ADAPTIVE CONFORMAL SETS UNDER DISTRIBUTION SHIFT: USING ENSEM-

BLE DISAGREEMENT AS AN EPISTEMIC NORMALIZER FOR EEL-

GRASS SEGMENTATION UNDER TEMPORAL DRIFT

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Statistics

by

Dillon Murphy

June 2025

COMMITTEE MEMBERSHIP

TITLE: Adaptive Conformal Sets Under Distribution Shift: Using Ensemble Disagreement as an Epistemic Normalizer for Eelgrass Segmentation under Temporal Drift

AUTHOR: Dillon Murphy

DATE SUBMITTED: June 2025

COMMITTEE CHAIR: Kelly Bodwin, Ph.D.
Associate Professor of Statistics and Data Science

COMMITTEE MEMBER: Jonathon Ventura, Ph.D.
Associate Professor of Computer Science and Software Engineering

COMMITTEE MEMBER: Andrew Fricker, Ph.D.
Associate Professor of Geography

ABSTRACT

Adaptive Conformal Sets Under Distribution Shift: Using Ensemble Disagreement as an Epistemic Normalizer for Eelgrass Segmentation under Temporal Drift

Dillon Murphy

Semantic segmentation of eelgrass (Zostera marina) from high-resolution drone-based imagery is essential for coastal habitat monitoring, restoration, and management, as these habitats see rapid changes due to climate change and human-induced influences. However, the reliability of generalizing these classification models not only relies on high-accuracy segmentation but also on rigorous uncertainty quantification that holds up when conditions change across years or locations. Conformal prediction (CP) converts a classifier's output into prediction sets with finite-sample marginal coverage; however, the standard score $s = 1 - p_y$ can under-cover in hard or out-of-distribution (OOD) regions under temporal drift. Inspired by heteroskedastic conformal regression, this study proposes normalizing conformal scores by an epistemic proxy, ensemble disagreement. These scores re-scale nonconformity by a data-driven estimation of local difficulty learned on the calibration set, so that the global quantile is conservative where OOD risk is high. This study evaluates (i) a parametric form $s' = (1 - p_y)/(1 + \lambda V)$, where V is ensemble variance of probabilities, and (ii) a nonparametric normalization $s' = s/\hat{\sigma}(V)$, where $\hat{\sigma}(V) = \mathrm{E}[s|V]$ is estimated through quantile binning and interpolation. On drone imagery of the Morro Bay estuary, California (2018-2022), the models were trained on 2018-2021, calibrated on 2021, and evaluated on 2022 on OOD points. Normalized scores recover much of the lost coverage seen by vanilla split conformal prediction and approximate target, increased the percentage of singletons, and reduced the spatial coverage variability. The study provides a simple-to-implement drop-in framework to regain coverage in difficult regions and out-of-distribution data without retraining or labels at test time.

Keywords: conformal prediction, adaptive sets, uncertainty, eelgrass, remote sensing.

ACKNOWLEDGMENTS

This page is not required, but if you have received funding for your research or assistance or guidance that you feel should be noted, it belongs on this page.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Chapter 2

BACKGROUND AND MOTIVATION

2.1   Uncertainty in Deep Learning

2.2   Deep Ensembles

2.3   Semantic Segmentation in Environmental Monitoring

2.4   Conformal Prediction Foundations

2.5   Adaptive or Shift-Aware CP

Chapter 3

METHODOLOGY

3.1 Data

3.2 Model Training + Pipeline

3.3 Uncertainty Analysis

3.4 Conformalizing

3.4.1 Split CP for Classification

3.4.2 Variance-Aware Score Normalization

a) Parametric linear shrink

b) Nonparametric normalization

3.5 CP Evaluation

Set Composition, Spatial Equality, and Conditional Coverage

Chapter 4

RESULTS

4.1 Ensemble Results

4.2 In-Distribution Evaluation

4.3 Temporal OOD (2022)

global coverage and set composition

4.4 Class Conditional Coverage

4.5 Spatial Robustness within 2022

4.6 Sensitivity

Chapter 5

DISCUSSION

Chapter 6

LIMITATIONS

Chapter 7

CONCLUSION

Chapter 8

APPENDICES