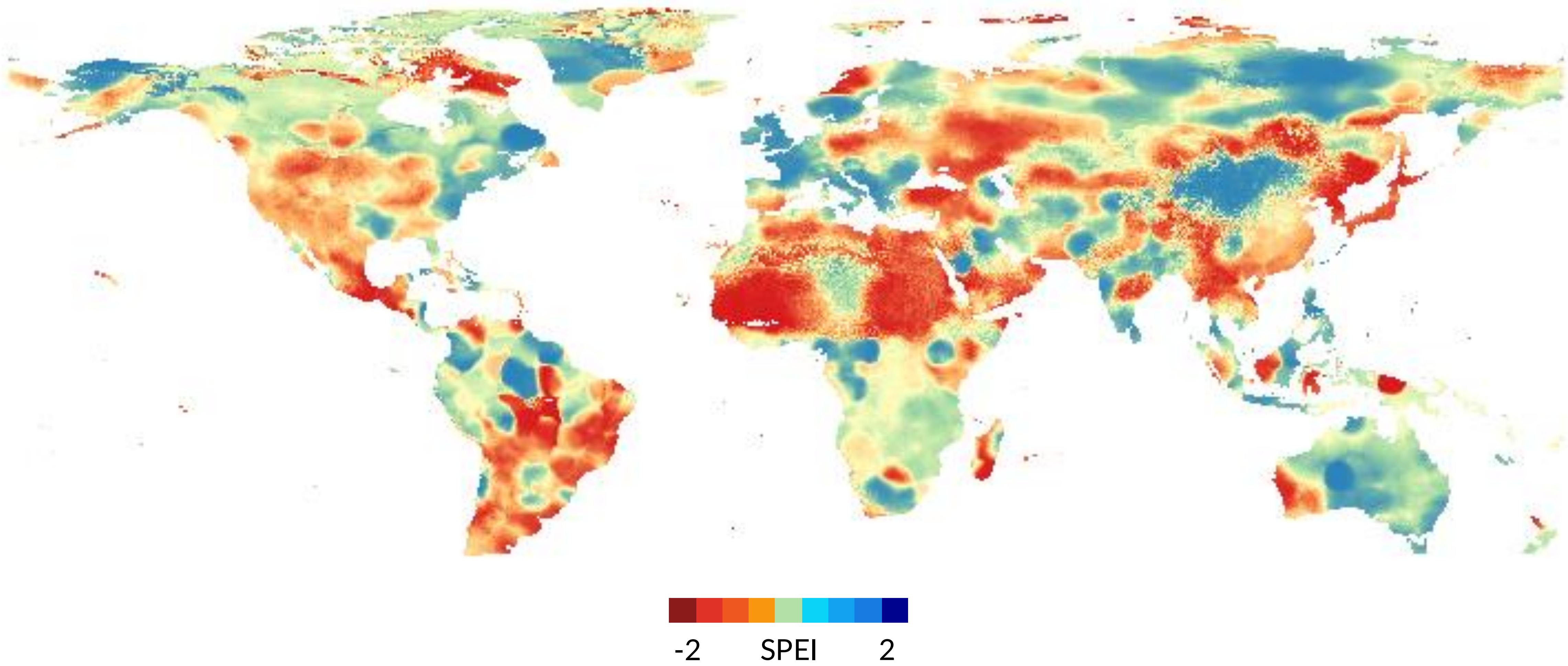
A MACHINE LEARNING FRAMEWORK TO DESIGN BASIN SPECIFIC DROUGHT INDEXES

Sami Miaari[1,2]
Advisor: Prof. Andrea Castelletti[1]
Co-Advisors: Dr. Elena Matta[1]
Co-Advisors: Dr. Matteo Sangiorgio[1]

(1) Department of Electronics, Information, and Bio-Engineering, Politecnico di Milano, Milan, Italy;
(2) Institute of Bio-Geosciences (IBG-3, Agrosphere), Research Centre Jülich, Jülich, Germany.

DROUGHTS



Source: Laboratory of Climate Services and Climatology (2022)

DROUGHTS

DEFINITIONS AND INDEXES

Meteorological

- SPI
- SPEI
- Precipitation
- Temperature

Agricultural

- NDVI
- VCI
- SSI
- SPEI
- Soil moisture

Hydrological

- SRI
- River discharge

Operational

- Reservoir storage level
- GDP
- Hydro-power production

OBJECTIVE

Drought monitoring and forecasting

Water stress

Open source data (e.g., remote sensing, reanalysis)

Reproduce a drought index adapting to different study areas

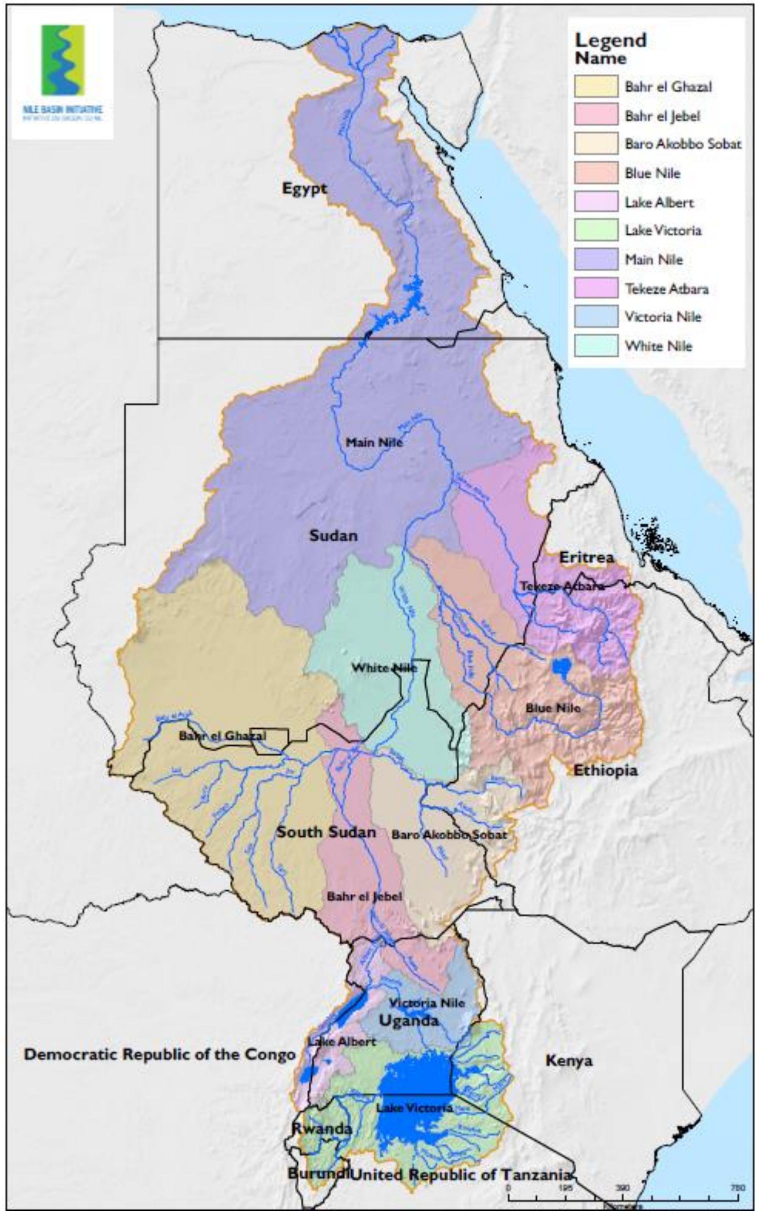
An easy-to-use drought monitoring index able to identify drought conditions

FRIDA

FRAMEWORK FOR INDEX BASED DROUGHT ANALYSIS

Methodology *(Zaniolo et al., 2018)*		Technical application	Performance evaluation
Basin characteristics	• Definition of target variable • Collection of input variable predictors	• NDVI as the target variable • Hydro-meteorological variables • Drought indicators	
Feature extraction	• Input variable selection • Definition of Pareto-efficient subsets	• W-QEISS • Borg MOEA • Extreme Machine Learning	Symmetric Uncertainty • Accuracy • Cardinality • Relevance • Redundancy
Drought index modeling	• Choice of preferred subset • Calibration of the selected model class	• Linear regression model • Feedforward Artificial Neural Networks	• Coefficient of determination (R^2) • Root Mean Squared Error (RMSE)

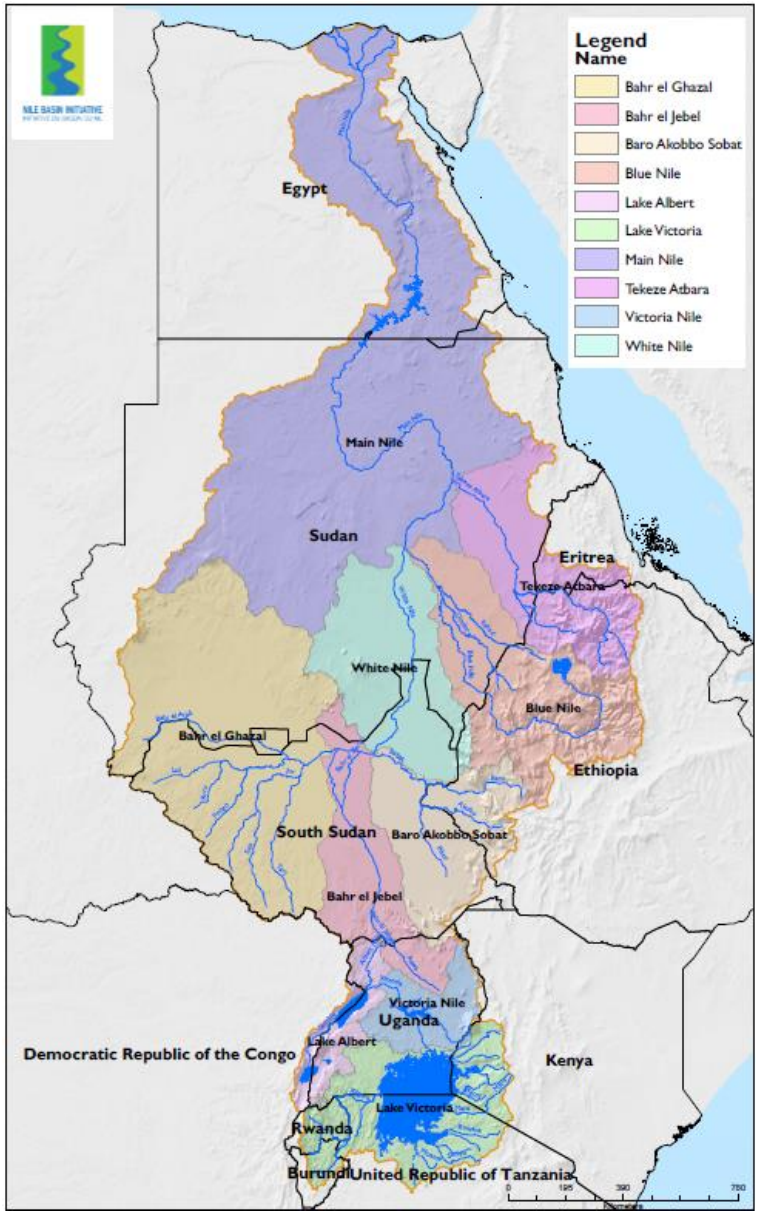
CASE STUDY: THE NILE RIVER BASIN



Nile River major sub-basins (NBI, 2016)

- The basin climate varies from south to north
- Complex hydrology and different topographic areas
- Climate change and population growth
- Most of the Nile River water is allocated for agriculture
- Lack of efficient water management coordination among the basin countries
- Transboundary waters conflicts

