# Compact Models for Adaptive Sampling in Marine Robotics

Journal Title
XX(X):1–15

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

**Trygve Olav Fossum[1,7], John Ryan[2], Tapan Mukerji[3], Jo Eidsvik[4], Thom Maughan[2], Martin Ludvigsen[1,5] and Kanna Rajan[6,7,8]**

**Abstract**

Finding high-value locations for *in-situ* data collection is of substantial importance in ocean science, where diverse bio-physical processes interact to create dynamically evolving phenomena. These cover a variable spatial extent, and are sparse and difficult to predict. Autonomous robotic platforms can sustain themselves in harsh conditions with persistent presence, but require deployment at the right place and time. To that end, we consider the use of remote sensing data for building compact models that can improve skill in predicting sub-mesoscale features and inform onboard sampling. The model enables prediction of regional patterns based on sparse *in-situ* data, a capability that is essential in regions where use of satellite remote sensing in real time is often limited by cloud cover. Our model is based on classification of sea-surface temperature (SST) images, but the technique is general across any remotely sensed parameter. Images having similar magnitude and spatial patterns are grouped into a compact set of conditional means representing the dominant states. The classification is *unsupervised* and uses a combination of dictionary learning and hierarchical clustering. The method is demonstrated using SST images from Monterey Bay, California. The consistency of the classification result is verified and compared with oceanographic forcing using historical wind measurements. The established model is then shown to work in a real application using measurements from an autonomous surface vehicle (ASV), together with forecast and sampling strategies. Finally an analysis of the model prediction error is presented and compared across different paths and survey duration.

**Keywords**

Machine Learning, Sampling, Ocean Modeling, Marine Robotics

## 1 Introduction

Effective and informative sampling of the ocean requires data gathering strategies that can resolve the spatial and temporal variations of phenomena. This is a formidable challenge due to the dynamic and unstructured nature of the ocean, with spatio-temporal scales spanning many orders of magnitude, making it unrealistic to observe the dynamics in detail. Additionally, coastal waters are often heterogeneous in nature due to interactions between bathymetry, river discharge, oceanic circulation, as well as endogenous processes (e.g. biology). Methodologically this drives a requirement for using compact spatial models to inform more effective sampling strategies, capable of running on robotic platforms, utilizing prior and current *in-situ* observations. There are numerous ways to build spatial models. The essential goal is to exploit the underlying spatial correlation structures and try to reconstruct the environment, so that future sensing locations can be determined accordingly.

Earth observing satellites offer the possibility to observe a large spatial extent for a range of ocean parameters, such as sea-surface temperature (SST), sea-surface height (SSH), salinity, and ocean color. From such data, a number of ocean processes can be discerned and characterized, including algal blooms, fronts, eddies, internal waves, and numerous water quality parameters (Johannessen et al. 2000). Consequently, there is enormous potential for using such information to perform automated analysis of the spatial patterns in the ocean. However, a major challenge using modern machine learning techniques is the reliance on labeled data sets (Gonalves et al. 2008). In remote sensing applications, this challenge is further exacerbated in having to work with a limited number of training samples (Mountrakis et al. 2011), especially when concentrating on a specific area of scientific interest. It is therefore valuable to use unsupervised methods that provide the ability to learn the inherent structure of the data without using explicitly provided labels.

[1]Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.
[2]Monterey Bay Aquarium Research Institute, California, US.
[3]Stanford University, Department of Energy Resources Engineering and Department of Geophysics by courtesy, California, US.
[4]Department of Mathematical Sciences, NTNU, Trondheim, Norway.
[5]University Centre in Svalbard (UNIS), Longyearbyen, Norway.
[6]Department of Engineering Cybernetics, NTNU, Trondheim, Norway.
[7]Centre for Autonomous Marine Operations and Systems (AMOS), Trondheim, Norway.
[8]Underwater Systems and Technology Laboratory, Faculty of Engineering, University of Porto, Portugal.

**Corresponding author:**
Trygve Olav Fossum, Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway.
Email: trygve.o.fossum@ntnu.no

