Kiel University
Christian-Albrechts-Universität zu Kiel

CRSLR Research Group

UCF Coastal
UNIVERSITY OF CENTRAL FLORIDA

Wave downscaling using machine learning

Sara Santamaria-Aguilar & Thomas Wahl

Goal & Motivation

Kiel University
Christian-Albrechts-Universität zu Kiel

CRSLR Research Group

UCF Coastal
UNIVERSITY OF CENTRAL FLORIDA

Problem

- Ensemble nearshore wave projections are required to assess uncertainties in **future wave climate**
- Ensemble wave projections are available offshore
- Wave downscaling from offshore to nearshore using numerical models requires high computational capacity -> **Wave propagation involves non-linear processes**

Solution?

- **Can machine learning models be an efficient tool for downscaling wave projections?**
  - Condition: A representative set of nearshore and offshore wave data is needed in order to train the model

Methods

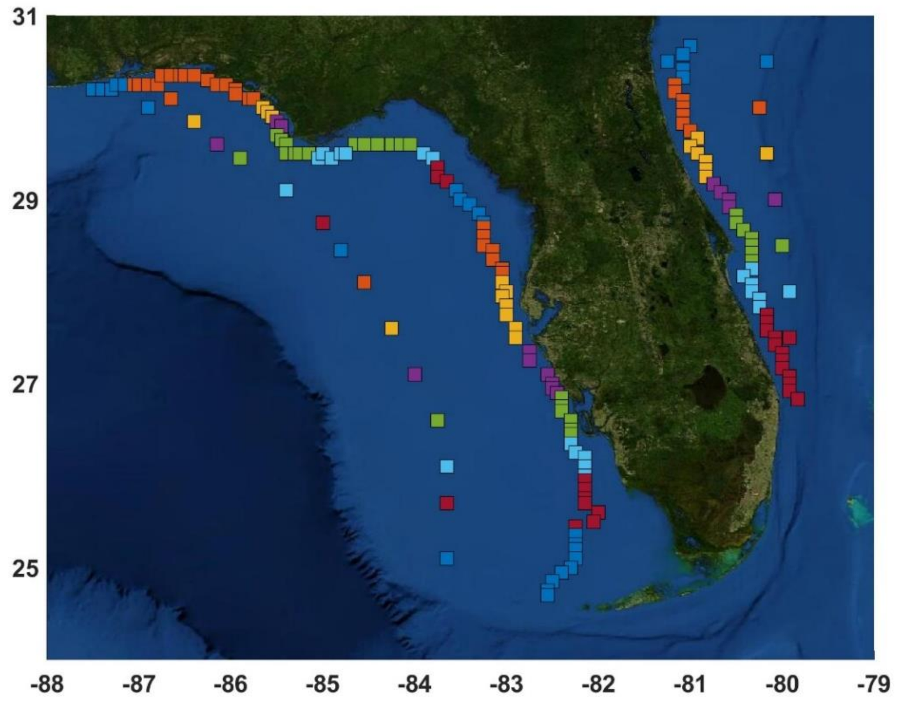
- We test the **performance** of 4 models on representing the links between offshore & nearshore waves:
  - **Multi Linear Regression** (MLR)
  - **Random Forest** (RF)
  - **Multivariate Adaptive Regression Splines** (MARS)
  - **Artificial Neural Networks** (ANN)

Machine Learning Models

- **Inputs & Outputs:** Offshore & Nearshore wave parameters

$$\mathbf{X}(Hs, Tm, Tp, Dir)_{OFFSHORE} \longrightarrow \mathbf{Y}(Hs, Tm, Tp, Dir)_{NEARSHORE}$$

- **Data**: Wave Information Studies Hindcast (WIS) of US Army Corps of Engineers
  - Hourly Sea States from 1980 to 2014



- **Input & Output stations**: Correlation between offshore and nearshore Stations

- **Performance**: 10-Fold cross validation
  - RMSE, R^2 (bias, scatter index,…)