CASE STUDY: THE NILE RIVER BASIN



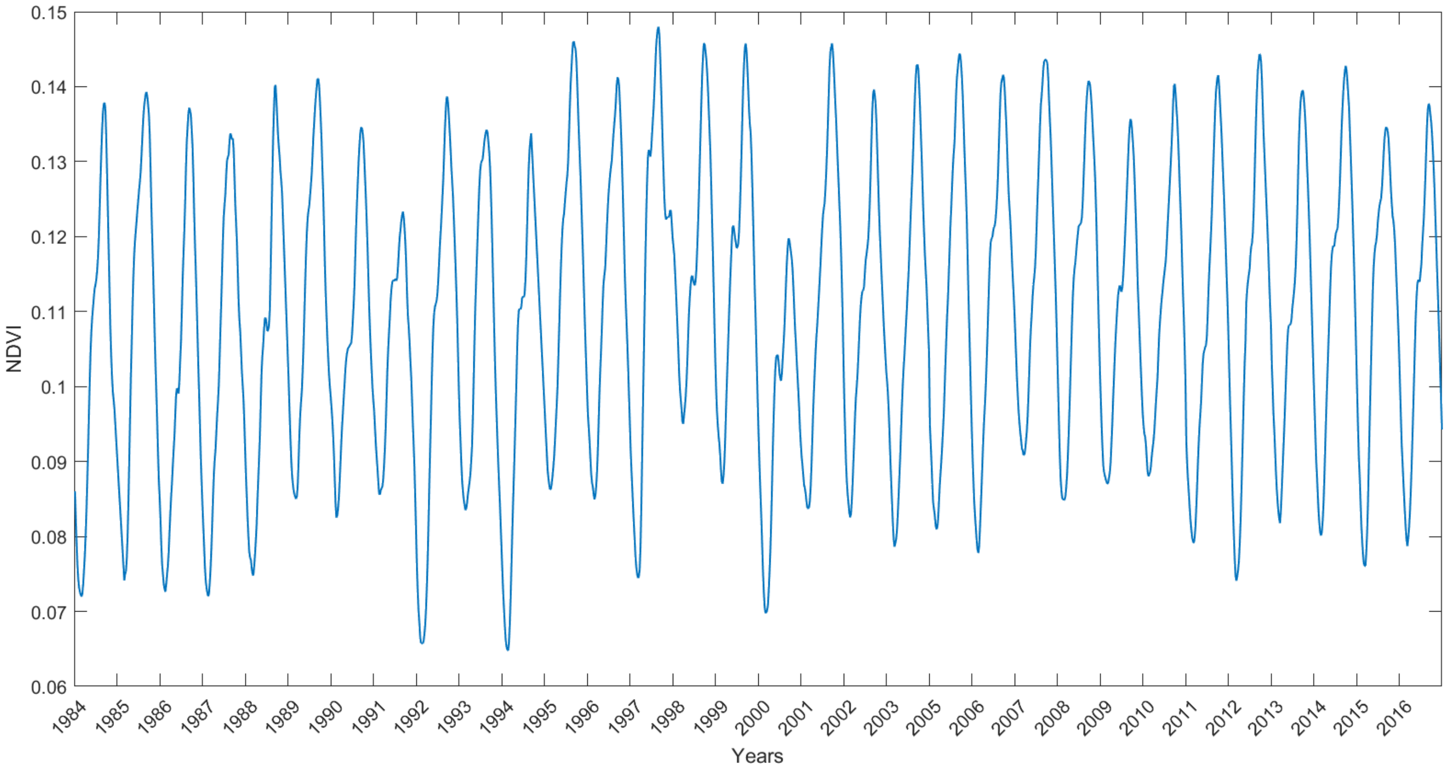
Nile River major sub-basins (NBI, 2016)

Input variables used for the W-QEISS

Input features	Source	Spatial resolution	Temporal resolution	Time aggregation (weeks)
Year	-	-	-	Not aggregated
Week	-	-	-	Not aggregated
Precipitation	CHIRPS	0.05°x0.05°	Daily	1,2,4
Tmin				1,2
Tmax	CHIRTS	0.05°x0.05°	Daily	1,2
Tmean				1,2,4
Evapotranspiration	ERA5	0.1°x0.1°	Hourly	1,3,6,16
River discharge	GloFAS	0.1°x0.1°	Daily	1,2,4,16
Soil moisture	MERRA-2	0.5°x0.625°	Hourly	1,16,52
SPI	Precipitation based (CHIRPS)			1,3,6,16,52
NDVI	NOAA STAR	4kmx4km	Weekly	Target variable

TARGET VARIABLE: NDVI

- Able to
  - detect agricultural and meteorological droughts
  - describe vegetation health conditions and stress level
- Highly dependent on land cover

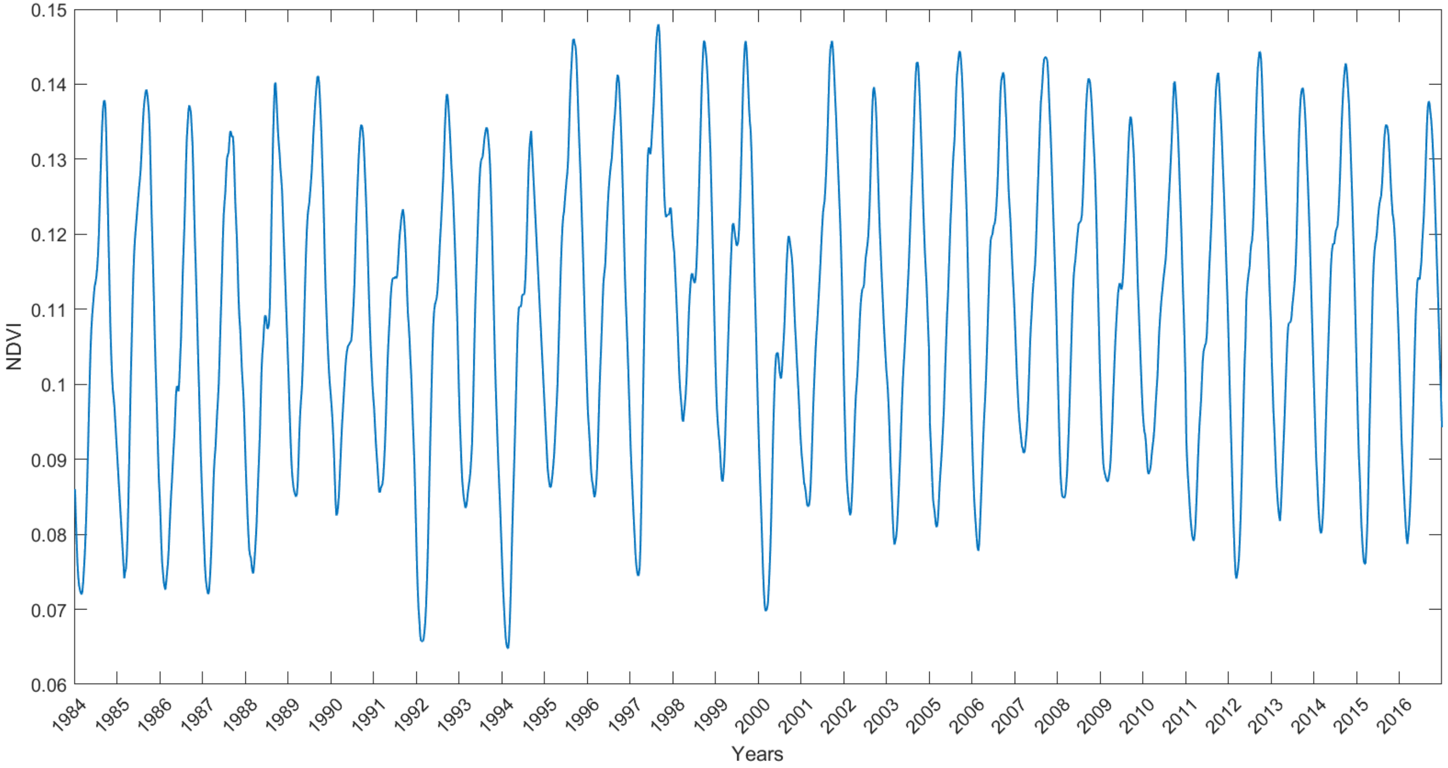


Mean NDVI values for the Nile River Basin (STAR)

The Nile River major sub-basins

TARGET VARIABLE: NDVI

- Able to
  - detect agricultural and meteorological droughts
  - describe vegetation health conditions and stress level
- Highly dependent on land cover
- Averaged excluding water bodies and desert

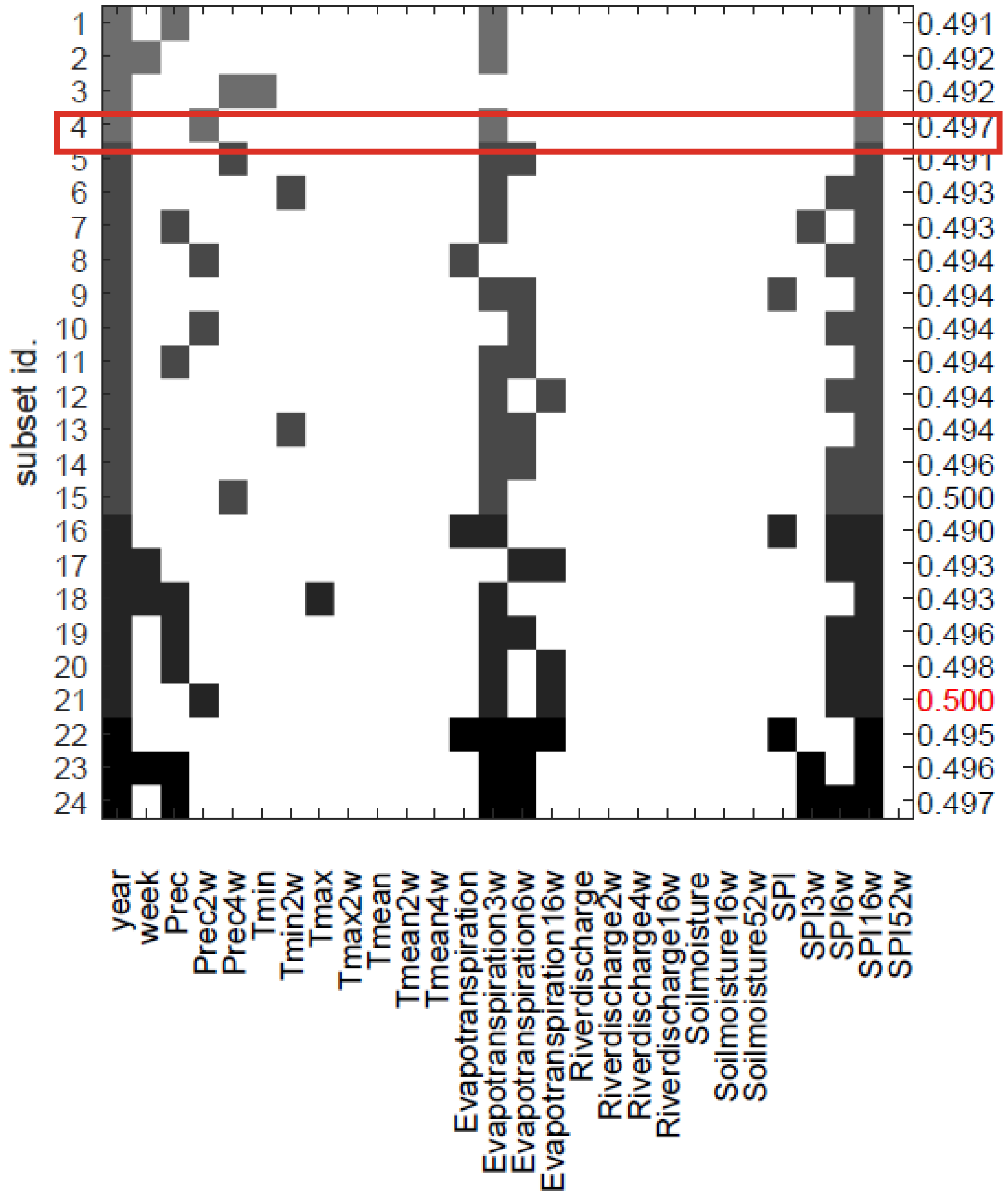


Mean NDVI values for the Nile River Basin (STAR)

