COMPARING PHYSICS-BASED AND EMPIRICAL-STATISTICAL METHODS OF REPRESENTING HYDROLOGY

Eric Swain, Melinda Lohman, and Julieta Gomez

Principal Investigators:

Eric Swain, Research Hydrologist, Davie Office, FL

Melinda Lohman, Hydrologist, Davie, Office, FL

Julieta Gomez, Hydrologist, San Juan Office, PR

Outline

- Physics-based models incorporate equations that express hydrologic dynamics and transport
- Empirical-statistical models derive functional relationships between hydrologic variables
- Each method has advantages and disadvantages
- The choice of method depends on the data available and the purpose of the simulation

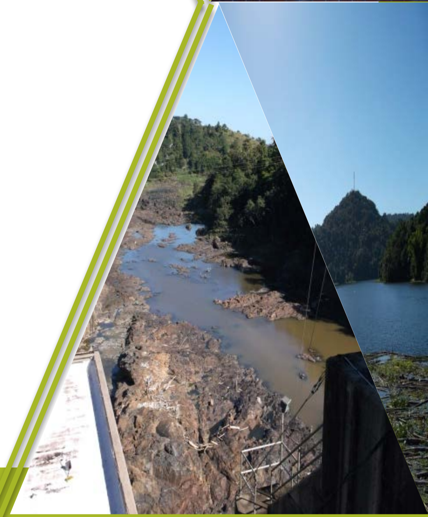
Advantages of each method

- The physics-based methods should represent the actual factors that affect the hydrology, and therefore be more robust in representing various conditions

- The empirical-statistical methods can be fit to only the relevant parameters while accommodating lack of information so it is simpler and not as easily damaged by bad data

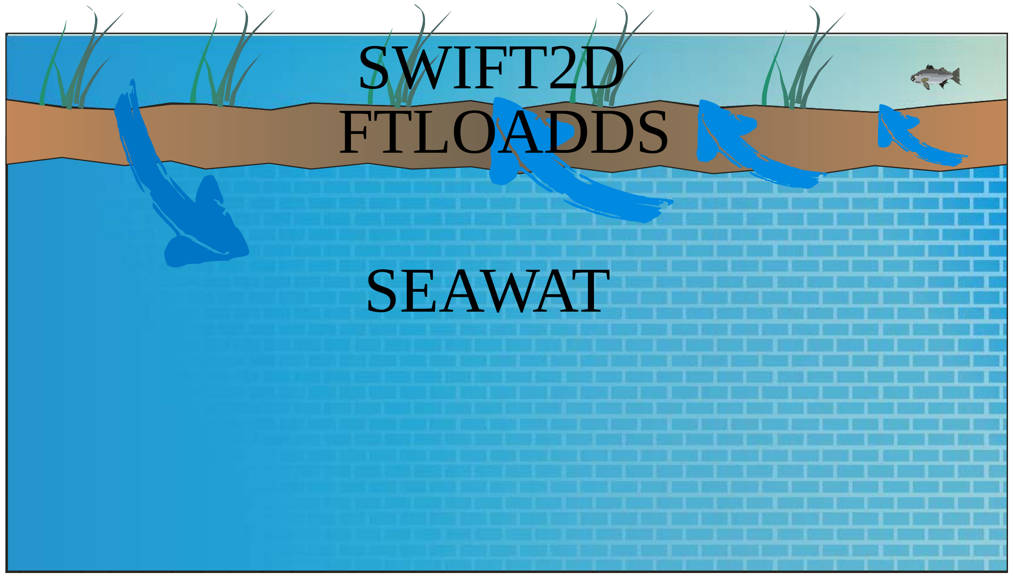
Disadvantages of each method

- The physics-based methods require an accurate delineation of all the relevant quantities, so missing data and unknown input data errors can create difficulties
- The empirical-statistical methods do not account for the actual physics, so when hydrologic conditions change the applicability of the model becomes quite uncertain

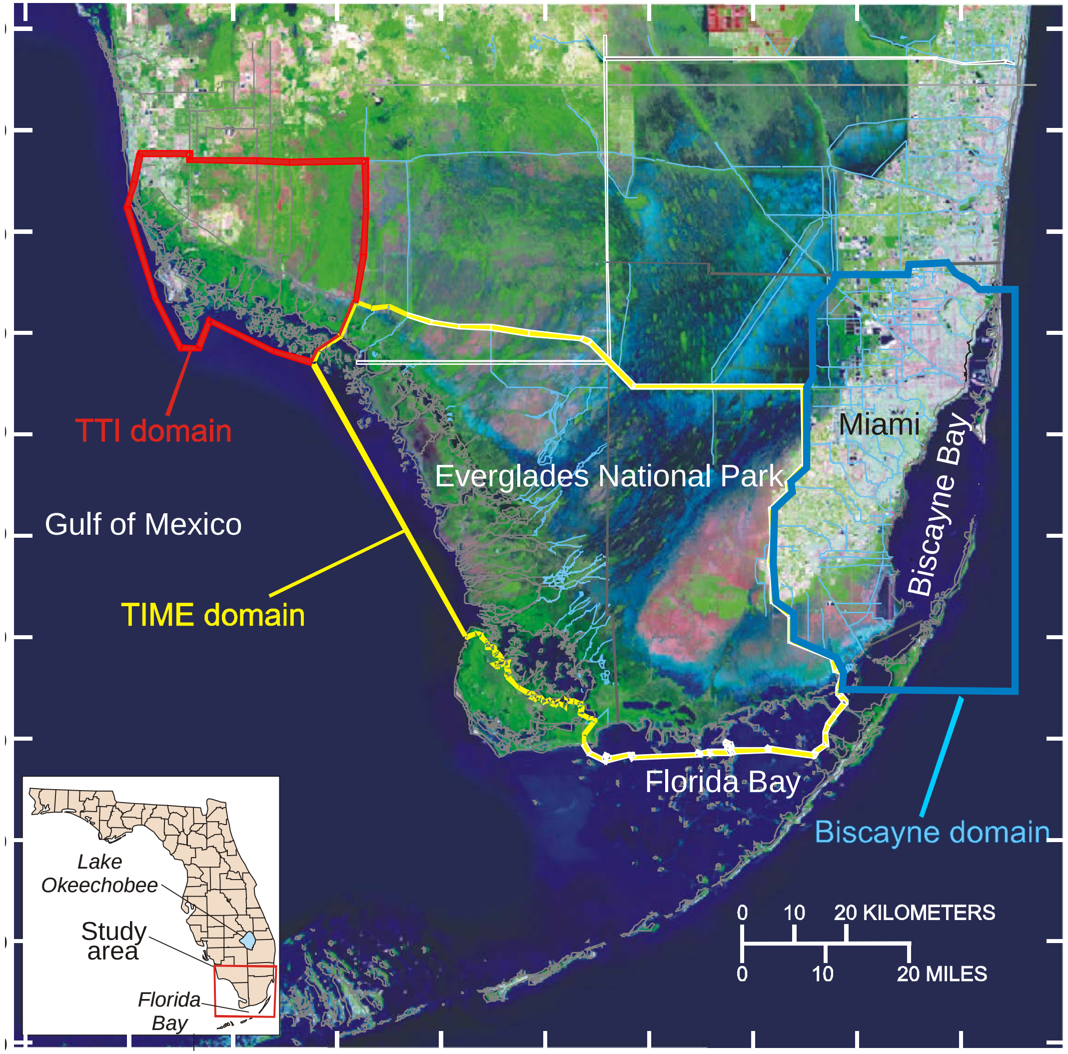


My example of a physics-based model

- FTLOADDS (Flow and Transport in a Linked Overland/Aquifer Density Dependent System) Combines:
  - **SWIFT2D** hydrodynamic surface water code
  - **SEAWAT** variable density ground-water flow and transport code
- Satisfies requirements for modeling South Florida
  - Hydrodynamic representation of surface water in two-dimensions
  - Three dimensional representation of groundwater
  - Salinity transport is represented in each model and passed with leakage
- Modifications
  - Heat Transport
  - Interfaces with other models



Project History of FTLOADDS Code



Code first applied to Florida Bay area in Southern Inland and Coastal Systems (SICS) model to examine coastal interactions

Tides and Inflows in the Mangrove Ecotone (TIME) model developed for Everglades National Park area

Application to Ten Thousand Islands area including heat transport for temperature computation