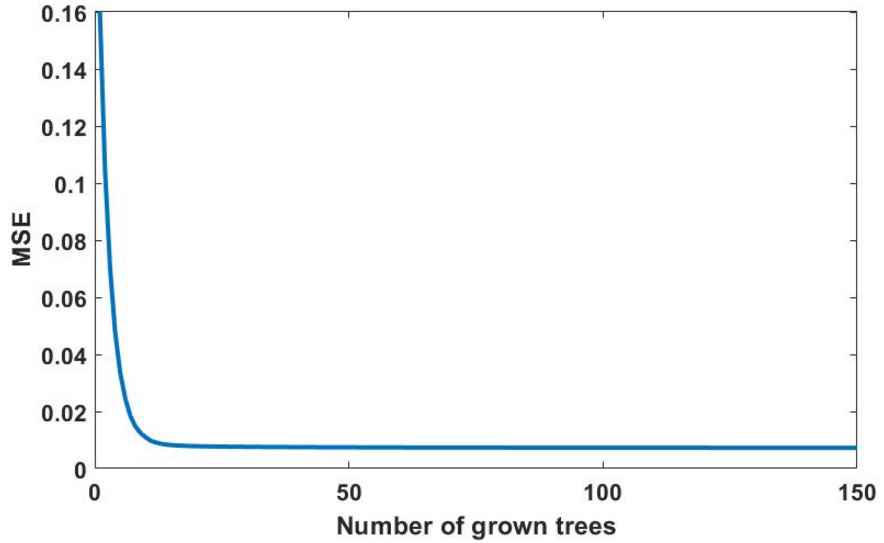
Machine Learning Models

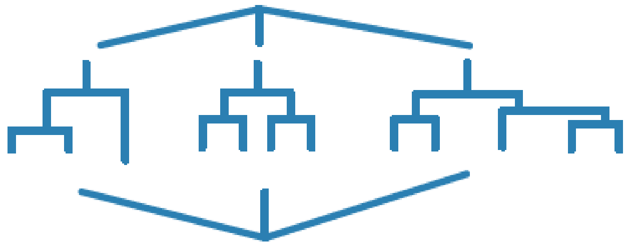
- **Multi Linear Regression**

$$\boxed{\mathbf{Y}(Hs,Tm,Tp,Dir)_{NEARSHORE} = \beta \cdot \mathbf{X}(Hs,Tm,Tp,Dir)_{OFFSHORE}}$$

  - Pros: Easy implementation and interpretation
  - Cons: Non-linear wave propagation processes

- **Random Forest**
  - A number of decision trees (*bagged*) are trained independently on bootstrapped data from the input dataset.
  - Pros: Fast algorithm, easy implementation, able to capture non-linearities and provides quantiles of the response variable
  - Cons: Difficult interpretation

