# Statement of Purpose

*Xurxo Rigueira Diaz* — *Civil, Environmental and Sustainable Engineering*
*xrigueira.github.io* — *December 8, 2022*

My research goal is in the field of statistics and computer science for environmental applications. Through my research work, I would like to leverage and develop mathematical and computational methods to help us understand how natural systems behave and predict how they evolve. To this end, I hope to pursue a Ph.D. in Civil, Environmental and Sustainable Engineering at Arizona State University.

## 1 Research experience

I have worked as a research assistant in the Department of Natural Resources and Environmental Engineering at the University of Vigo in Spain. During this time, I participated in several projects in functional data analysis and machine learning applied to environmental modeling. This opportunity has helped me gain knowledge in these fields and become acquainted with the fundamentals to do quality research.

**Univariate functional data analysis**. The goal of this project was to introduce functional data analysis as a suitable approach for the study of environmental information. Functional data is obtained through the careful approximation of discrete data points into functions with the help of basis functions. The resulting functions are studied in a Hilbert space through various statistical measures. In this work, we proposed the use of directional outlyingness, which considers the shape and the magnitude of each function compared to the whole set, allowing for the detection of outliers. In this project, I was responsible for suggesting the use of state-of-the-art functional methods, developing an error-free implementation, validating the methods proposed, making the algorithm available online, and writing and submitting the research paper. This effort resulted in my first publication and provided me with a great theoretical understanding of functional methods as well as their practical implications. Moreover, I learned how to correctly organize and present new findings in a research paper and collaborate with other researchers.

**Outlier detection**. This research aimed to reduce the number of errors in the detection of outliers in functional data. The current detectors proved to be faulty in our applications, creating a real need for a new method. Starting from the values of shape and magnitude of each function, the proposed approach analyzes their distribution to determine if the presence of outliers in the data is true. If this is the case, the algorithm engages two different advanced methods for outlier detection -Isolation Forest and Minimum Covariance Determinant- to create a single output. As for my role in this project, I formulated the architecture of the new outlier detector, developed its implementation, and validated the results obtained with a Monte Carlo simulation. This was an instructive experience that provided me with meaningful insights into the design of outliers detectors and how to solve complex problems efficiently.

**Multivariate functional data analysis**. The last project was on escalating the previous functional methods to a multivariate plane. The motivation for this project stemmed from the intrinsic multivariate characteristics of natural processes. In short, our approach analyzes the shape of the functional data set through an estimator of its quantile cross-spectral density. This, combined with the weighted average of the magnitude of each time frame, enables the analysis of a theoretically infinite number of variables and the detection of outliers. In this case, I was tasked with implementing the method for shape analysis, developing the algorithm for the magnitude analysis, and designing and validating a new outlier detector for this methodology.

The results of the last two projects were presented at international conferences, WWEM and MME&HB, respectively. This excellent opportunity helped me improve my scientific communication skills and strengthen my ideas through the exchange of opinions with experienced researchers. Besides research, teaching has also been key in motivating me to pursue a Ph.D. Nowadays, I lead two outreach talks our school offers to high schools, in which I give a lecture on smart materials or artificial intelligence. This experience, combined with my role as an after-school tutor, has helped me learn how to communicate complex concepts and structure lessons to make them compelling for a broad range of students at different academic levels.

## 2 Research interests

My main interest is machine learning for environmental modeling. In particular, I would like to apply and develop the techniques in this field to leverage the continuous flow of environmental data and help decipher how the natural world behaves.

Machine learning has the remarkable capacity to extract valuable information from this type of data. In order to improve the current methods, I intend to implement deep learning models and optimization techniques to assess challenges such as high dimensionality, internal variability, and high-order correlation among variables. This would help define statistically causal links in the data and identify which variables hold a higher weight in predicting how the system evolves, therefore reducing uncertainty when physical equations cannot provide precise results.

Additionally, I will strive to make models more reliable through the extraction of clean subsets in the training phase when dealing with noisy data and by analyzing the patterns of the missing values and their possible relationship with the target variable in prediction tasks. Moreover, one of the most notable flaws of machine learning-powered environmental models is their low adaptability to other scenarios. I intend to implement transfer learning to tackle this issue. This implies that the model is to be trained with massive publicly available databases, while the last layer of the neural network would be trained specifically for each scenario, reducing the training time and maintaining or increasing accuracy.

Lastly, machine learning systems can also work with images. I aspire to improve the resolution of current climate ensembles through statistical downscaling. In order to do this, techniques such as neighborhood analysis can be used to generate reliable information on those areas with scarce or unreliable data.

## 3 Conclusion

Given my research interests as well as potential changes in direction to keep in mind, Arizona State University would be a perfect fit to pursue my doctoral studies. More specifically, I hope to work with **Assistant Professor Tianfang Xu** as my research interests align well with her area of expertise. Reading her research paper entitled "Data-driven methods to improve baseflow prediction of a regional groundwater model" has helped me learn how these methods can provide information about precipitation uncertainty and, consequently, more robust prediction intervals. Doubtlessly, her profound knowledge in the field of numerical simulation, uncertainty quantification, and machine learning in hydrology could help me advance the research I intend to conduct.

After finishing my Ph.D., I would like to stay in academia and continue my path to eventually have a professor role. I believe I can contribute positively to your university and the whole academic community, as my goals are to perform state-of-the-art research and continue to share the knowledge I acquire with younger generations of students. For those reasons, I am confident that Arizona State University is the perfect place to pursue my Ph.D.