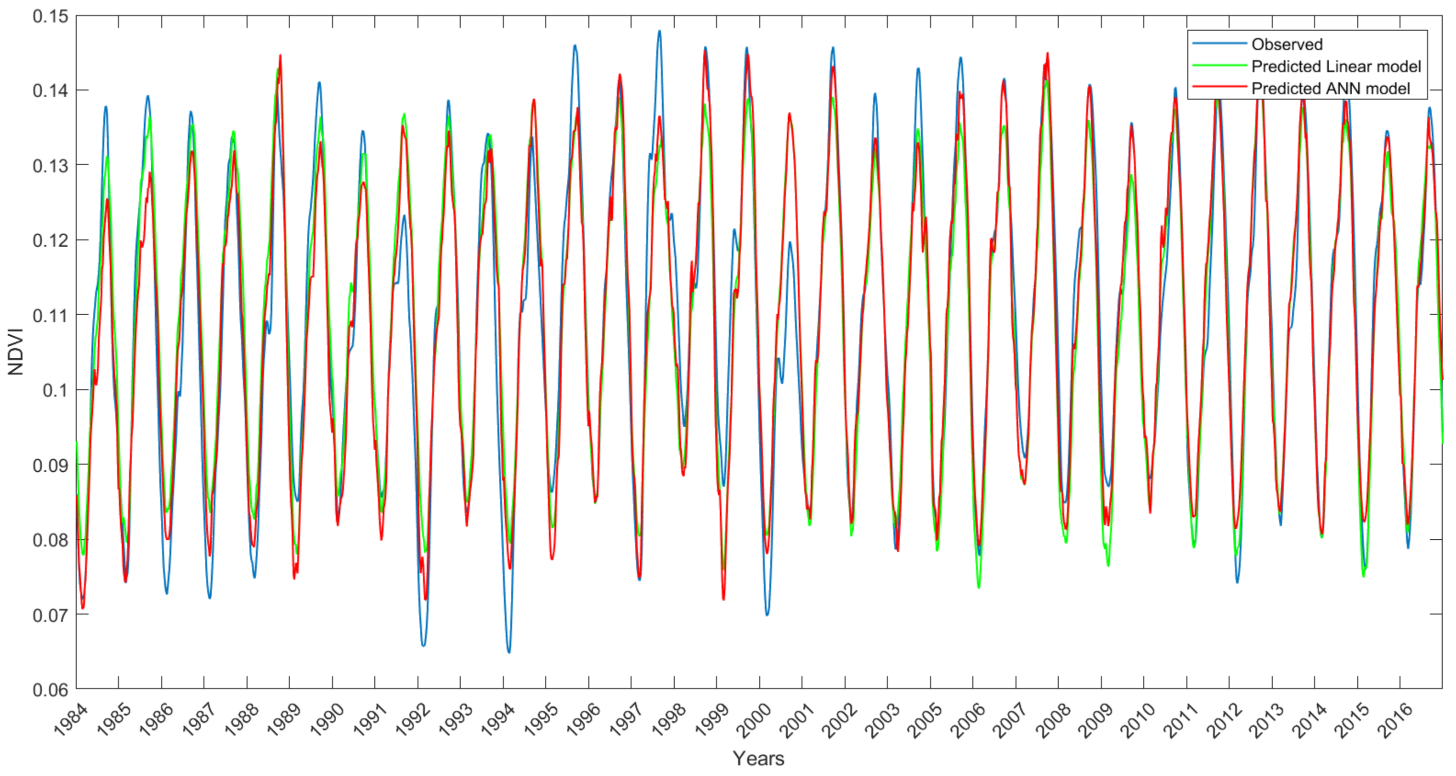
The Nile River major sub-basins excluding water bodies and desert

Nile River Basin scale

RESULTS: THE NILE RIVER BASIN



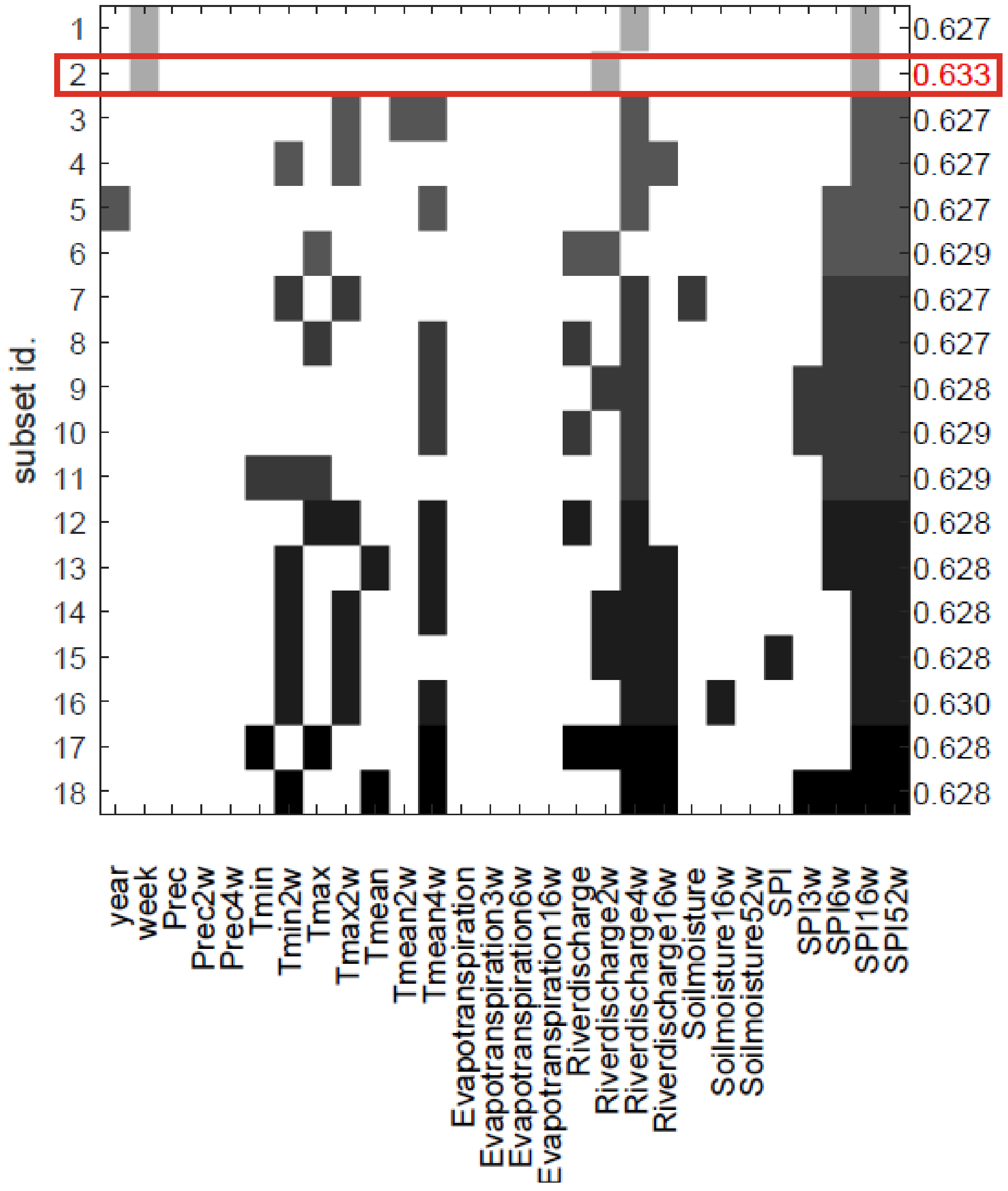
Feature selection results for the Nile River Basin



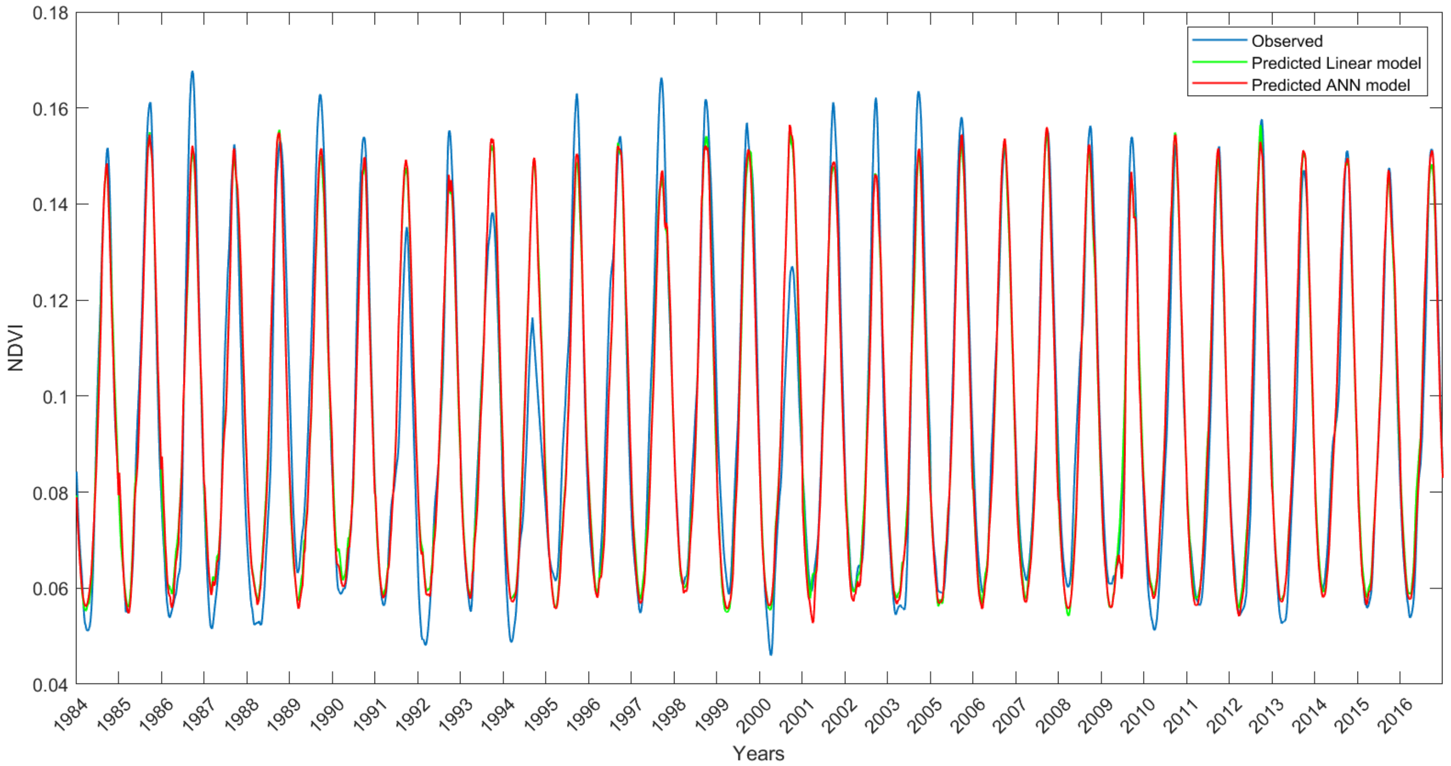
Observed and predicted NDVI values over the Nile River Basin with R^2_{Linear}=0.887 and R^2_{ANN}=0.9057