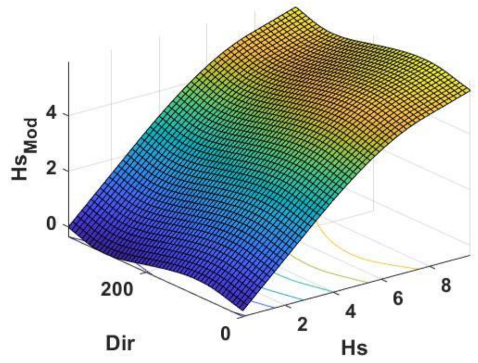




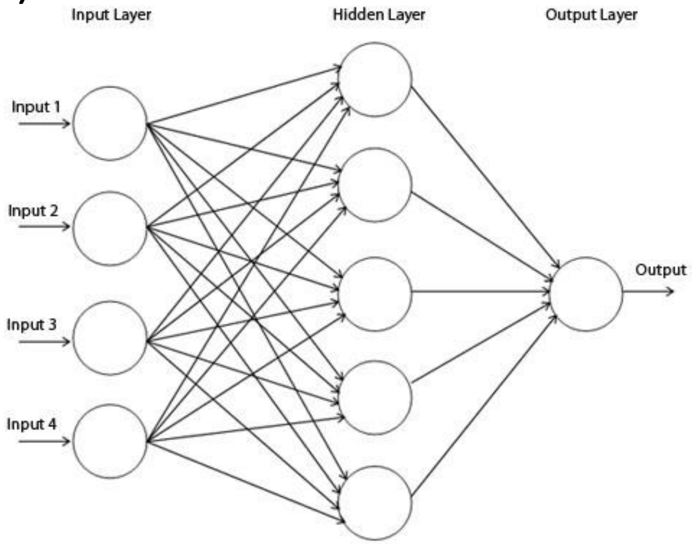
Machine Learning Models

Kiel University
Christian-Albrechts-Universität zu Kiel

- **Multivariate Adaptive Regression Splines**
  - The algorithm automatically selects the cutpoints (*Knots*) of the predictors for fitting cubic regressions where the smallest error is achieved.
  - Pros: Automatically captures non-linear relationships and easy interpretation
  - Cons: Computational expensive compared to MLR and RF

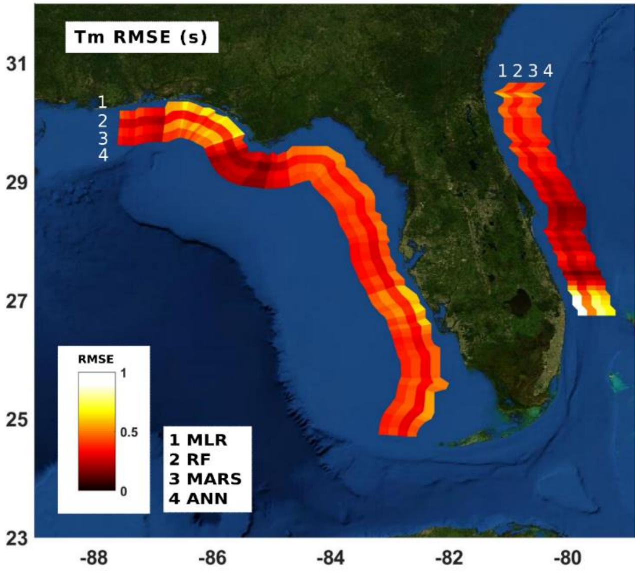


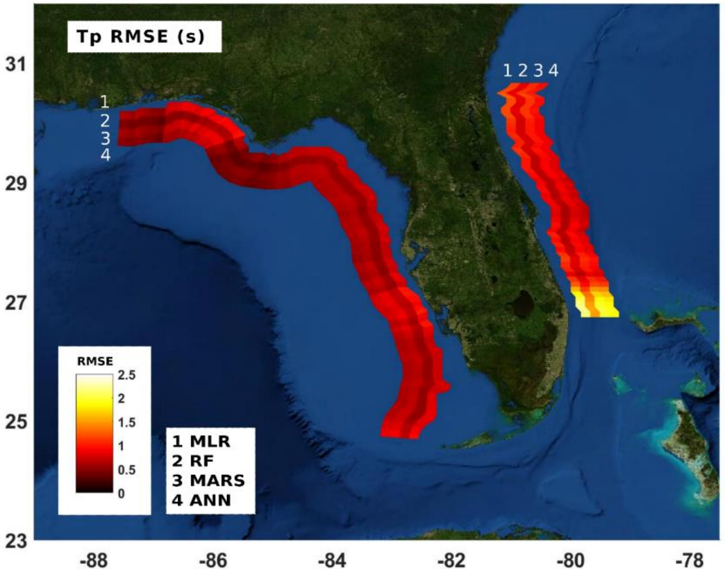
- **Artificial Neural Network**
  - Connected networks of neurons that are iteratively trained (by modifying the weights of the connections) to relate the inputs (*predictors*) to the output (*response*)
  - Network architecture:
    - 1 Hidden Layer with 10 neurons
    - Transfer function hidden layer: tan sigmoid
    - Transfer function output layer: purelin
  - Pros: Automatically captures non-linear relationships
  - Cons: Not computational efficient compared to MLR and RF, network architecture has to be defined in order to obtain good performance