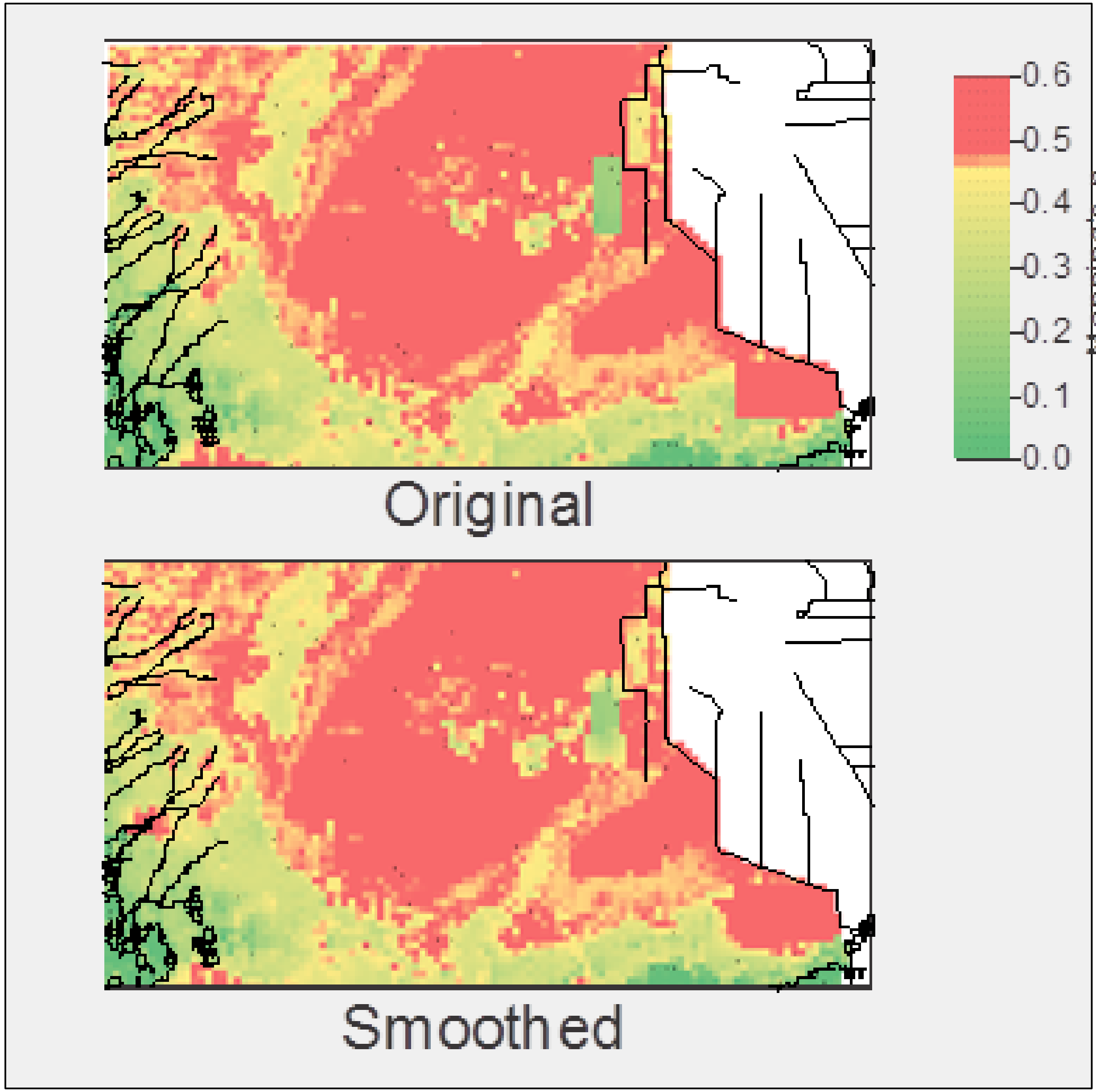
BISCAYNE application to the coastal and urban area of Biscayne Bay to examine hypersalinity events

TIME and BISCAYNE applications combined to produce Biscayne Southern Everglades Coastal Transport (BISECT) model

Field studies determine model parameters

Evapotranspiration
Vegetative frictional resistance
Land-surface elevation
Wind frictional effects
Measurement of coastal discharge
Aquifer properties
Vegetation mapping
Airborne resistivity measurements
Salinity mapping
Stage and salinity timeseries data

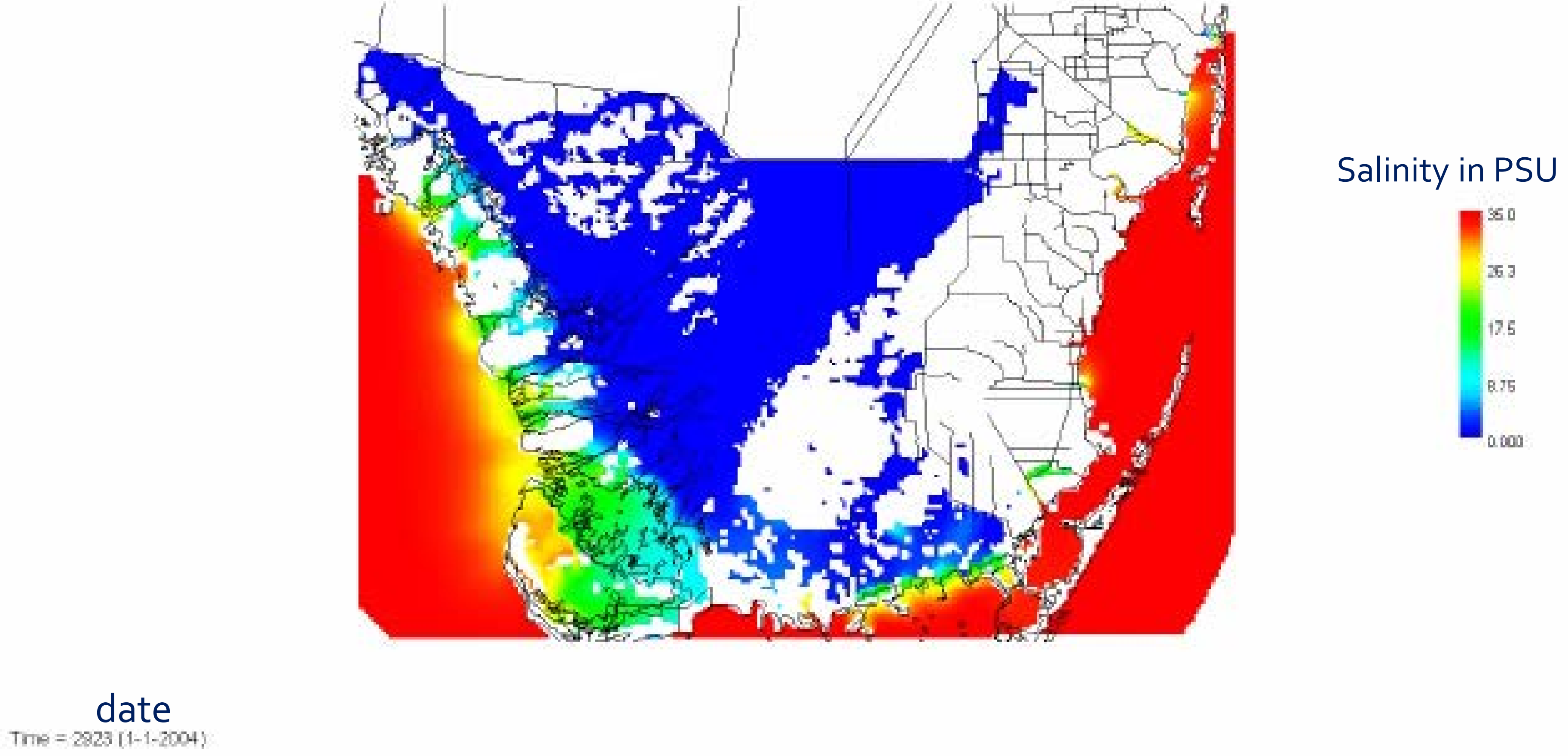
Field studies determine model parameters



Calibration and smoothing of Mannings *n* values

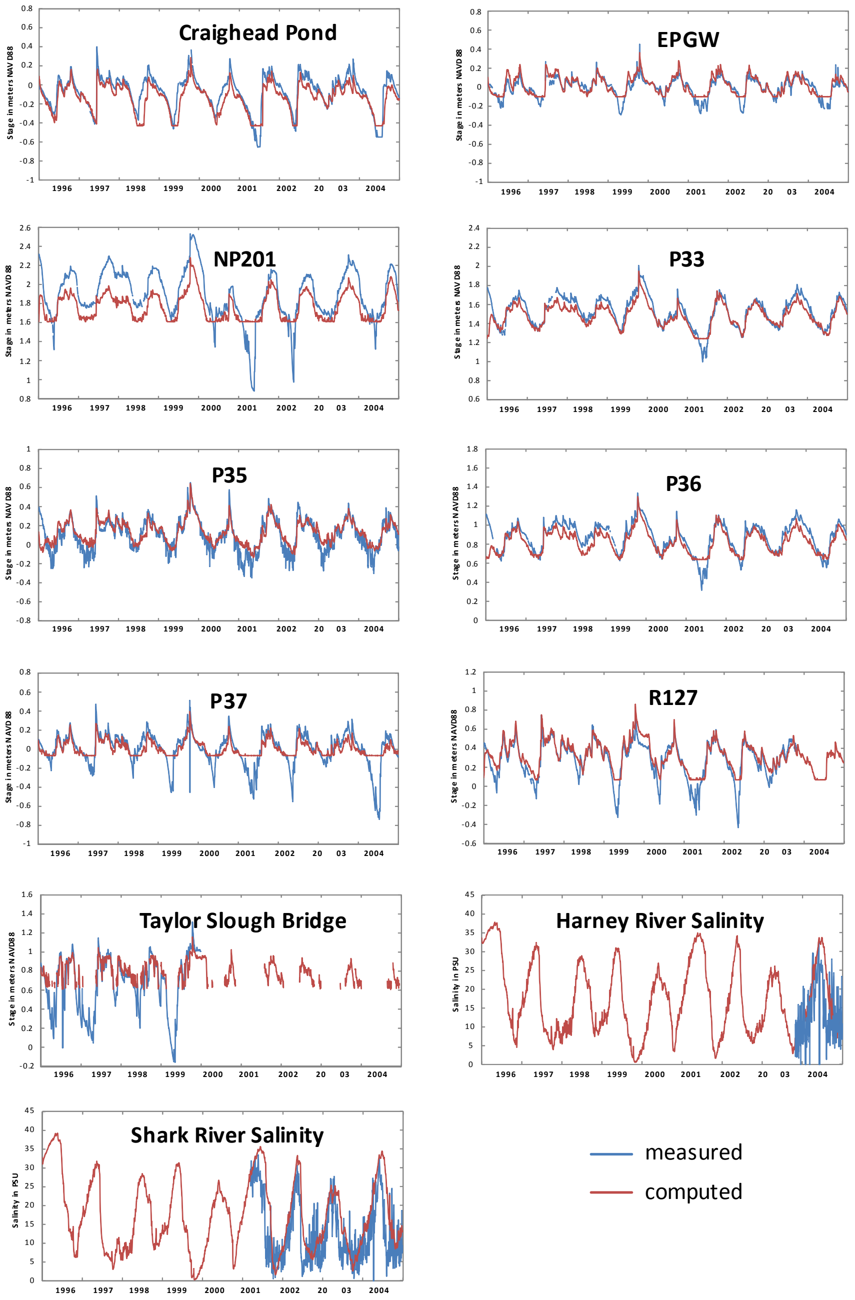
River	Original percent error in cumulative flows	Refined percent error in cumulative flows
Broad	+11.61	+2.34
Chatham	+7.89	+4.06
Harney	N.A.	-4.02
Long Sound	-6.93	-2.53
Lostmans	N.A.	-9.30
McCormick	+1.32	-0.88
Mud	+2.11	+0.35
North	N.A.	+0.02
Shark	+20.78	-0.60
Taylor	-0.54	+2.19
Trout	-2.02	-0.88
Total error	+34.22	-9.25