Sub-basin scale

RESULTS: BLUE NILE

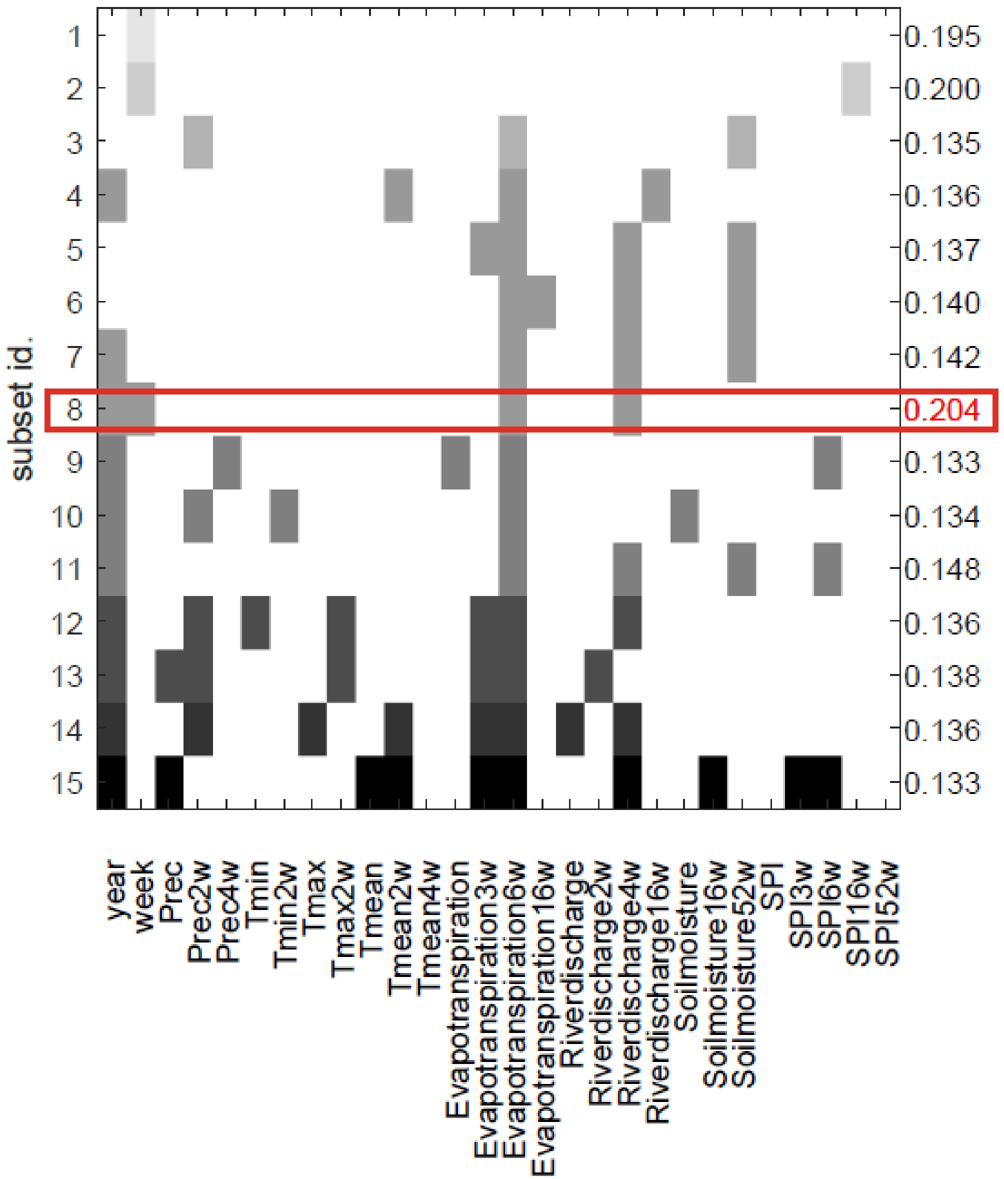


Feature selection results for the Blue Nile sub-basin

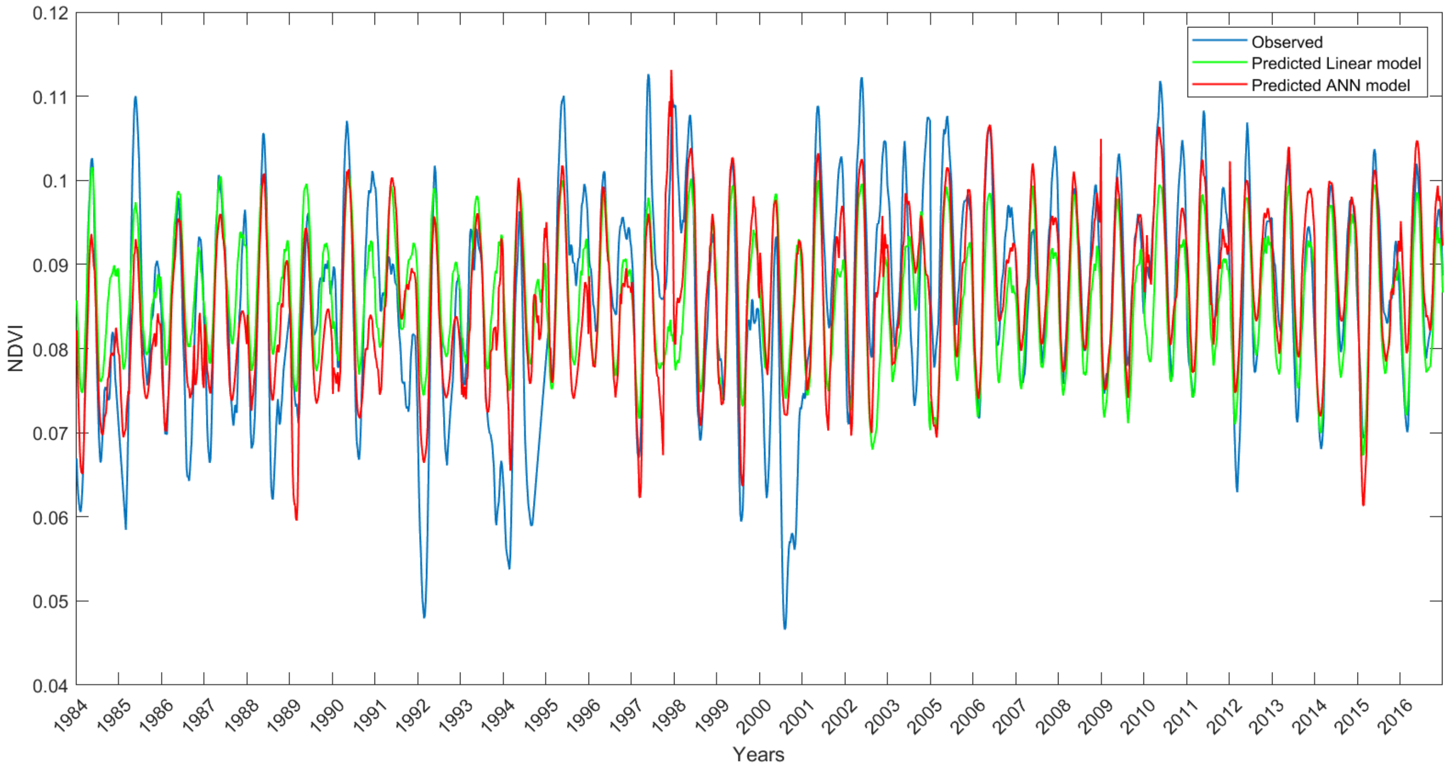


Observed and predicted NDVI values over Blue Nile with R^2_{Linear}=0.928 and R^2_{ANN}=0.9288

RESULTS: LAKE ALBERT



Feature selection results for the Lake Albert sub-basin



Observed and predicted NDVI values over Lake Albert with R^2_{Linear}=0.4151 and R^2_{ANN}=0.5498

RESULTS: ALL SUB-BASINS

Selected subsets for the major sub-basins

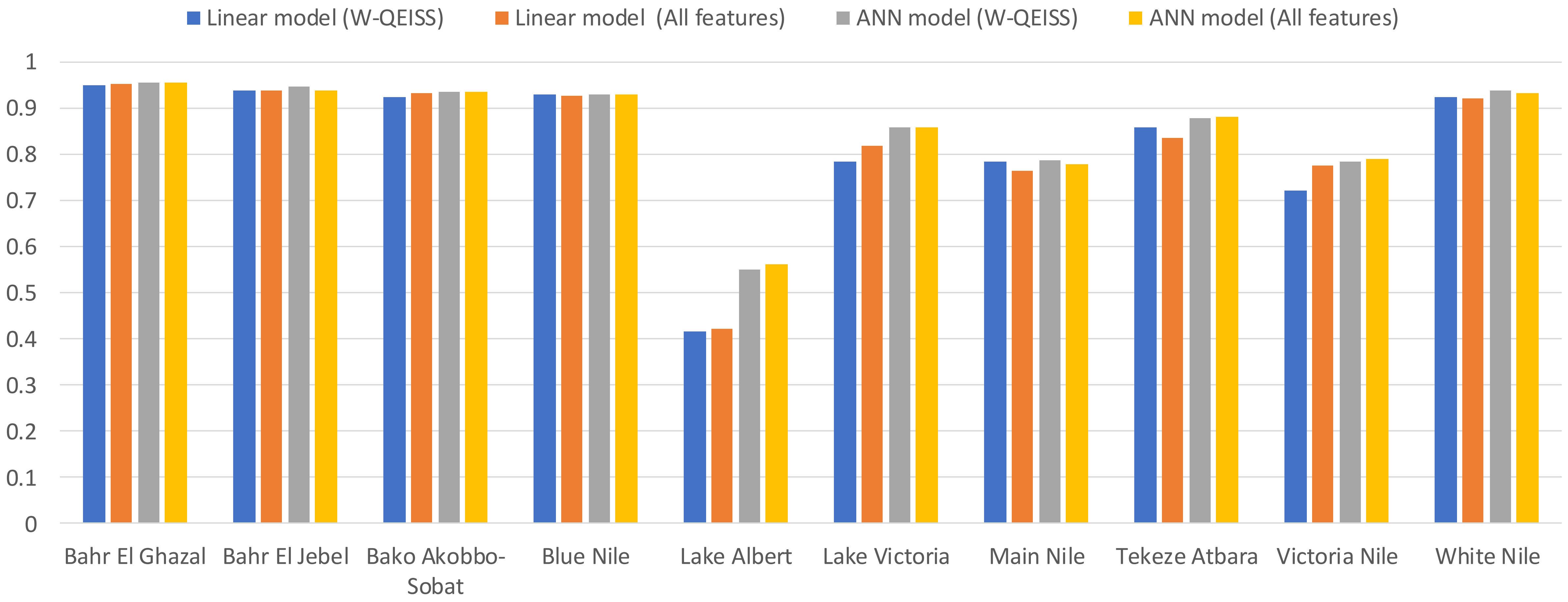
Sub-basin	Year	Week	Precipitation	Tmin	Tmax	Tmean	Evapo-transpiration	River discharge	Soil moisture	SPI	SU
Bahr El Ghazal		X	4w							16w	0.675
Bahr El Jebel		X								16w	0.627
Bako Akobbo-Sobat	X	X	2w				1w		1w	6w,16w	0.606
Blue Nile		X						2w		16w	0.633
Lake Albert	X	X					6w	4w			0.204
Lake Victoria	X						6w	4w	52w		0.342
Main Nile			1w	1w,2w		4w		1w,16w		3w,6w,16w	0.371
Tekeze Atbara		X	2w							3w,16w	0.495
Victoria Nile	X	X				4w	6w				0.248
White Nile	X								1w	6w,16w	0.603

RESULTS: ALL SUB-BASINS

Selected subsets for the major sub-basins

Sub-basin	Year	Week	Precipitation	Tmin	Tmax	Tmean	Evapo-transpiration	River discharge	Soil moisture	SPI	SU
Bahr El Ghazal		X	4w							16w	0.675
Bahr El Jebel		X								16w	0.627
Bako Akobbo-Sobat	X	X	2w				1w		1w	6w,16w	0.606
Blue Nile		X						2w		16w	0.633
Lake Albert	X	X					6w	4w			0.204
Lake Victoria	X						6w	4w	52w		0.342
Main Nile			1w	1w,2w		4w		1w,16w		3w,6w,16w	0.371
Tekeze Atbara		X	2w							3w,16w	0.495
Victoria Nile	X	X				4w	6w				0.248
White Nile	X								1w	6w,16w	0.603
					0%					70%	