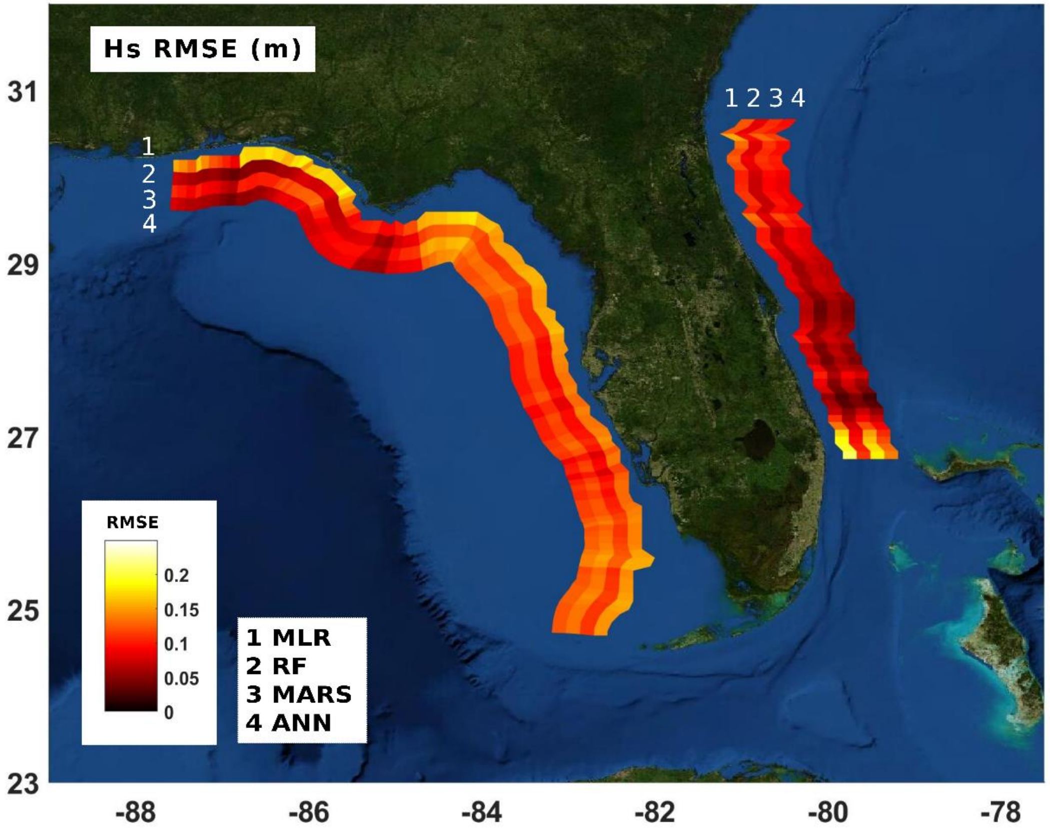


Models' Performance

- **RF outperforms** the other models, it is easy to implement and computational efficient
  - Hs is simulated with average error of 11% along the entire coast of Florida and 6% in the extremes
  - Similarly, Tp and Tm are simulated with errors between 5% to 6%