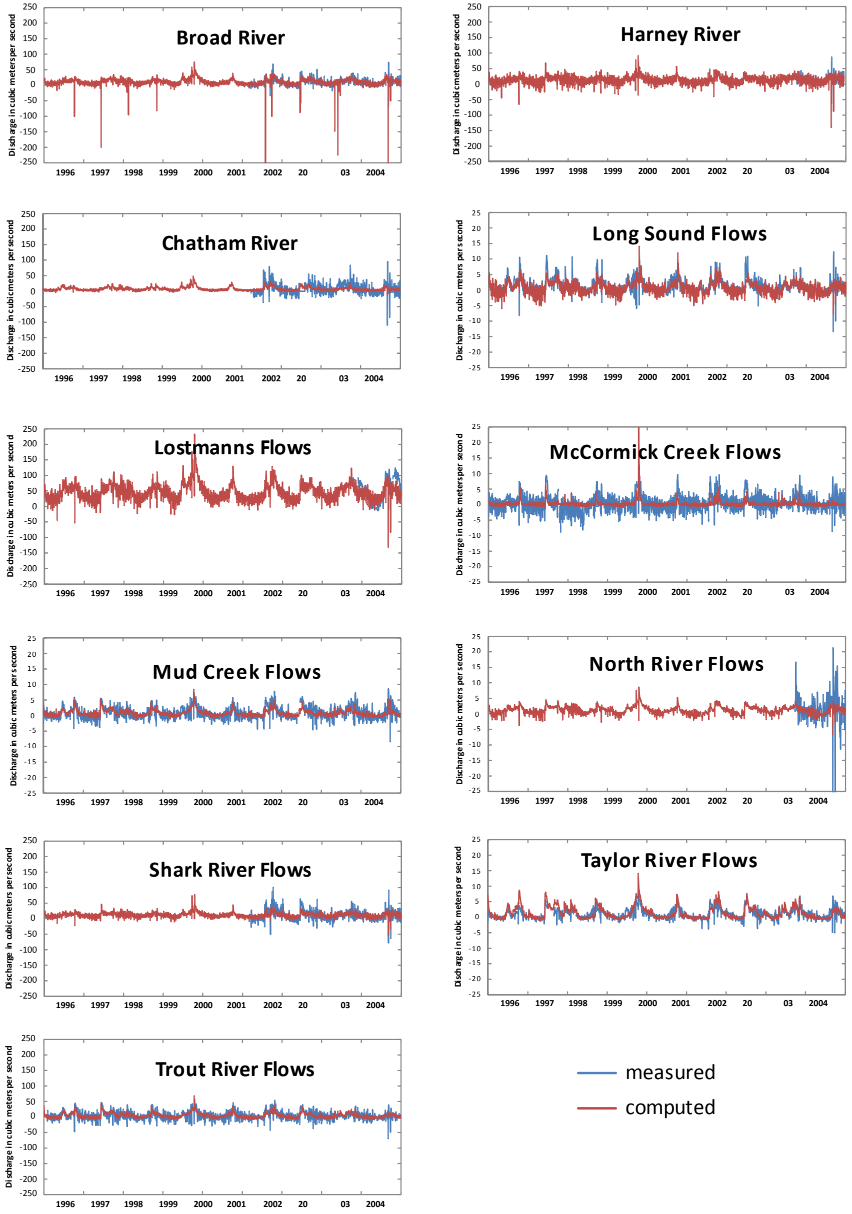
Additional frictional adjustments made at coastal rivers



BISECT simulated salinity and inundation

Measured and simulated stages and salinities





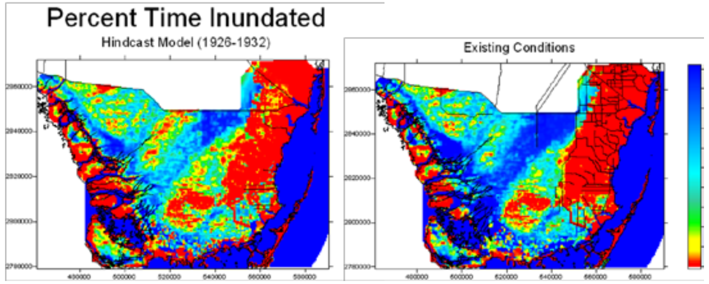
Measured and simulated flows

Stage, Discharge, Salinity, and Temperature Statistics

Location and Parameter	N-S Mean Adj,	N-S	RMSE	ME
Craighead Pond Stage	0.774	0.637	0.106 m	-0.065 m
EPGW Stage	0.625	0.624	0.059 m	0.003 m
NP-201 Stage	0.786	0.064	0.193 m	-0.170 m
P-33 Stage	0.831	0.721	0.074 m	-0.046 m
P-35 Stage	0.741	0.714	0.089 m	0.027 m
P-36 Stage	0.811	0.634	0.082 m	-0.057 m
P-37 Stage	0.721	0.465	0.066 m	-0.046 m
R-127 Stage	0.821	0.754	0.073 m	0.038 m
Taylor Slough Bridge Stage	0.720	0.715	0.084 m	0.012 m
Broad River Discharge	0.413	0.331	8.808 m^3/s	3.073 m^3/s
Chatham River Discharge	0.438	0.400	13.006 m^3/s	3.263 m^3/s
Harney River Discharge	0.388	-0.125	13.119 m^3/s	-8.860 m^3/s
Long Sound Discharge	0.465	0.324	1.691 m^3/s	-0.772 m^3/s
Lostmans River Discharge	0.296	-0.099	41.008 m^3/s	-24.573 m^3/s
McCormick River Discharge	0.085	0.075	2.503 m^3/s	-0.264 m^3/s
Mud Creek Discharge	0.441	0.438	1.403 m^3/s	0.105 m^3/s
North River Discharge	0.355	0.355	3.795 m^3/s	0.035 m^3/s
Shark River Discharge	0.374	0.373	13.039 m^3/s	-0.455 m^3/s
Taylor River Discharge	-0.172	-0.350	1.864 m^3/s	0.677 m^3/s
Trout Creek Discharge	0.366	0.366	10.199 m^3/s	-0.265 m^3/s
Harney River Salinity	0.224	0.021	7.107 PSU	3.235 PSU
Shark River Salinity	0.501	0.191	7.331 PSU	4.540 PSU
P-33 Temperature	0.921	0.746	2.058 ^{o}C	-1.710 ^{o}C
Shark River Temperature	0.693	0.250	3.180 ^{o}C	-2.445 ^{o}C

