

Hi everyone, I'm Justin Ting, and the topic I'm going to be talking about today is Large Scale Probabilistic and Multi-Output Benthic Habitat Mapping.

## Motivation ~1.5m

The reason we want to perform benthic, or deep ocean habitat mapping, is that the our oceans cover more than 70% of the Earth's surface, influence our climate, and life on Earth in general - to be able to manage them properly, we need to be able to understand their state and how it is changing over time.

Benthic habitat mapping has evolved in the past few decades to allow easy collection of large amounts of data, compared to the past where sediment samples would have to be collected manually by divers, and experts would draw crude maps based on sparse information and their existing knowledge.

Nowadays, we can collect bathymetry data spanning large areas on board ships, such as the depth information in the top right, and use underwater autonomous vehicles, or AUVs, to collect image data in smaller areas, such as these here, which we can group into class labels using clustering methods. By modeling a relationship between the data where both sources exist, we can predict the labels where we only have bathymetry data, by applying this relationship in those areas. We want to be able to perform these predictions during expeditions themselves, so we not only need to generate maps, but do so efficiently while utilising all the data available.

## Related Works ~1.5m

One of the most common methods used in benthic habitat mapping are random forests, known for being robust in classification problems in machine learning in general. They work by using multiple decision trees that each operate on a random subset of the full data where each one provides a class prediction, and the most common result or mode is taken as the final prediction for the overall random forest. While it achieves a moderately high accuracy in studies its used in, it is a deterministic method that deals in absolutes.

But, the purpose of mapping is to be able to potentially take actions for management and conservation purposes that may inherently carry serious risks. To properly assess such risks, we need to quantify the uncertainty of predictions, which deterministic methods are unable to do.

Use of probabilistic methods in the area is relatively uncommon, but a study in 2012 showed that considering the probabilistic labels from clustering the image data rather than the most likely ones resulted in a drop in both mean error and variance, where the variance allowed the uncertainty of every prediction to be

considered. However, a major restriction present was that only a few thousand points could be used for training to make the experiments feasible to run, due to Gaussian process models requiring the inversion of the covariance matrix that scales cubically in the number of data points.

## **Approach - GP Approx (~1.5m)**

This is where our work begins, as we try to remove this constraint that dictates use of only subsets of data. One way to do this is to use ensemble approximation methods - the ones we considered were the product of Gaussian process experts, and generalised product of Gaussian process experts. These function by using a collection of separate Gaussian processes that each model unique subsets of the dataset that are then multiplied back together and weighted by their variance as an approximation to the parameters of the Gaussian over all the data. Let's say we had 20-thousand points and 100 separate GPs - instead of having to invert a single 20-thousand by 20-thousand matrix, we would only need to invert 100 separate 200 by 200 ones - a difference in complexity of four orders of magnitude. The difference between the two is that the standard product of experts weights experts by their variance directly, while the generalised version divides by the number of experts, preventing the cancellation of variances that results in overconfidence of some predictions.

## **Approach - Intro to multi-output methods ~0.5m**

However, these GP approximations use a single label during model fitting, a simplification that doesn't represent the original data. As we can see, even if point originally contained 5 counts of habitat 1 and 4 counts of habitat 2, to simplify it to be compatible with single-output machine learning algorithms, we take the most commonly occurring label and say that the area is only habitat 1.

## **Approach - Dirichlet Multinomial Regression ~1m**

To deal with the rich distribution of labels per point, we propose the use of Dirichlet multinomial regression. As every point corresponds to a category count over all labels, each can be represented by a multinomial distribution, just as you would a weighted n-sided die. But to determine the parameters of the multinomial at each point, we draw from a Dirichlet that can provide a distribution over the space of all possible multinomial parameters. To demonstrate the behaviour of a Dirichlet's parameters in 3 dimensional space - values must sum to 1, and larger

values will result in a most concentrated distribution of values. The activation function that we used to link the Dirichlet's parameters with the data was the softmax - a more numerically stable, normalised version of the exponential. This means that the underlying behaviour of Dirichlet multinomial regression actually contains similarities to linear regression.

## **Results - GP vs DM vs Deterministic Methods (~0.5m)**

Something to keep in mind here is that for the Gaussian process and Dirichlet multinomial, to be able to compare them their rich outputs to the deterministic methods, the most probable and most frequent labels per point were taken respectively. We can see that the Gaussian process performed the best across both sets of labels in both accuracy and f-scores, while SVM performed the worst in most cases.

## **Results - Time ~0.25m**

Because we want predictions to be able to be generated during expeditions, we also took time into account - between standard GPs, GP ensemble approximations, and Dirichlet multinomial regression, there is at least an order of magnitude.

## **Results - Deterministic Maps (~0.25m)**

Although the cross validation scores for some of the deterministic methods were quite similar, their full predictive maps don't agree with each other, but we also don't have a way to decide which one to trust more for certain areas - hence the need for probabilistic predictions.

## **Discussion - Gaussian process variance (~0.25m)**

Whereas for the Gaussian process maps, we have the associated standard deviations at each point, showing us large areas that had a especially low or high variance, giving us a measure of how much to trust predictions in those areas.

## Evaluation - GP vs GPoGPE (~1m)

- GPoGPE matches DM results more
- GP needs to learn a single set of hyperparameters for the training data
- GPoGPE is more flexible, many GPs learning different HPs for different sections of data (corresponding to different regions)
  - reflected in results - though GPoGPE was on paper worse with cross-val scores, it was much closer to the DM maps

## Discussion - Dirichlet Multinomial Biodiversity (~1m)

- able to see more visual richness even in the argmax graph taking MAP approximation
- DM can be used to determine biodiverse areas
- e.g. this blank area here with no biodiversity if you remember the bathymetry depth map, is very deep and is just all sand
- this higher point though, is a large raised area rich with different types of coral

## Discussion - Dirichlet Multinomial using parameters from full distribution (~0.5m)

- MAP only takes max approx from posterior distribution
- we can take draws of weights using MCMC and see if the maps are mostly in agreeance

## Discussion - Entropy (~0.5m)

- Because the weights used to calculate the alpha parameters of the Dirichlet were calculated using MAP, we ignored the possible set of weights in the posterior distribution
- Sampling the weights using MCMC gives us an idea of which areas are more certain than others, i.e. change less
- We can also use the Dirichlet's entropy to determine these less certain areas
- Quantify certainty of predictions - sure vs unsure

## **Contribution/Conclusion**

- able to speed up probabilistic methods while using approximations that allow more flexibility in model
- model multi-output data to allow collection of information such as biodiversity

## **Future work**