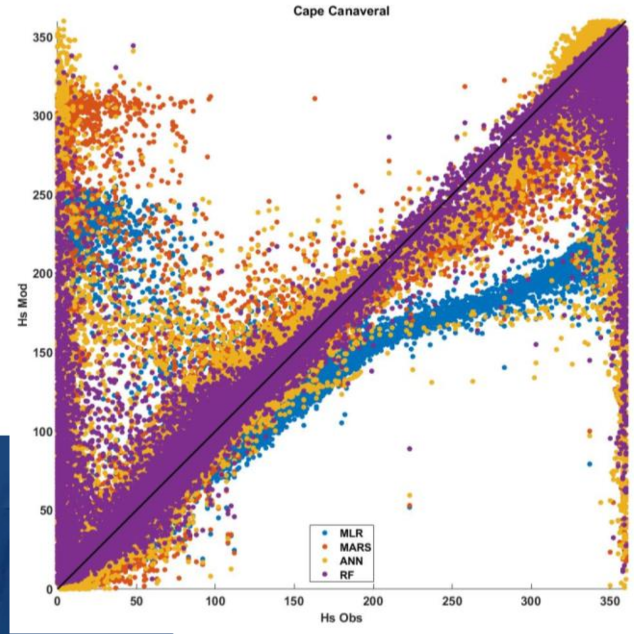




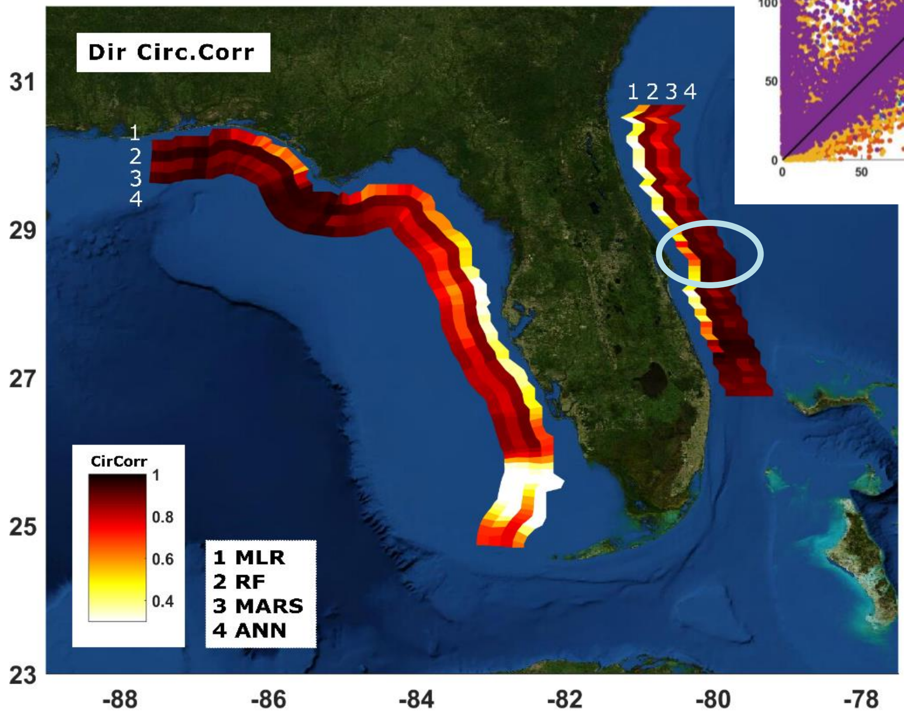


Models’ Performance Dir

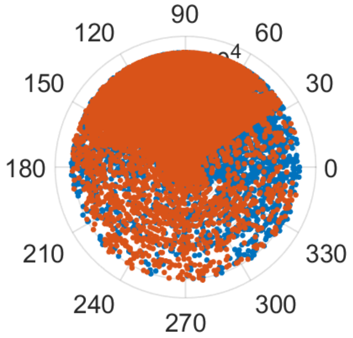
- Poor performance modelling the Dir by all models
- Models are not able to capture the behavior of directions within the North sector



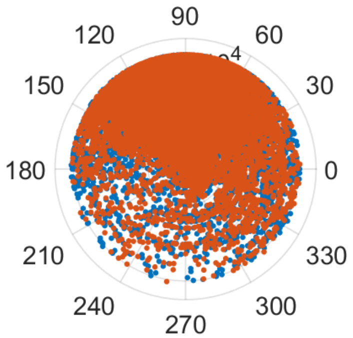
- <u>**Dir is a Circular Variable**</u>
- **Transformation of Dir** into 2 variables: **Sine & cosine** improves model performance



No Transformation



Transformation sin & cos



Take-Home Message

Kiel University
Christian-Albrechts-Universität zu Kiel

CRSLR Research Group

UCF Coastal
UNIVERSITY OF CENTRAL FLORIDA

Machine learning models are an efficient tool for downscaling wave projections, which are still omitted in the majority of coastal flood assessments

- **RF outperforms the other models and requires lower computational time**
- **Circular variables such as the Dir require a transformation into two variables in order to accurately model the North sector**