Applications of BISECT

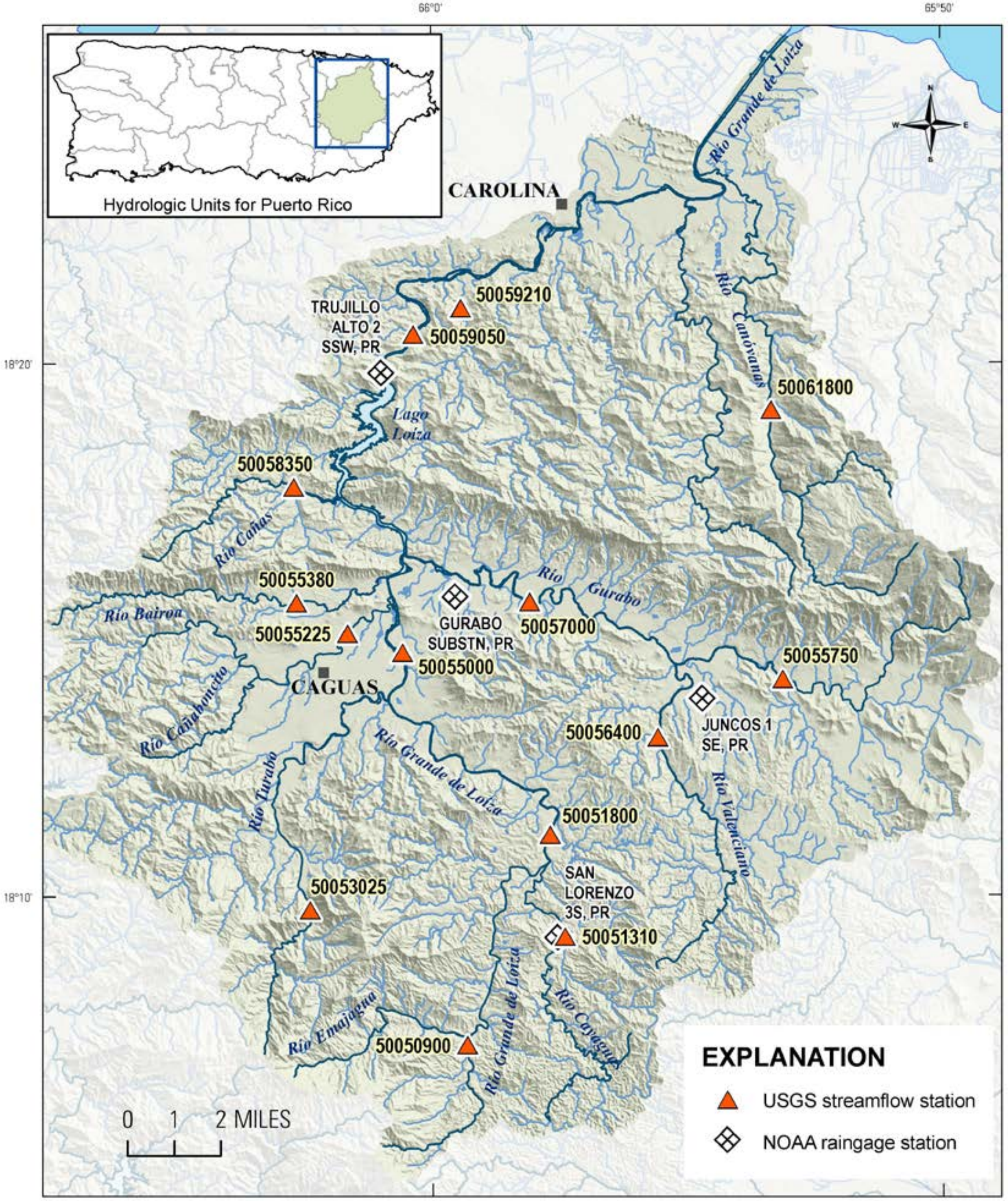
- Hindcast (1926-1940) with vegetation dynamics analyses
- Storm events in various historic periods
- Futurecast (2038-2057) using downscaled Global Climate Model data
- Effects of sea-level rise and restoration changes



My example of a empirical-statistical model

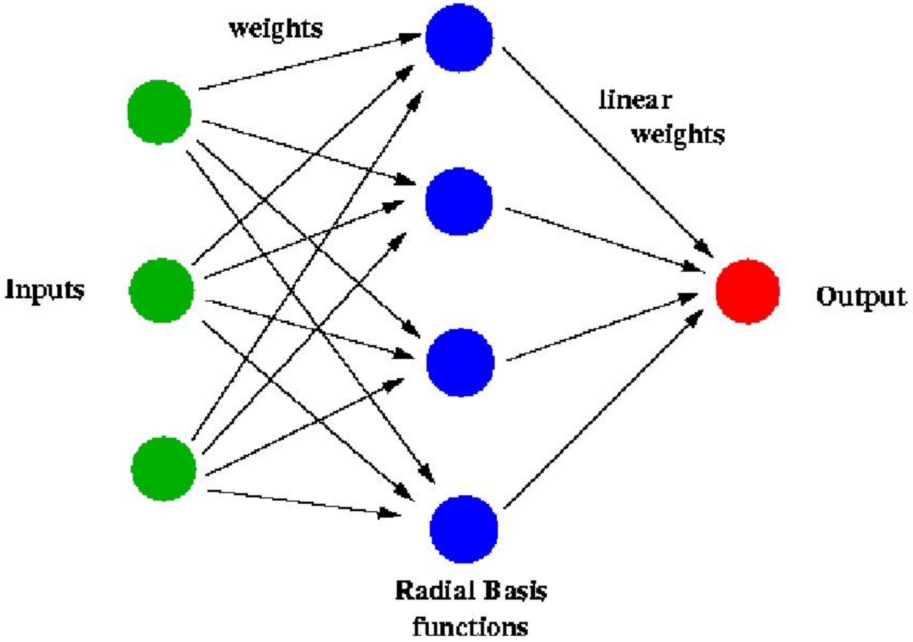
Artificial Neural-Network application to Lago Loíza Reservoir, Puerto Rico

- Rio Grande de Loíza hydrographic basin
- 4 NOAA raingage stations
- USGS Streamflow stations



Artificial Neural Network (ANN)

- A family of statistical learning models inspired by biological neural networks used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown.



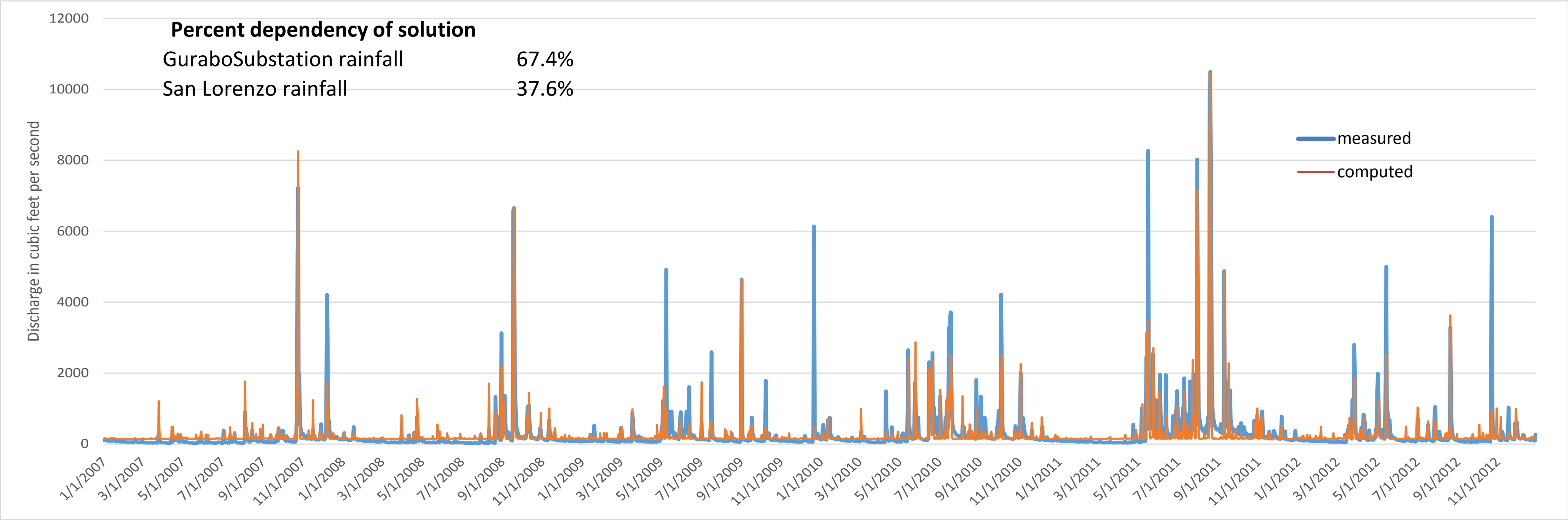
Rainfall Dataset

- Raingages operated by NOAA
  - Trujillo Alto 2 SSW , Gurabo Substation , Juncos 1 SE and San Lorenzo 3S
- Streamflow data of 5 USGS-gaged stations.