RESULTS: FEATURE SELECTION VS ALL FEATURES



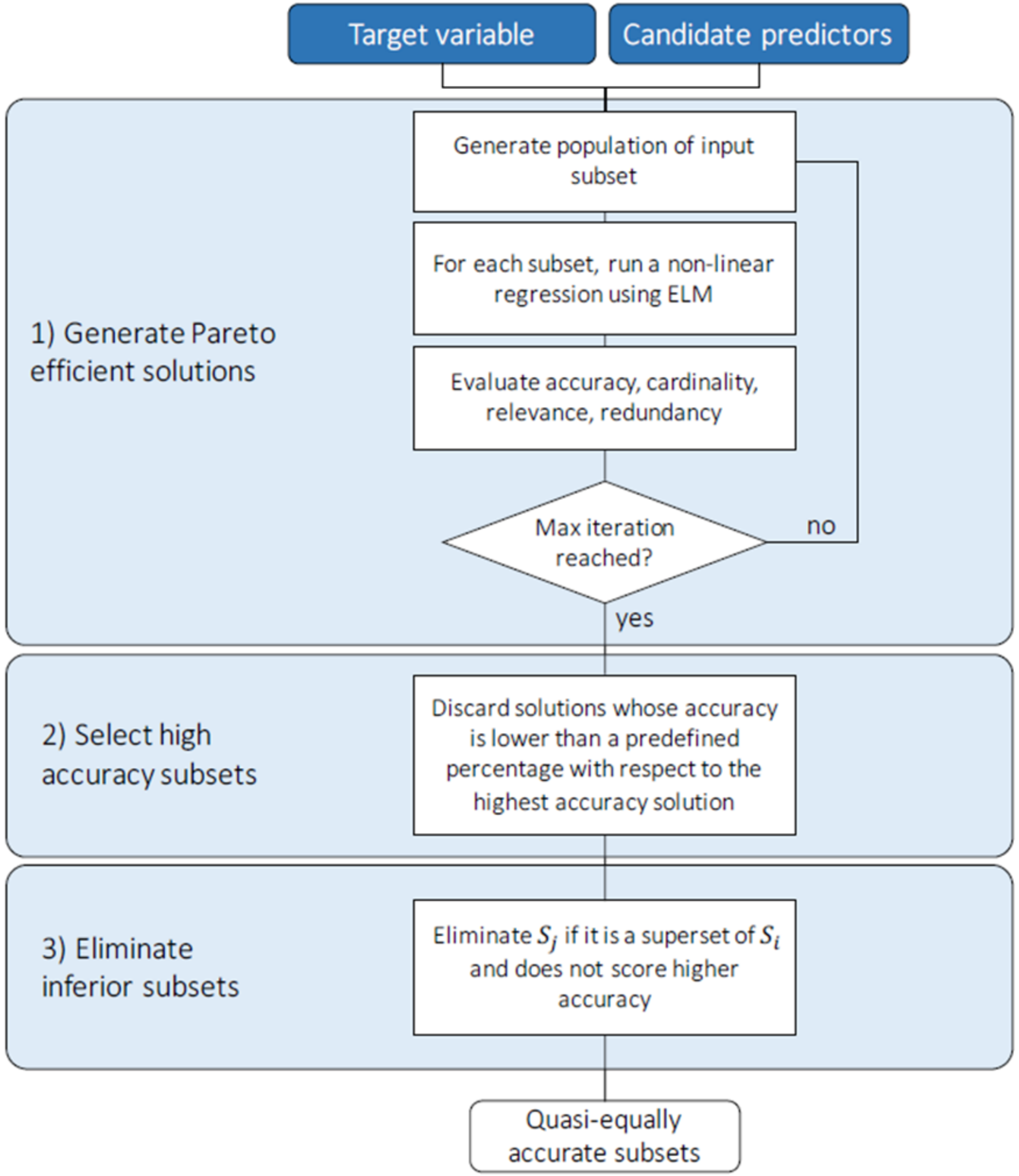
Regression models performance using the selected input variables compared with using all predictors

CONCLUSIONS & OUTLOOK

- The designed drought index can well reproduce the target variable for most of the Nile subbasins
- Both ANN and linear model perform similarly with a preference for the **non-linear model**
- The **efficiency of W-QEISS** is confirmed by comparing the selected subsets with all feature
- The selected subsets identify the **correlated variables** for the considered study area
- A **consistent basin subdivision** based on specific criteria (e.g., climatic characteristics, land cover, and topography) is recommended
- Input variables' spatial resolution as an additional criteria for basin subdivision can ensure noise filtering
- FRIDA has demonstrated its transferability to complex case studies
- Introducing additional prediction variables such as groundwater table and air humidity may improve FRIDA's performance

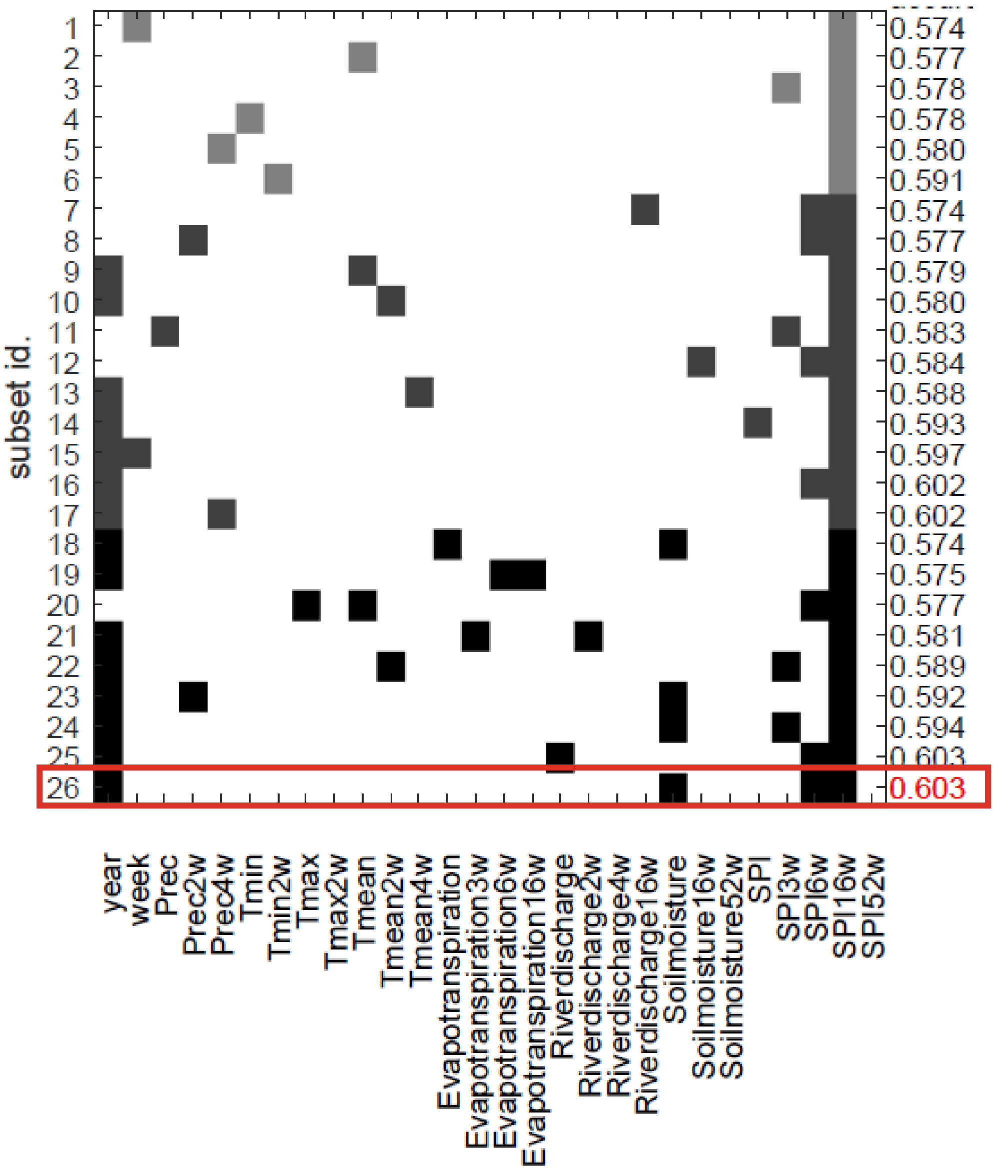
Thanks for your attention

W-QEISS METHODOLOGY

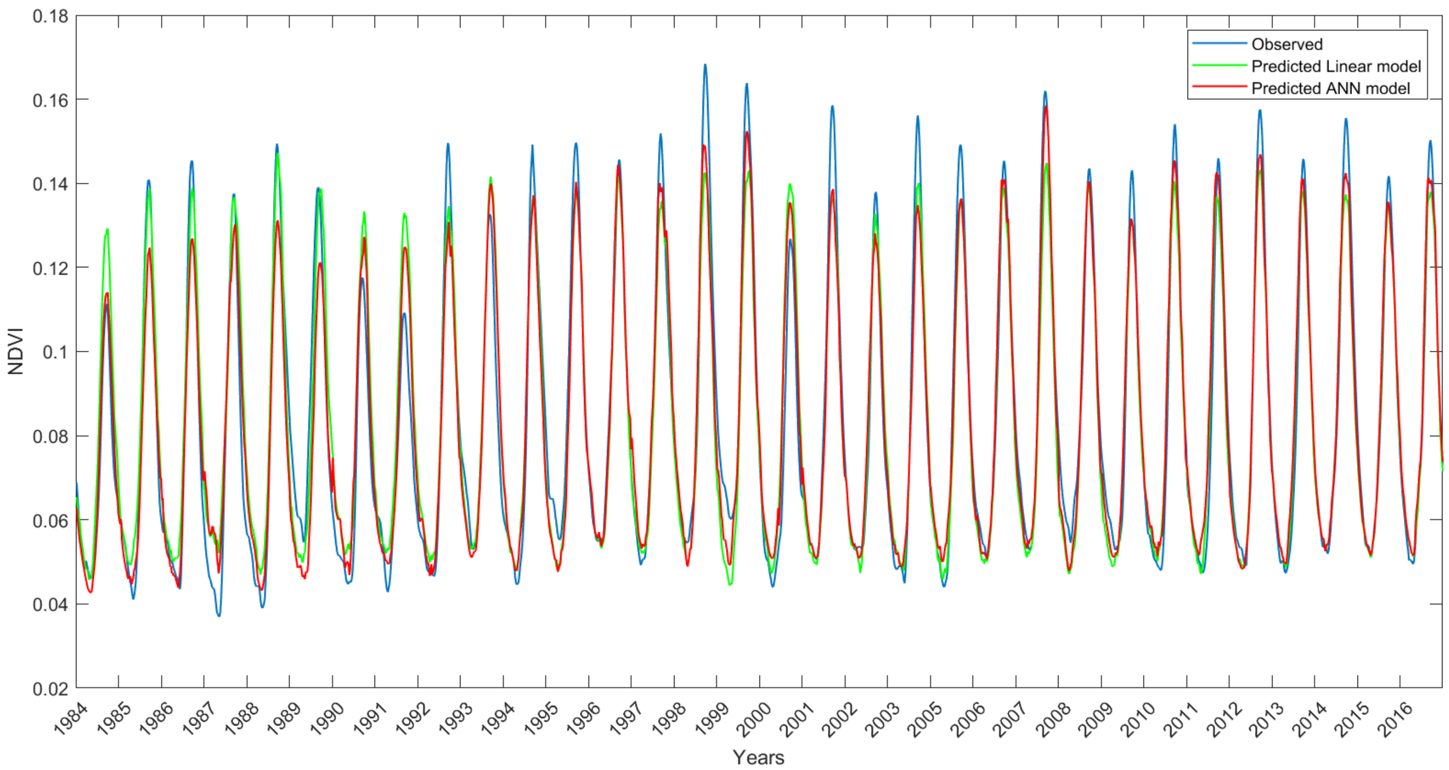


W-QEISS methodology (Zaniolo et al., 2018)

