8.9%	6.7%	5.7%	5.2%	0.8%
TensorFlow	8.7%	7.2%	6.2%	5.4%	0.9%
	R^2				
	EMM	JP	RS	CO	X2
MATLAB	0.997	0.998	0.998	0.999	0.999
TensorFlow	0.997	0.997	0.997	0.999	0.990

Comparing training performances of ML and TF ANNs in salinity (EC) estimation under 15 cm SLR scenario.

	RMSE				
	EMM	JP	RS	CO	X2
MATLAB	7.4%	8.1%	5.9%	4.8%	0.8%
TensorFlow	8.4%	9.3%	7.2%	4.9%	0.8%
	R^2				
	EMM	JP	RS	CO	X2
MATLAB	0.998	0.997	0.999	0.998	0.999
TensorFlow	0.997	0.996	0.999	0.996	0.999

Comparing training performances of ML and TF ANNs in salinity (EC) estimation under 30 cm SLR scenario.

	RMSE				
	EMM	JP	RS	CO	X2
MATLAB	7.8%	9.2%	6.4%	4.5%	0.7%
TensorFlow	7.7%	9.0%	6.4%	5.2%	0.8%
	R^2				
	EMM	JP	RS	CO	X2
MATLAB	0.997	0.997	0.997	0.999	0.999
TensorFlow	0.997	0.997	0.997	0.999	0.999

Conclusion

•**Successful Transition to TensorFlow**: The shift from MATLAB (ML) to TensorFlow (TF) for estimating salinity (EC) and X2 has streamlined the process while preserving accuracy, as shown in the CS3 case study.

•**Comparison Studies**: The CS3 scenarios with corresponding DSM2 targets were analyzed, showing that TF-trained ANNs provide results that are visually and statistically comparable to ML-trained ANNs.

•**Data Handling**: Daily EC and X2 values calculated by the ANN were averaged over months to align with the monthly nature of CS3, with ANN DLLs coded to handle conversions between daily and monthly values.

•**Comprehensive Results**: Analysis across training, validation, and full datasets indicates that TF-trained ANNs perform similarly to ML-trained ANNs, with no significant deviations observed in simulation outcomes.

•**Performance Consistency**: The analysis confirms that TF-based ANN DLLs perform comparably to ML-based models, ensuring consistent simulation results across various scenarios, including sea level rise (SLR).

Recommendation for future works/options in terms of development and applications

•Other Salinity Control sStations can be considered to compare performance of two training platforms

Antioch, Mallard Island, Los Vaqueros, Middle River, Victoria Intake, CVP Intake, CLFB Intake, Balden Landing, Martinez.

•Multivariate model vs single model

The current univariate ANN models predict EC values separately for each station, but transitioning to a multivariate approach with multiple output nodes could enhance efficiency by allowing simultaneous predictions across all stations in a single model

•Transferring to JAVA based application

The current method exports Fortran-based ANN models to DLLs for use in CS3 simulations, but an alternative approach suggests using standalone models in the JAVA environment with internal preprocessing, requiring the conversion of the ANN LineGen module for full integration.

•Perturbing the input dataset

Perturbing the training datasets and using them to retrain the ANN models will enhance the models' robustness and generalization, making them more capable of handling a diverse range of input scenarios.

•Other Suggestions: …

Follow us on social media

CADWR

CA_DWR

calwater

cadepartmentofwaterresources