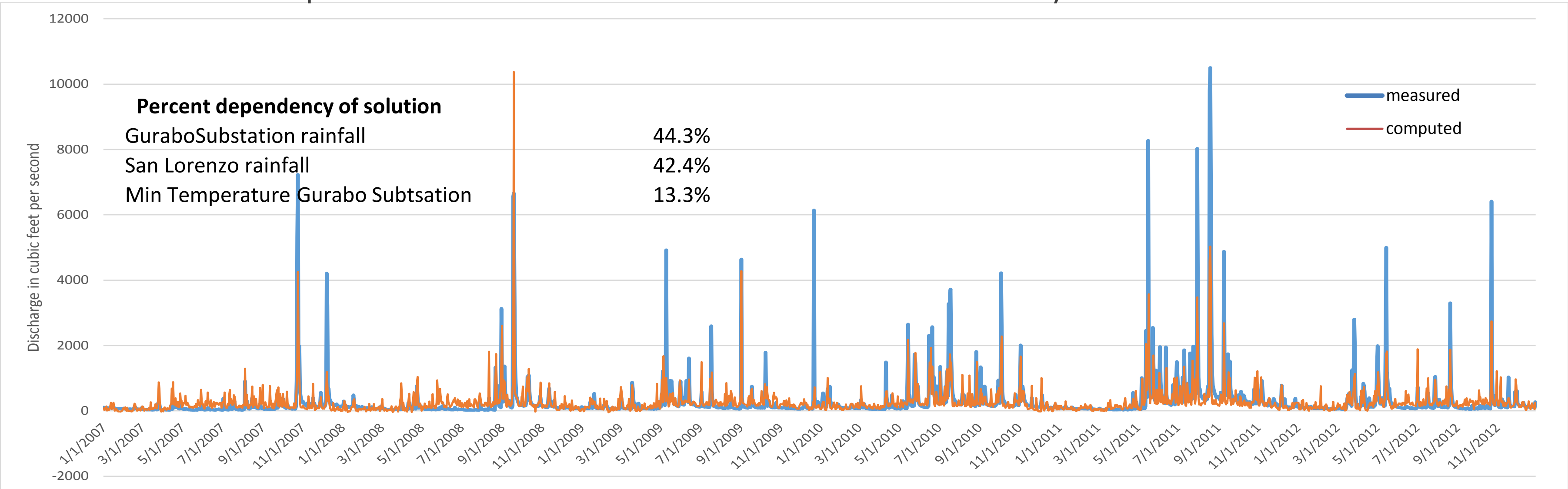
USGS Station Number	Station Name	Drainage area, in square miles	Total Annual discharge, in cubic feet per second (Water year 2014)	Annual Mean discharge, in cubic feet per second (Water year 2014)
50055000	Rio Grande de Loiza at Caguas, PR	89.8	81,970	224.6
50057000	Rio Gurabo at Gurabo, PR	60.2	40,060	109.8
50055225	Río Cagüitas at Villa Blanca at Caguas, PR	16.6	12,930	35.4
50058350	Río Cañas at Río Cañas, PR	7.53	4,668	12.8
50055380	Río Bairoa above Bairoa, Caguas, PR	4.74	2,824	7.74

Preliminary Results- station 50055000

- All variables are set as optional



- Minimum temperature at Gurabo substation is set as mandatory



Input Set Preparation

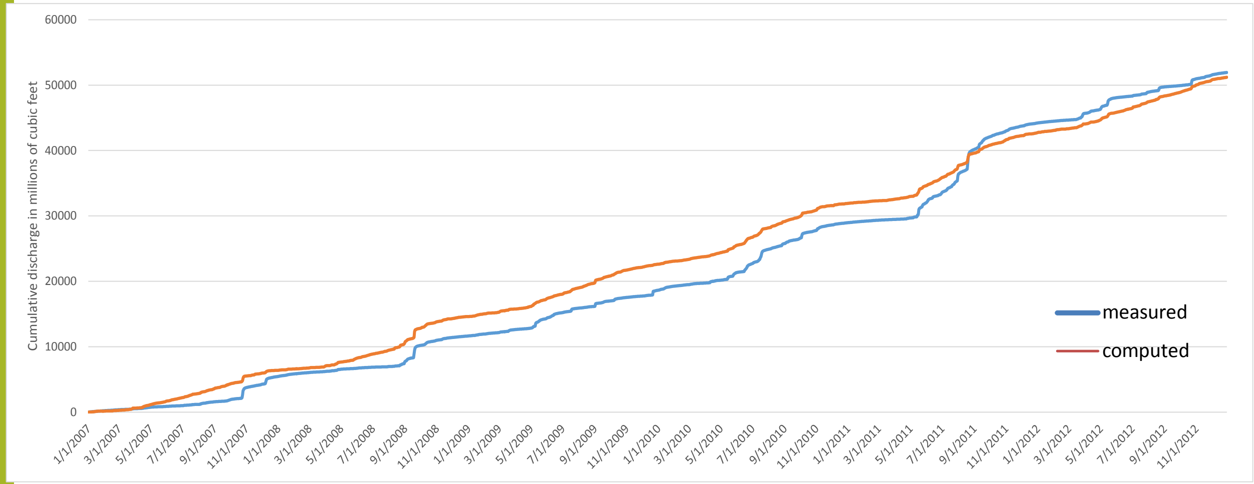
To provide a more consistent solution, an estimated evapotranspiration (ET) function is developed based on the Hargreaves equation (Hargreaves and Samani. 1985):

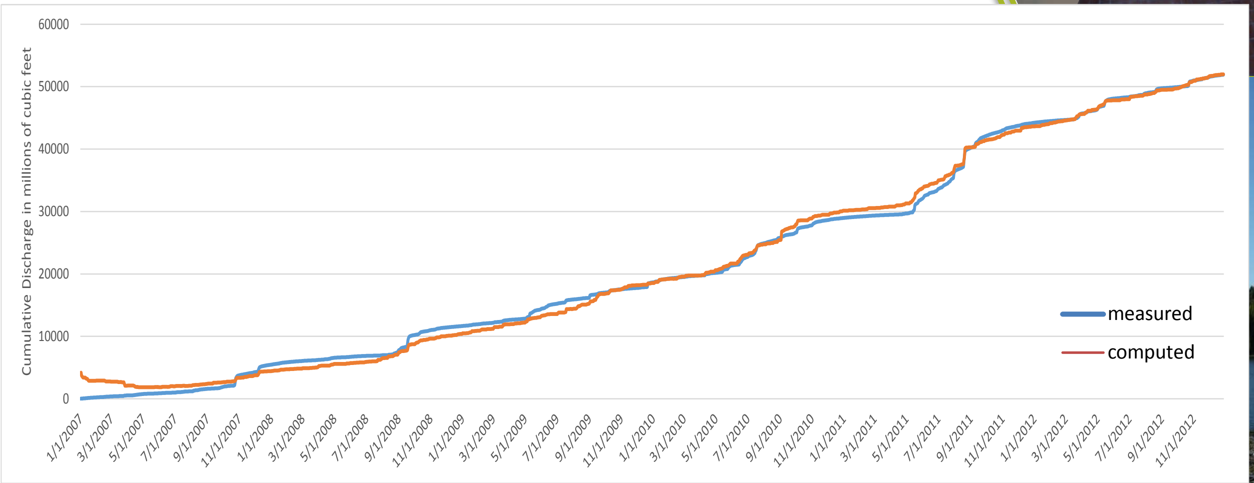
$$PET = C_1 R_a (T + C_2)\sqrt{T_{max} - T_{min}}.$$

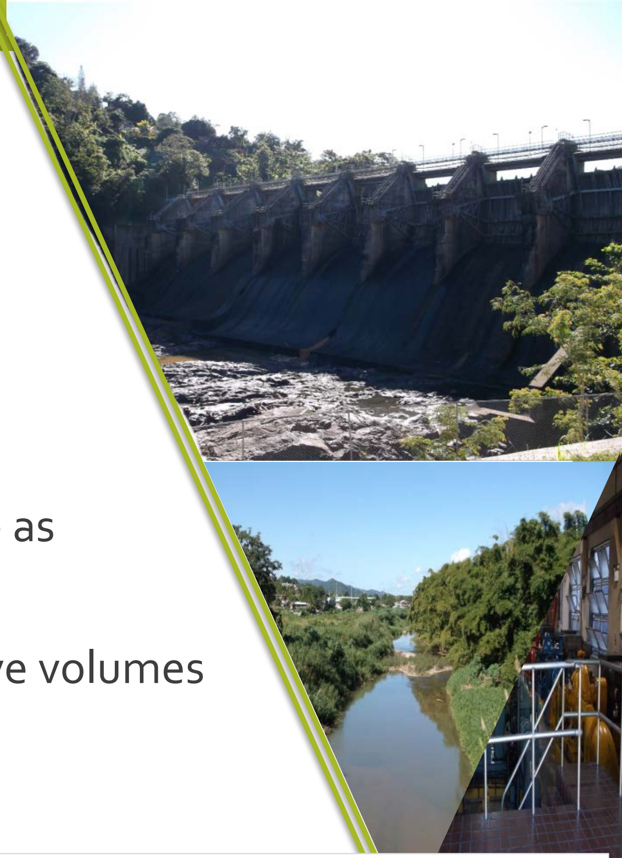
This is used in place of the measured temperatures in the model input

Accounting for cumulative discharge

- Cumulative discharge volume curves derived from simulated discharge at 50055000
- Further simulations included cumulative discharge volume as desired output
- Better matching of cumulative volumes with mean error of 5.322 cfs