RESULTS: WHITE NILE

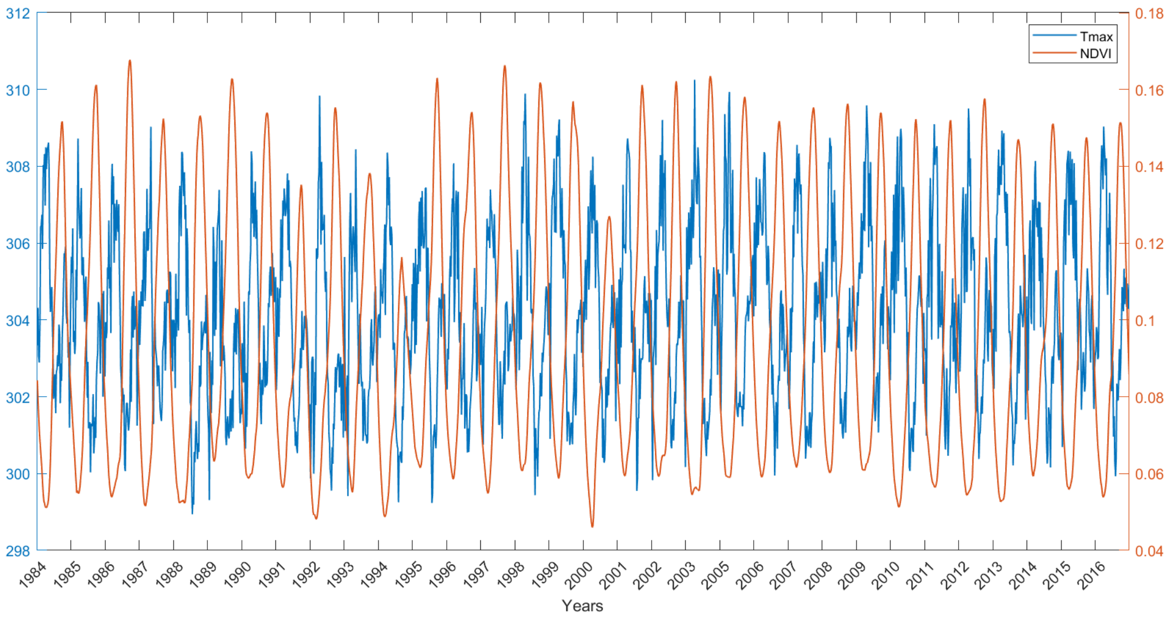


Feature selection results for the White Nile sub-basin

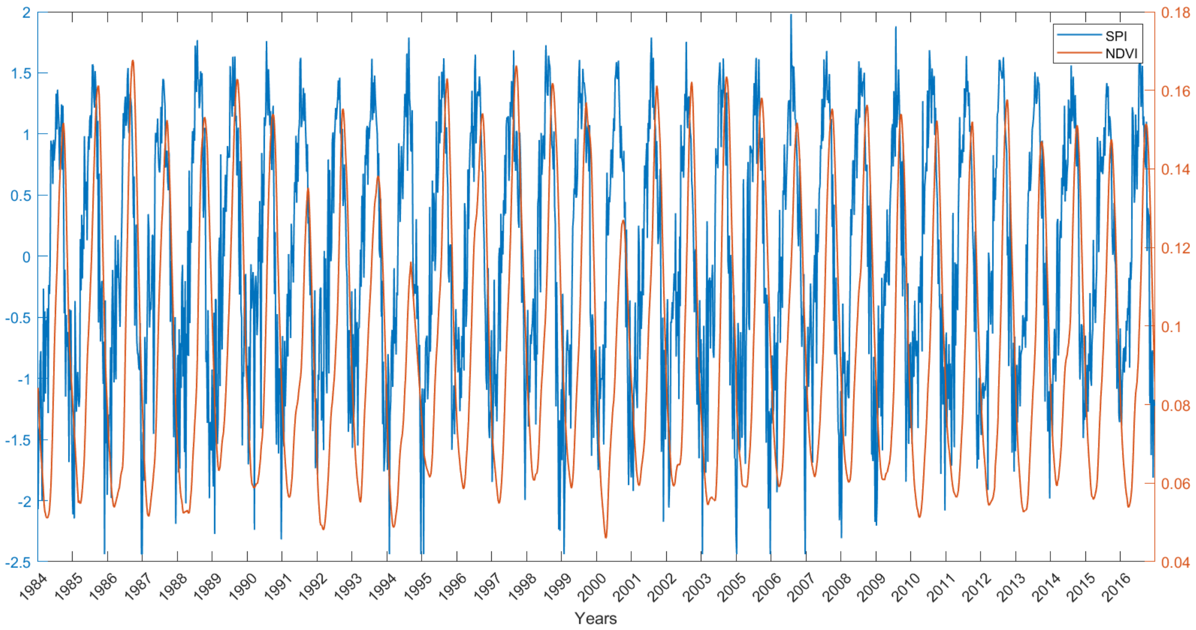


Observed and predicted NDVI values over White Nile with R^2_{Linear}=0.923 and R^2_{ANN}=0.937

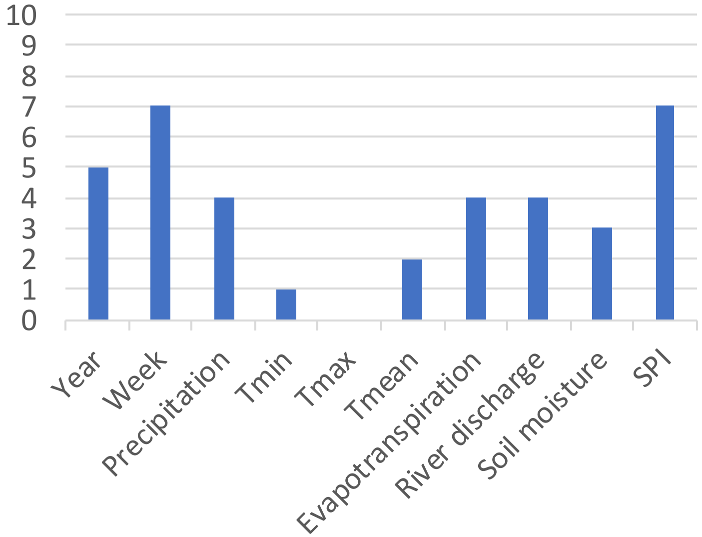
RESULTS: INPUT VARIABLES SELECTION



Tmax vs NDVI (Blue Nile)

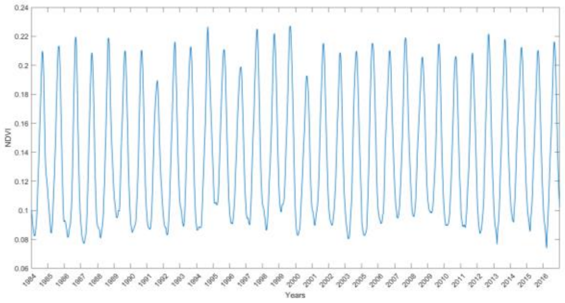


SPI vs NDVI (Blue Nile)

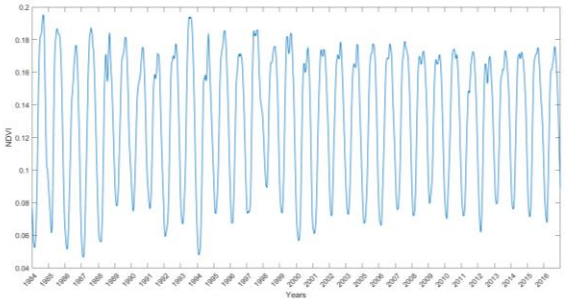


Input variables selection frequency

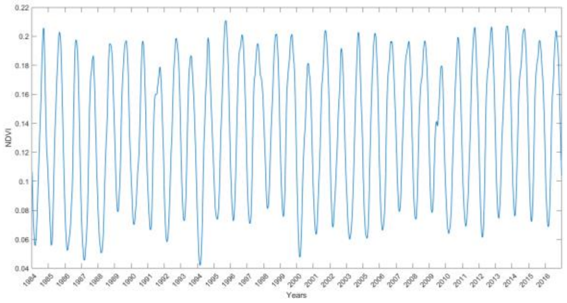
NDVI FOR ALL SUB-BASINS



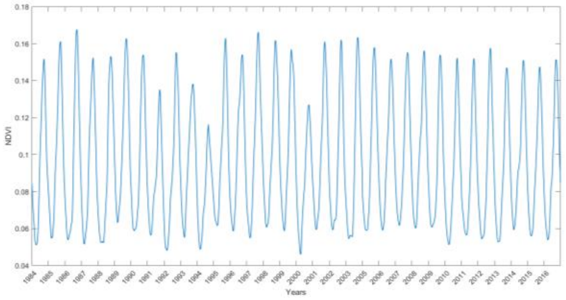
Bahr El Ghazal



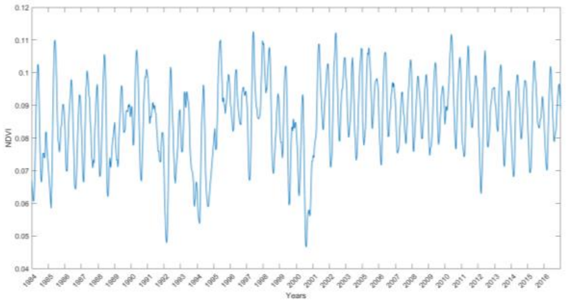
Bahr El Jebel



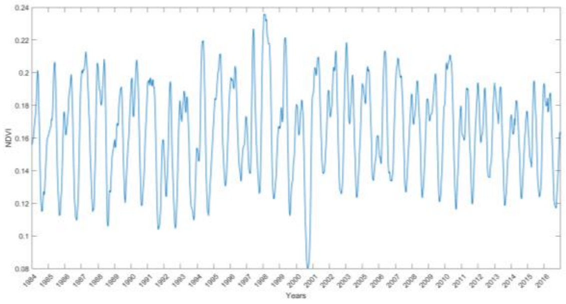
Bako Akobbo-Sobat



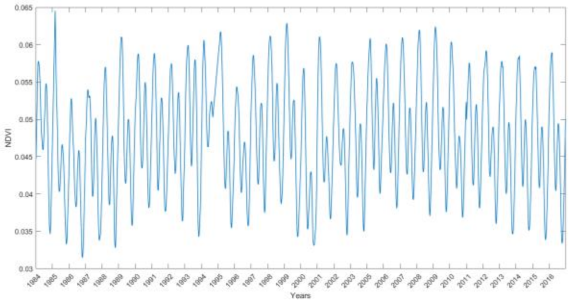
Blue Nile



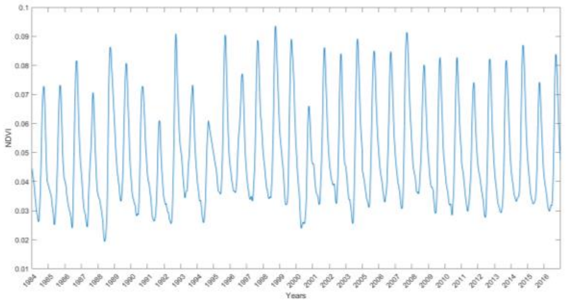
Lake Albert



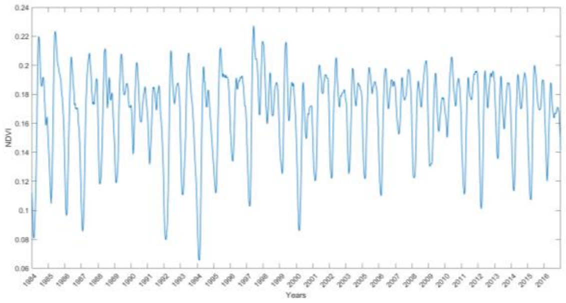
Lake Victoria



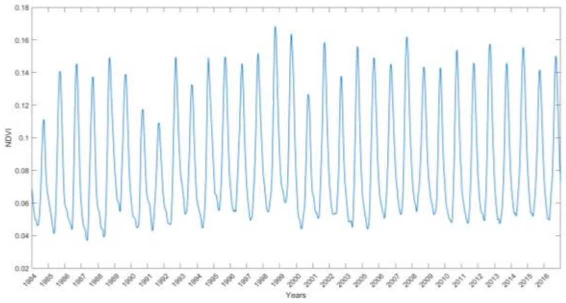
Main Nile



Tekeze Atbara

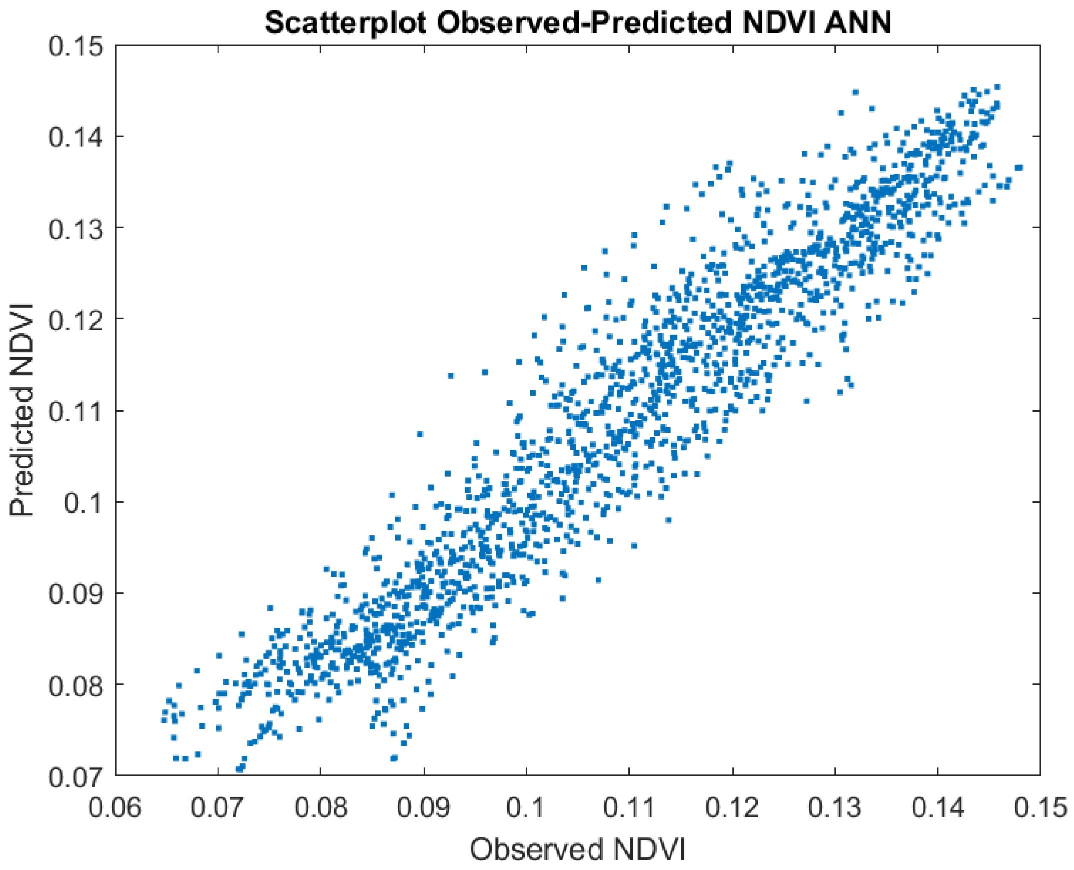


Victoria Nile

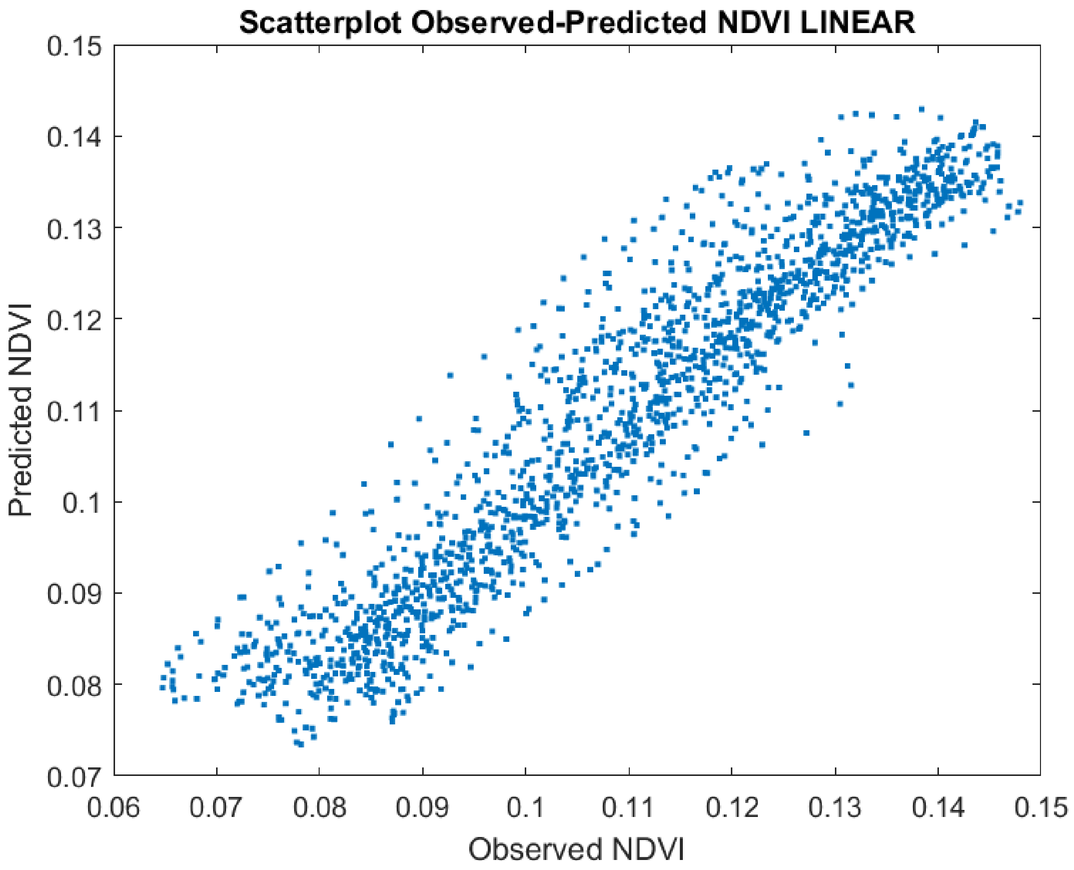


White Nile

THE NILE RIVER BASIN SCATTERPLOTS

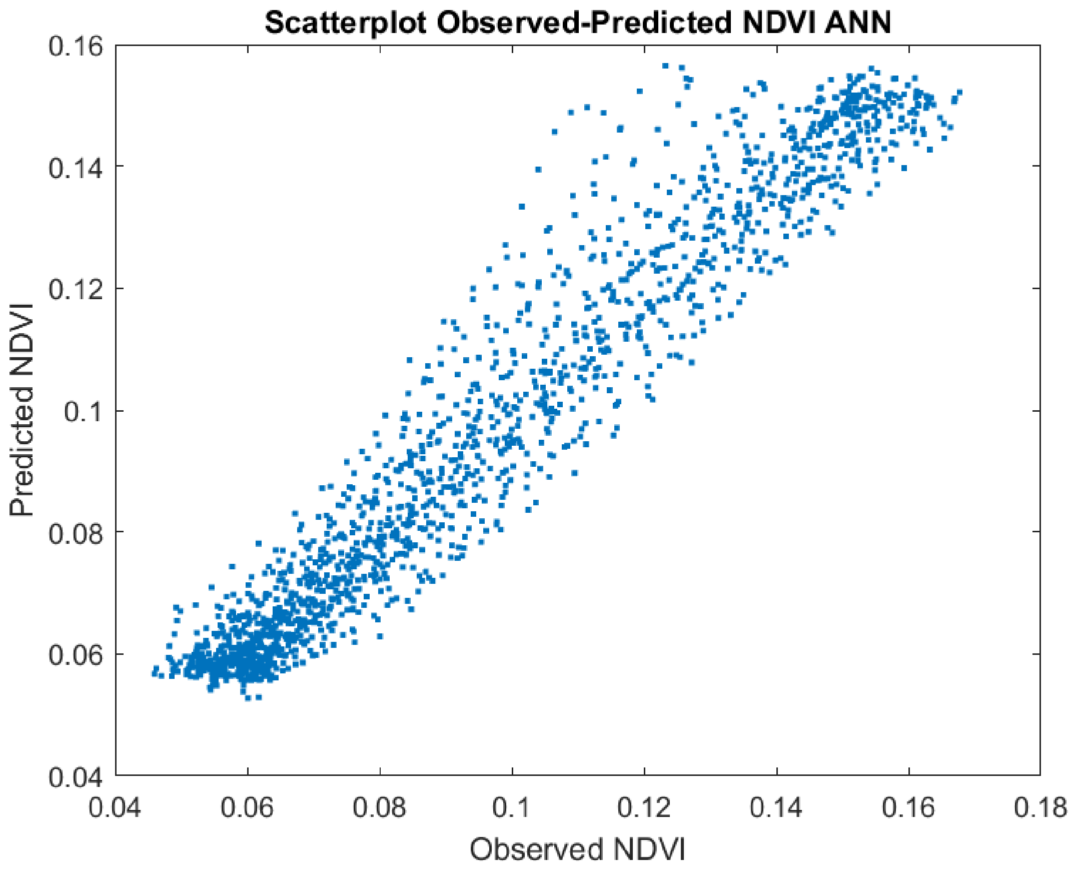


Scatterplots for NDVI observed and predicted using the ANN model