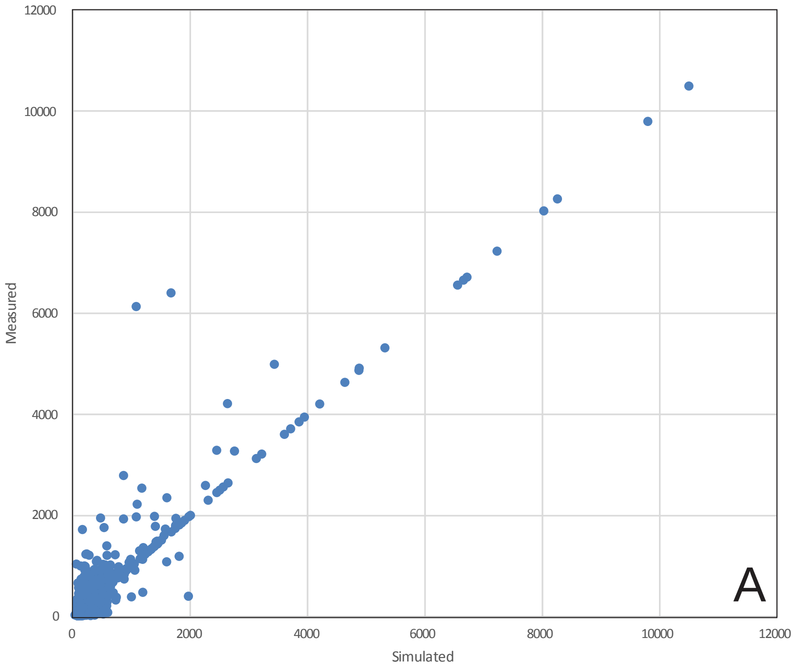




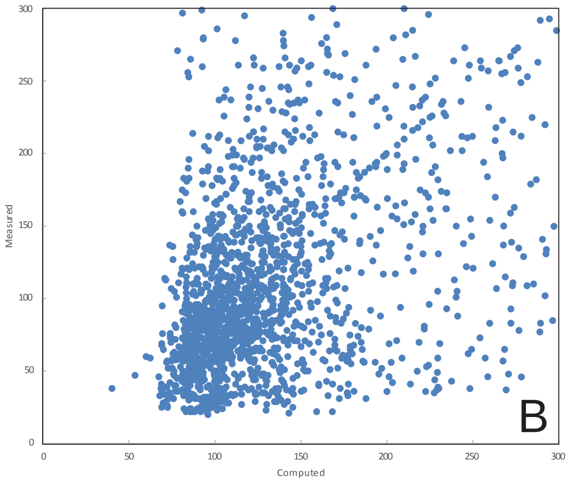
Simulation Identification	Input Dataset Descriptor	Period of record	Output variable
1A ANN training	Daily precipitation at selected raingage stations decorrelated to Gurabo Substation and 30-day average evapotranspiration	2007-2012	Daily discharge for selected USGS-gaged stations
1B ANN training	30-day cumulative daily precipitation and 30-day cumulative daily evapotranspiration		30-day cumulative discharge for selected USGS-gaged stations
2A ANN application	Daily precipitation at selected raingage stations decorrelated to Gurabo Substation and 30-day average evapotranspiration	1994-1997	Daily discharge for selected USGS-gaged stations
2B ANN application	30-day cumulative daily precipitation and 30-day cumulative daily evapotranspiration		30-day cumulative discharge for selected USGS-gaged stations
3A ANN application	Daily precipitation at selected raingage stations decorrelated to Gurabo Substation and 30-day average evapotranspiration	2050-2055	Daily discharge for selected USGS-gaged stations
3B ANN application	30-day cumulative daily precipitation and 30-day cumulative daily evapotranspiration		30-day cumulative discharge for selected USGS-gaged stations

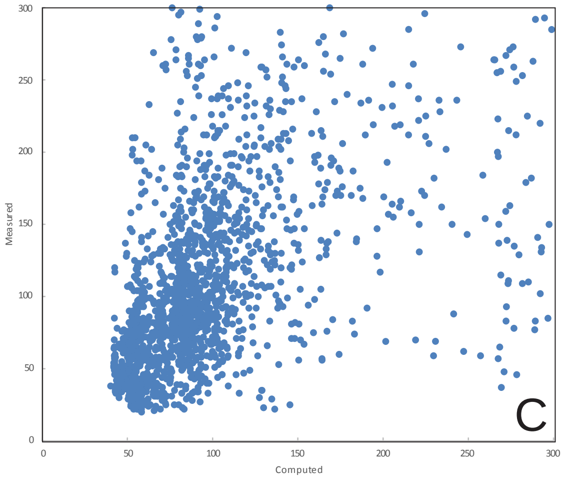
- "Two-stage" solution
  - Adjust the daily discharge data resulting in a reduction of low flow estimates
  - Values of the simulated daily flow less than the simulated average flow are compared with the 30-day average flow times a multiplier
  - Whichever value is lower is used as the final value, as the ANN solutions tend to overestimate low flows

- Solution Optimization
  - 30-day cumulative values is included in the analysis to counter measures the effects of preceding values in cumulative volume output desired scenarios
  - Two- stage method for minimum flows– scheme developed to post-processing ANN- simulated daily discharge and the 30- day cumulative values
- A data allocation of 90 percent training, 5 percent validation, and 5 percent evaluation.



Scatter plots of measured and simulated 5005500 2007-2012 daily discharge for A) entire dataset, B) discharges under 300 cfs, C) Two-Stage Method adjusted discharges.





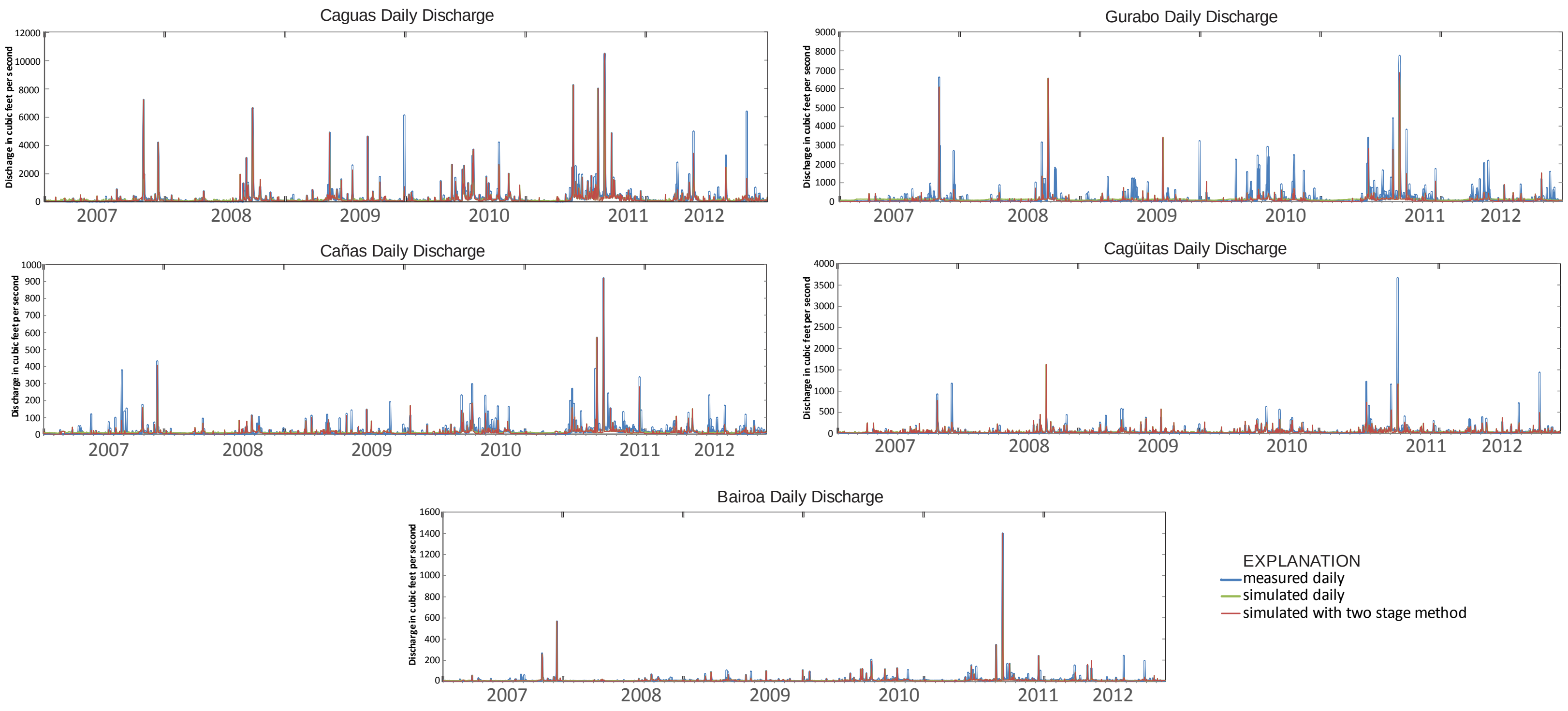
Results – Model training

- Daily and 30-day cumulative timeseries for 2007-2012 generated by the Artificial Neural Network model training for selected streamflow stations within the Lago Loíza reservoir drainage basin, Puerto Rico.

- Gurabo Substation is the most commonly used rainfall stations, and Trijillo Alto 2 SSW is the least used.
- Correlations values for the 30-day cumulative values for the 30-day cumulative best fits are consistently better than for the daily best fits

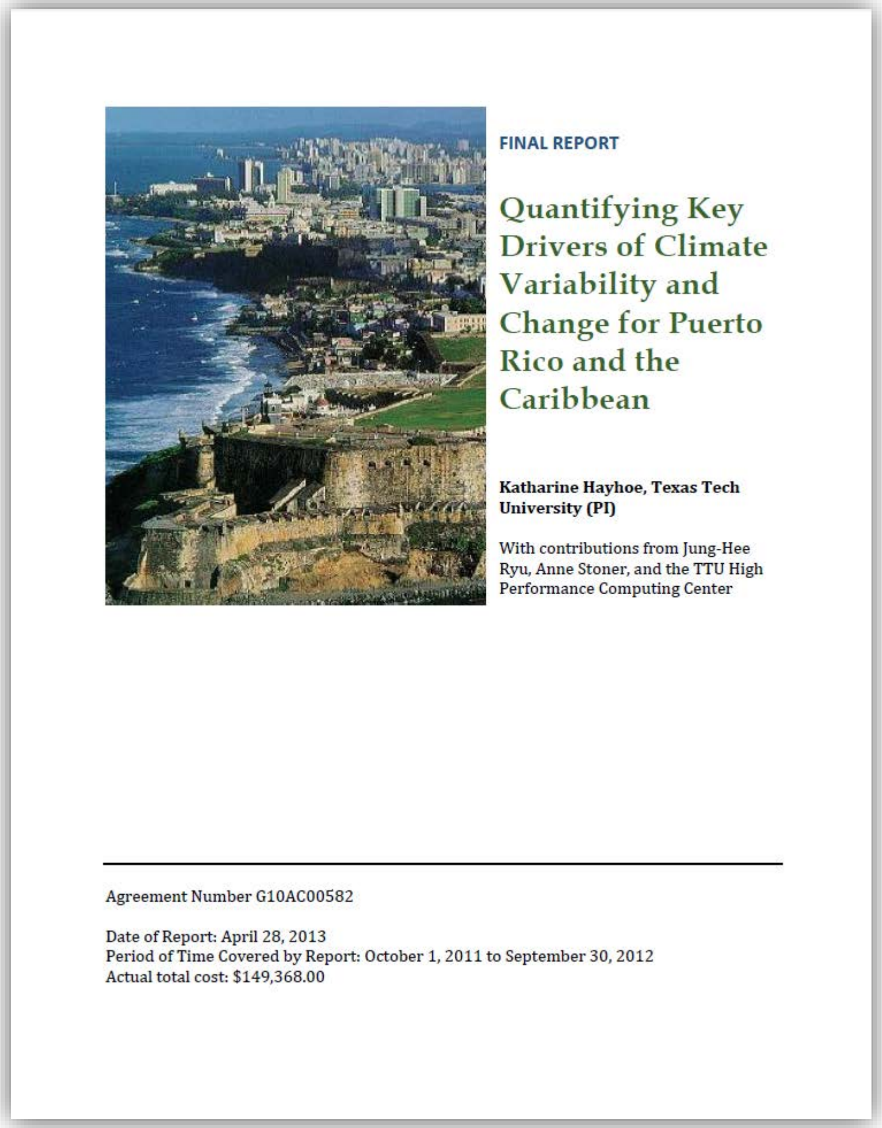
Model training -“Two-stage” Solution

- Two-stage daily discharges for 2007-2012 for selected streamflow stations within the Lago Loíza reservoir drainage basin, Puerto Rico



Climate Projections- Data downscaled from global climate models

- Assess the potential impact of climate changes in Puerto Rico and the Caribbean
- Downscaled simulations from global circulations models
  - Coupled Model Intercomparison Project version 3 and 5 (CMIP3 and CMIP5)
- Provided dataset of precipitation and temperature



FINAL REPORT

Quantifying Key Drivers of Climate Variability and Change for Puerto Rico and the Caribbean

Katharine Hayhoe, Texas Tech University (PI)

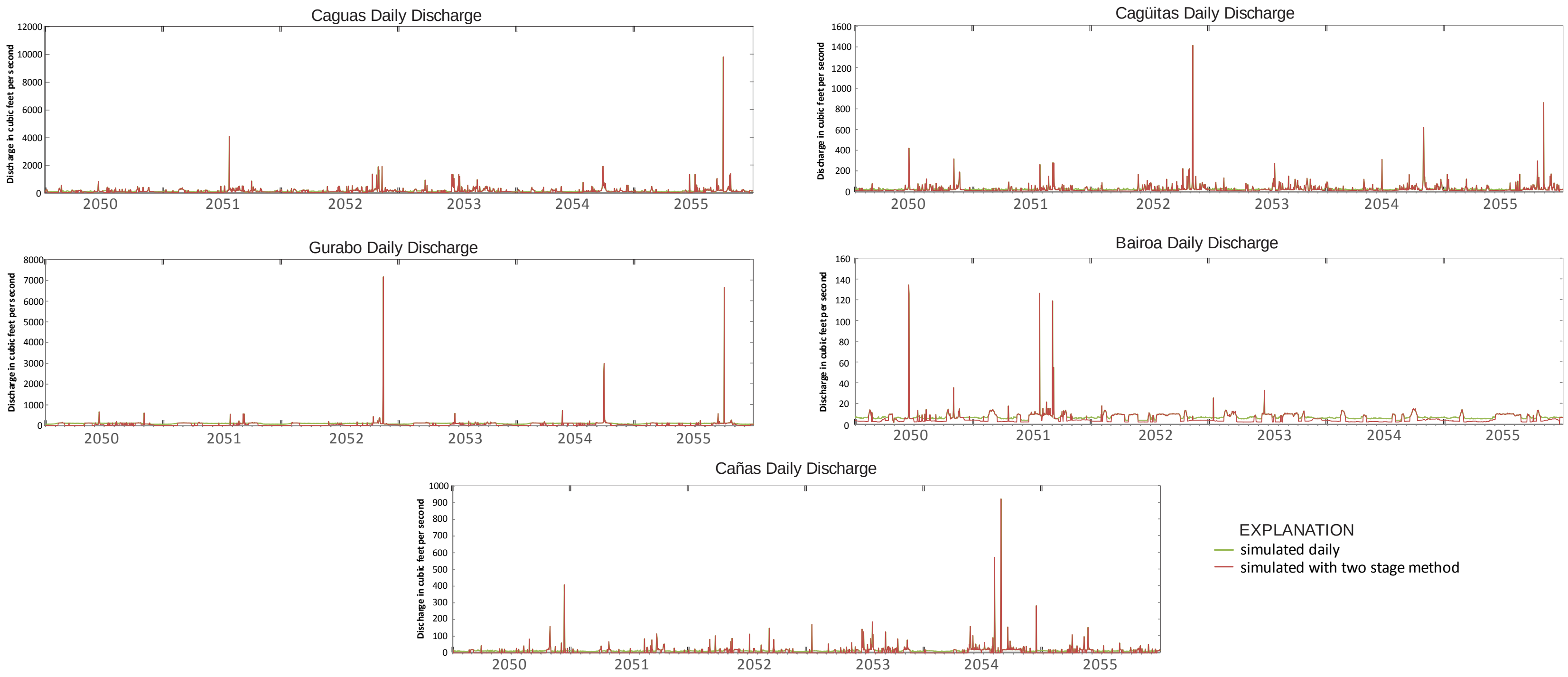
With contributions from Jung-Hee Ryu, Anne Stoner, and the TTU High Performance Computing Center

Agreement Number G10AC00582

Date of Report: April 28, 2013
Period of Time Covered by Report: October 1, 2011 to September 30, 2012
Actual total cost: $149,368.00

Results – Predictive Simulations

- Downscaled rainfall data (2050-2055)



- The total cumulative volumes at the discharge stations is consistently lower in 2050-2055 compared to 2007-2012
- The timeseries for 2050-2055 does not show as many simultaneous high-flow events among stations as does the 2007-2012 timeseries

Water Availability- Lago Loiza reservoir

Input Components

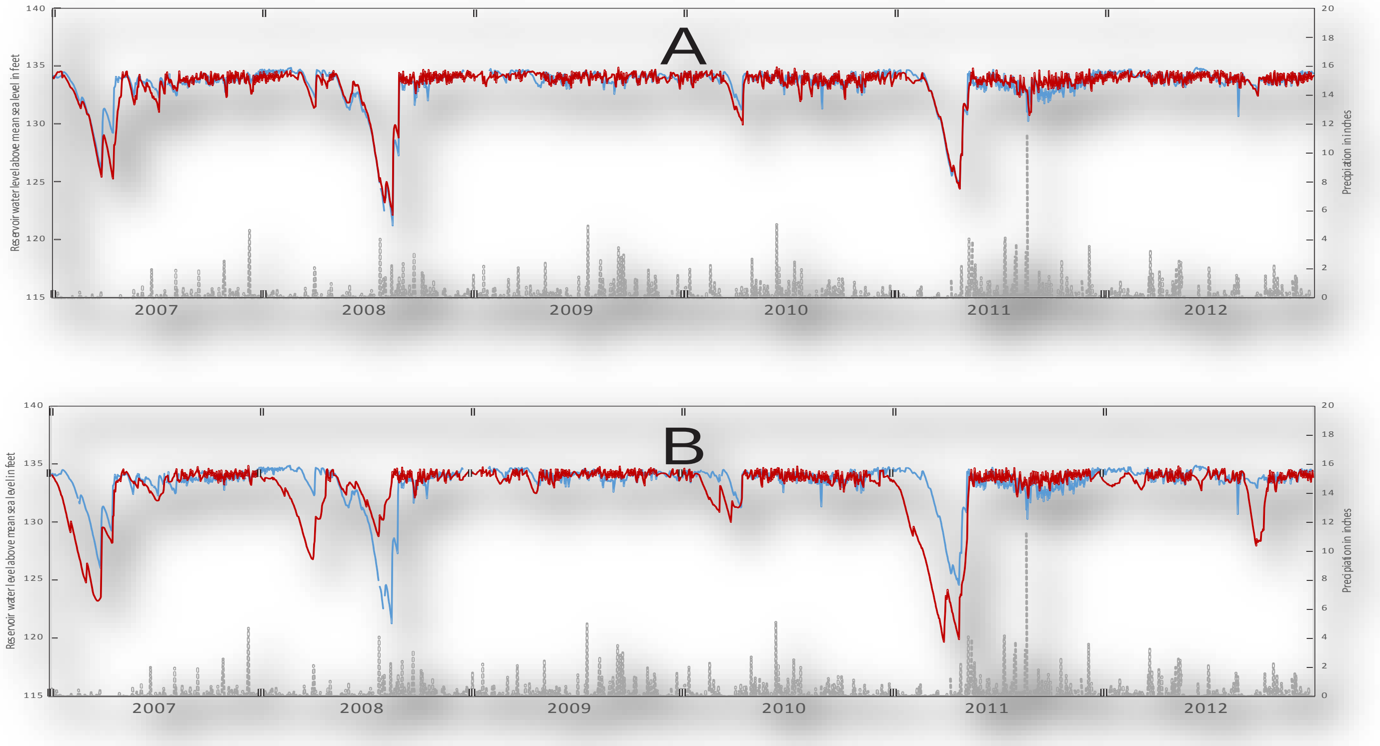
- Measured river inflows and Artificial Neural Network-computed river inflows
- Direct precipitation into reservoir- NOAA weather station (Trujilllo Alto 2 SSW
- Effluents from wastewater treatment plant (approx. 15 mgd)