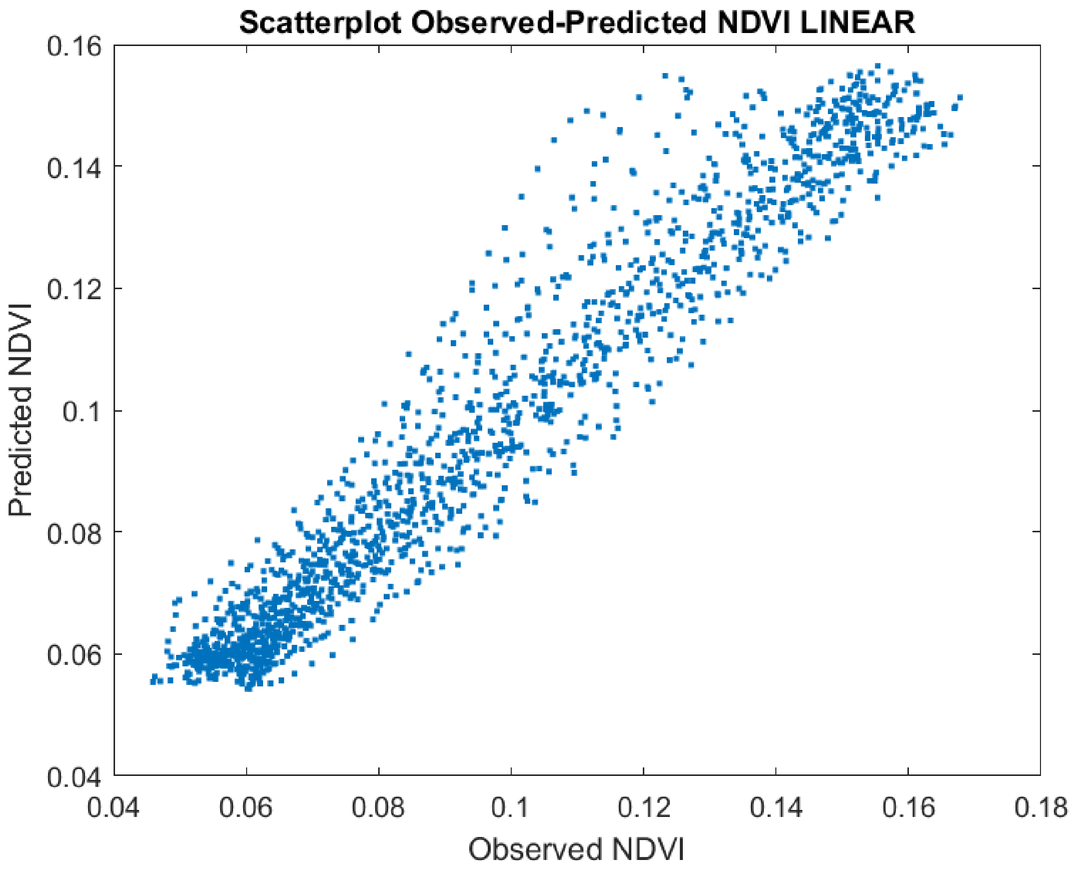


Scatterplots for NDVI observed and predicted using the linear model

BLUE NILE SCATTERPLOTS

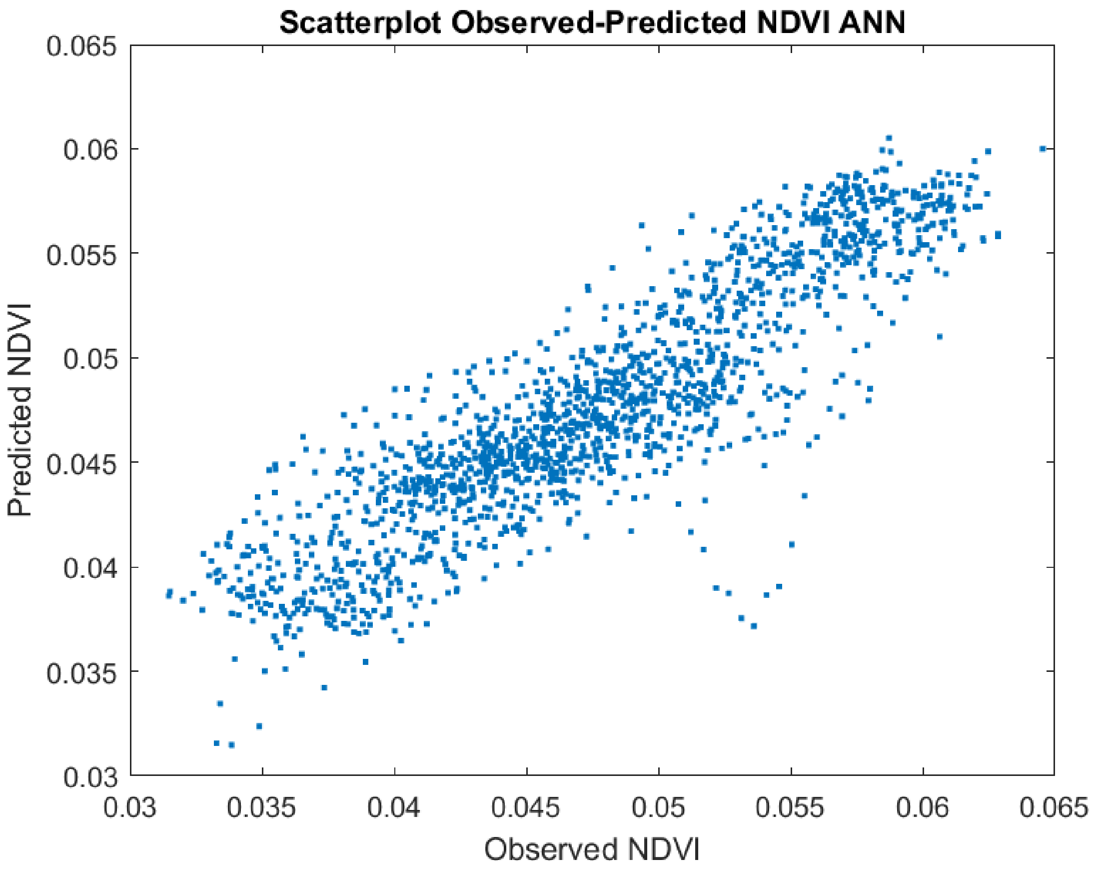


Scatterplots for NDVI observed and predicted using the ANN model

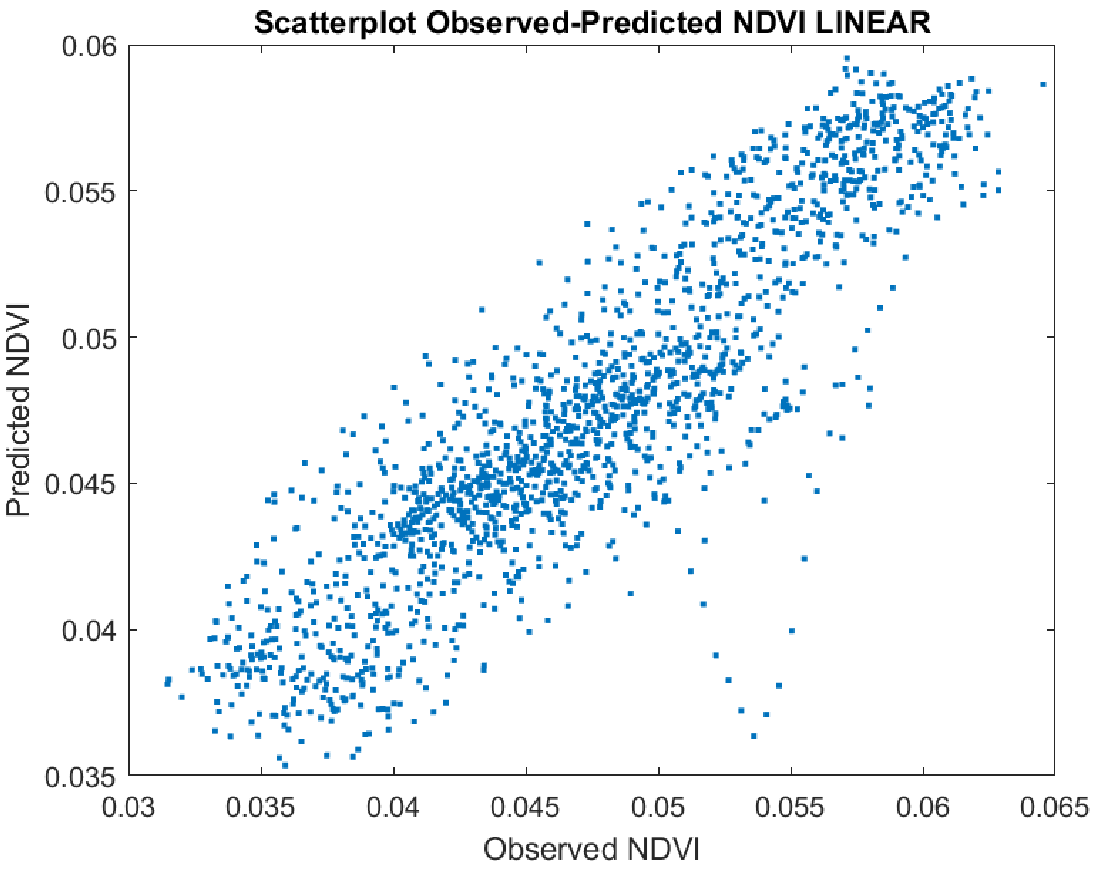


Scatterplots for NDVI observed and predicted using the linear model

MAIN NILE SCATTERPLOTS

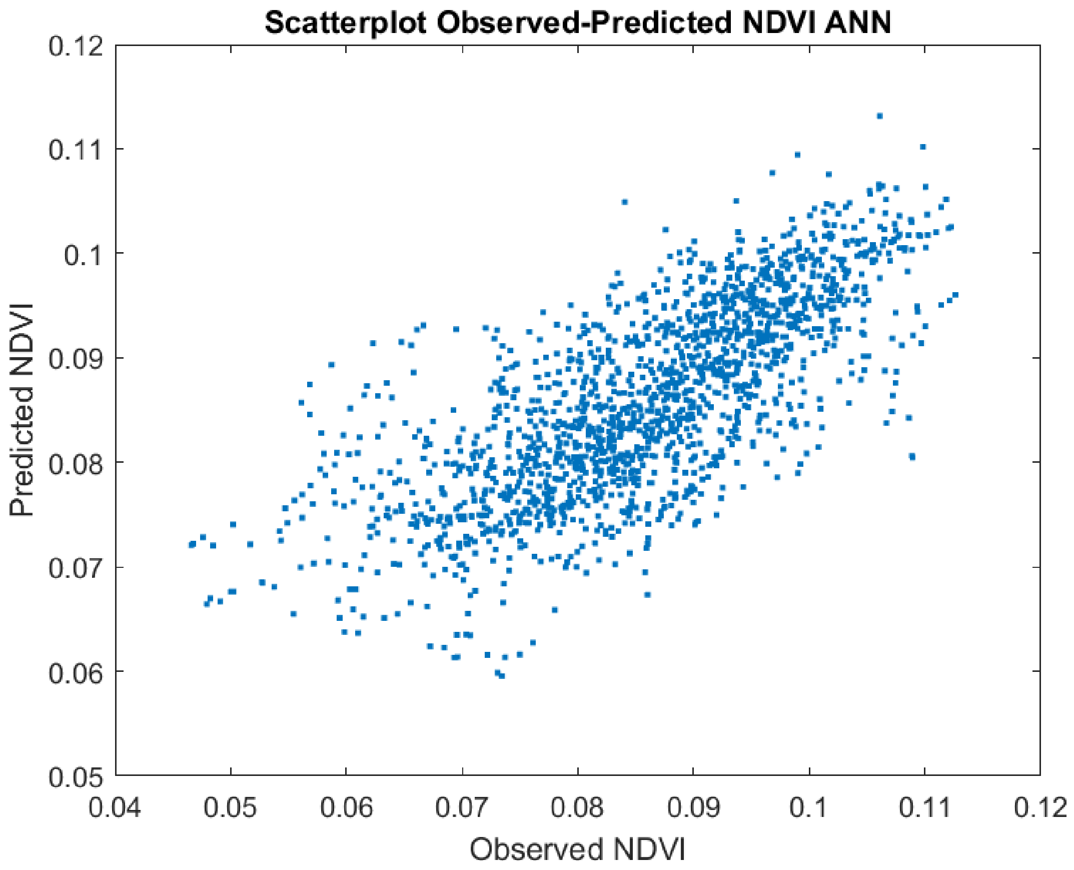


Scatterplots for NDVI observed and predicted using the ANN model

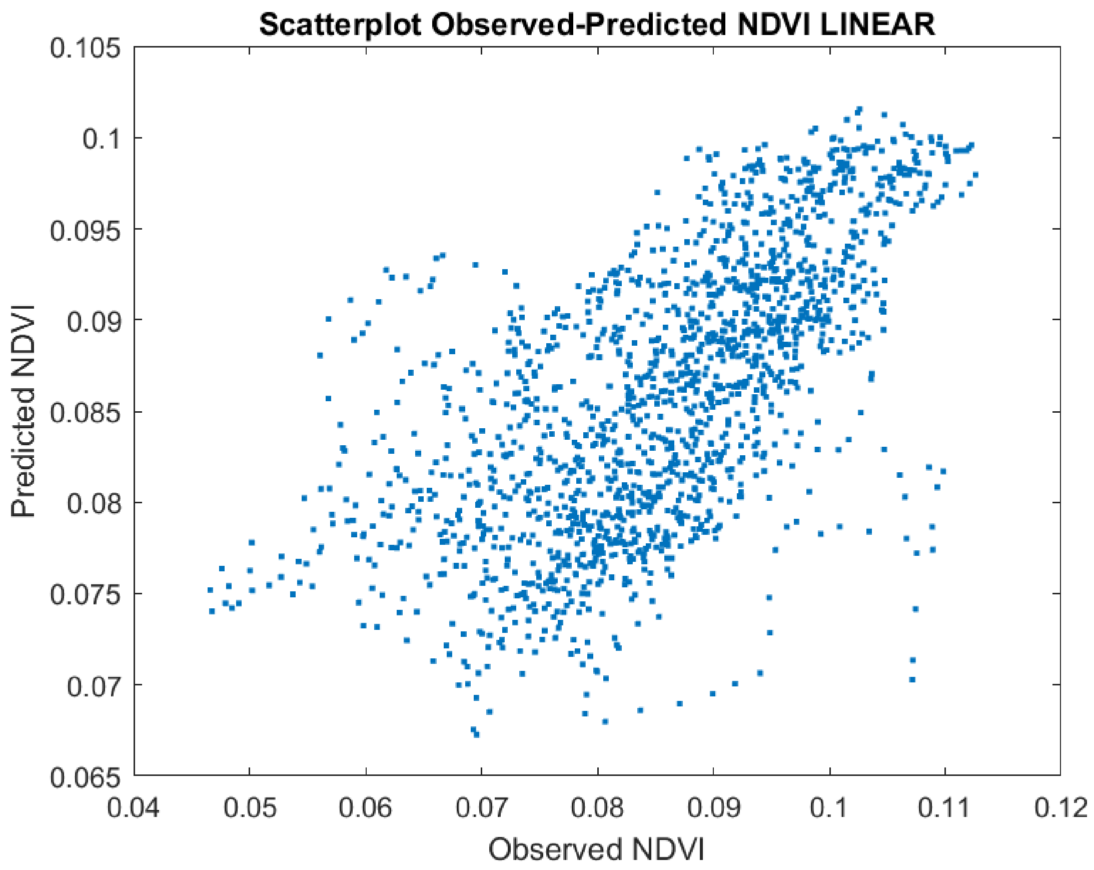


Scatterplots for NDVI observed and predicted using the linear model

LAKE ALBERT SCATTERPLOTS



Scatterplots for NDVI observed and predicted using the ANN model



Scatterplots for NDVI observed and predicted using the linear model