Output Components

- Water supply- Sergio Cuevas filtration plant (monthly variation)
- Reservoir release – lower part of Rio Grande de Loiza
- Evaporation- assumed from near NOAA weather station

Water Availability- Lago Loiza reservoir

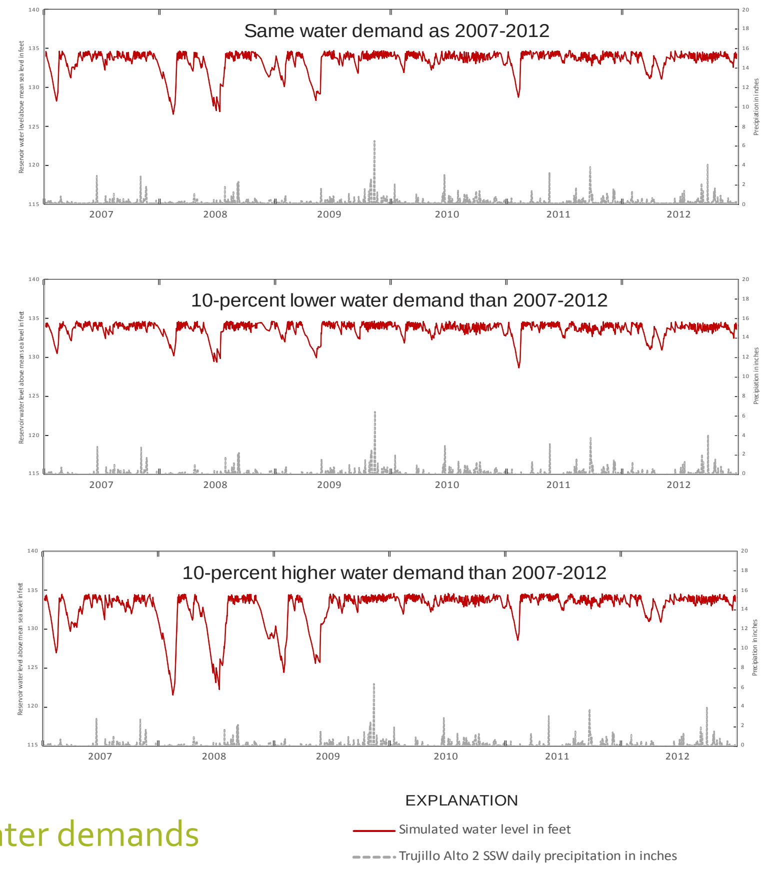
Simulated versus measured water level of Lago Loíza reservoir, Puerto Rico for the period from 2007 to 2012



A) measured river inflows and B) Artificial Neural Network-computed river inflowsBetter match during drop

Water Availability- Lago Loiza reservoir

Water Balance with ANN-simulated river discharges for 2050-2055

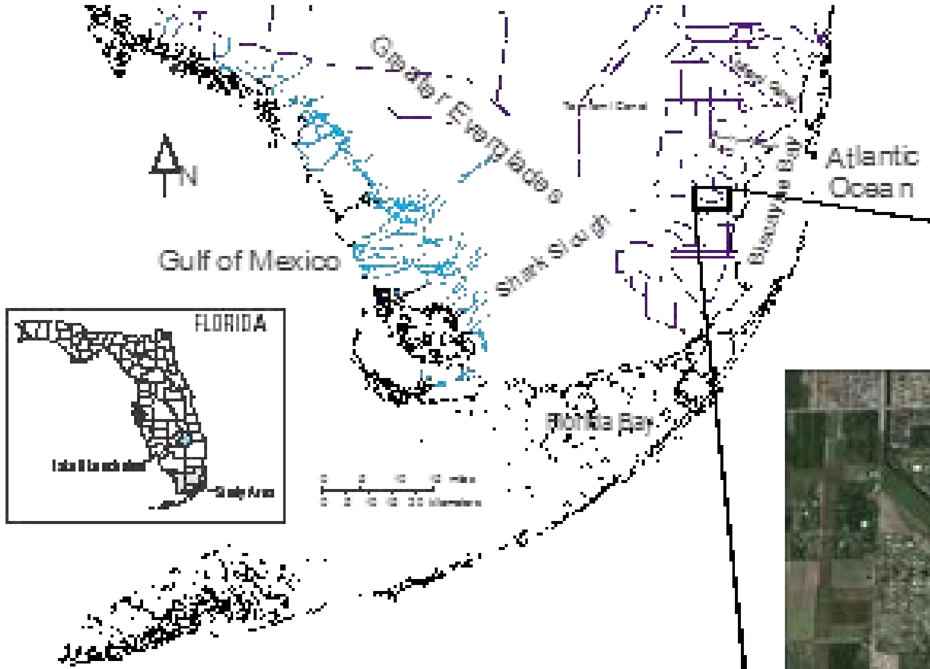


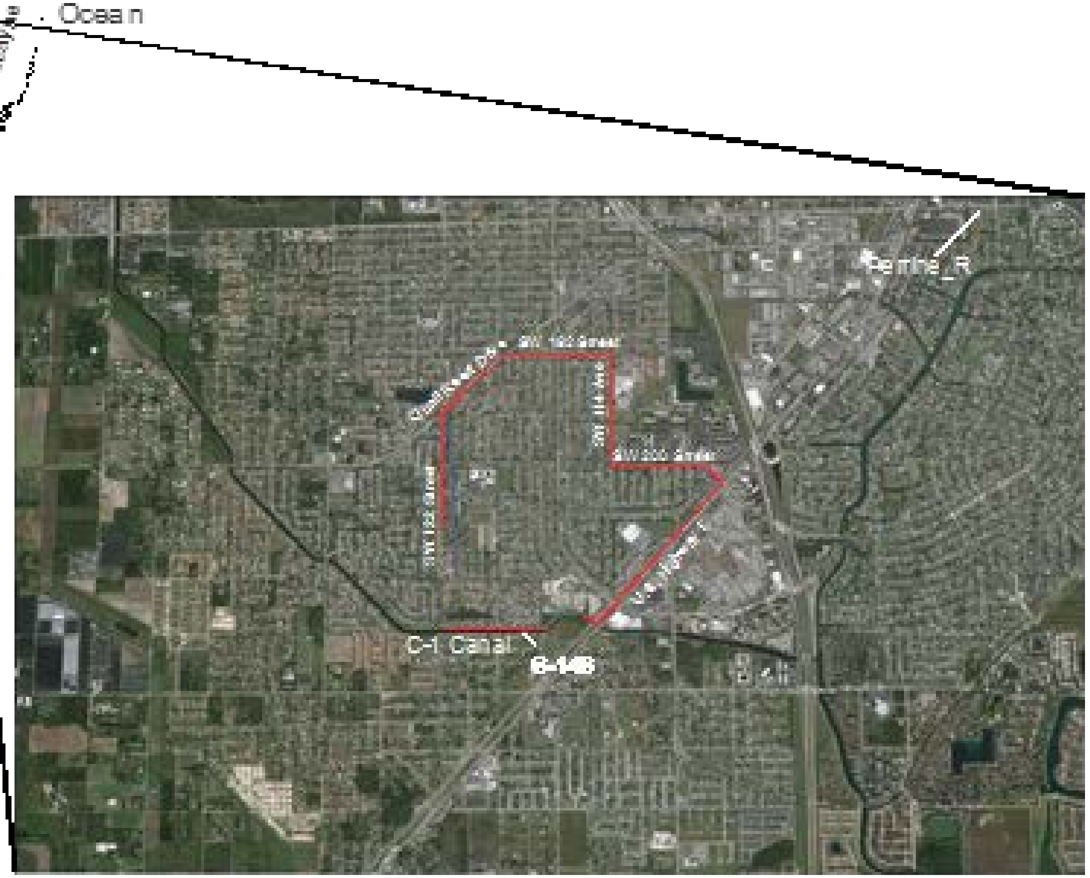
- Less effects on the last three years by varying water demands

So which is better?

- The physics-based model provides simulated data at a large number of locations (cells in the model)
- The empirical-statistical model requires much less input data to yield useful relationships and can be set up for different hydrologic scenarios with far less effort
- The physics-based model provides a more realistic concept of how the hydrologic system works, even if the results are not as precise
- The empirical-statistical model provides direct insight into the most relevant input parameters

Hybrid application of physics-based and empirical-statistical models





Neural network model used to define drainage characteristics in small area which are used in larger-scale physics-based numerical model

QUESTIONS?