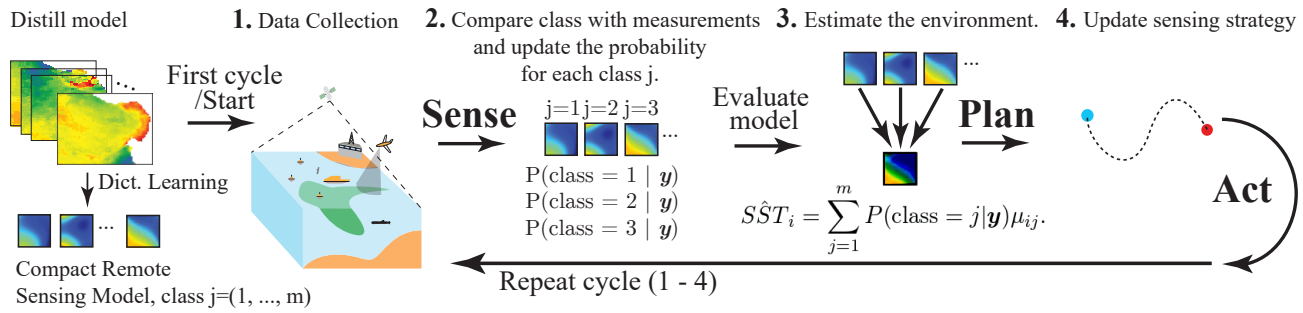



**Figure 1.** The main concept for integration of the compact model into a robotic sensing framework, following the *sense-plan-act* paradigm; details are explained in Section 6.

Given the challenge with *undersampling* in oceanography (Munk 2002) and the limited availability of accurate real time information, it is essential to call for capabilities that can estimate spatial and temporal variations in the ocean environment on the fly. Building compact spatial models using prior data sources such as remote sensing and ocean model output is therefore one such possibility. We present a method for building such a model using remotely sensed SST data, as well as examples of how this model can be used in a robotic sampling framework. Remote sensing images are chosen specifically due to their synoptic properties providing repeated large-scale surface observations with reasonably high resolution. The model aims at predicting the current state of the environment using a superposition of states, referred to as *classes* or *scenarios*, where each state is established from a similar set of SST images that represent recurring states of oceanographic conditions. The environment is therefore, assumed to be equal to such a superposition and is found by evaluating the likelihood of these states against observations. A conceptual view of this approach is presented in Fig. 1, showing how the model can inform sampling in a *sense-plan-act* structure.

It is important that this type of data reduction (unsupervised clustering) cover the common dominant spatial patterns seen in the images, i.e. in the form of several distinct classes of *conditional means*, that is comparable to the variation in the underlying environment. Using a combination of dictionary learning (Aharon et al. 2006), sparse coding (Mairal et al. 2009), and (agglomerative) hierarchical clustering (Everitt et al. 2011) we propose an unsupervised classifier that can group images with similar oceanographic characteristics, using spatial patterns and the magnitude of temperature from SST images. The idea is not only to automate this process, but also to provide new insight into the underlying processes themselves. Subsequently, the classified SST images are used to distill a compact model of the dominant features by computing conditional means within these classes. One potential drawback to this approach is the lack of uniformity in data acquisition due to cloud cover and lack of satellite coverage. These factors limit the ability to do repeated and systematic observations of a region of interest, which in the worst case can impede the construction of such spatial models. As a mitigation, data from a three year period in spring is used (see Fig. 2), such that we obtain a relatively continuous data coverage for a season. Combining this approach with other synoptic sources of data such as numerical ocean models can also be investigated and is discussed further in Section 7.

The applicability of such compact models is an effective characterization of key ocean states, especially when SST data from satellites is not available. In this work we show an example of this, determining regional conditions for a given day, using data gathered from a WaveGlider ASV (shown in the corner of Fig. 3b). We also examine the model prediction error, by computing the expected mis-classification rates for different sampling design strategies.

The structure of this paper is as follows: in Section 2 we introduce related work, along with some of the challenges that underpin the motivation for this paper. Section 3 presents some background on both the data sources and methods used. Section 4 presents the proposed classification methodology, with the results shown in Section 5. Section 6 presents model implementation and usage towards adaptive sampling. We conclude in Section 7 with a summary, followed by a discussion of future work.

## 2 Related Work

Much of the work on automated analysis of remote sensing data is focused on thematic mapping for terrestrial applications, with numerous applications such as in agriculture (Mulla 2013), mapping of urban environments (Saritha and Kumar 2017) and vegetation (Xie et al. 2008), and is usually geared towards change detection (Walter 2004). Typically, the general objective is to categorize regions of an image into one of various land cover classes or themes. Dictionary based classification has been explored for terrestrial hyper-spectral thematic mapping in Chen et al. (2011). Similar approaches such as sparse reconstruction-based classification have been applied to high resolution images of the sea floor, taken with synthetic aperture sonar McKay et al. (2016). Despite an abundance of approaches, Wilkinson (2005) demonstrates through evaluation of fifteen-years of remote sensing research related to classification, that no technique displays any significant advantage over another, showing no trend in improvement of the classification results, including results from artificial neural networks. Limited contexts, increasing complexity of data sets, lack of embodiment in best practices, the difference between low-level features and high-level user requirements, and imperfect human processes (such as labeling and ground truthing) are some explanations proposed for this lack of progress.

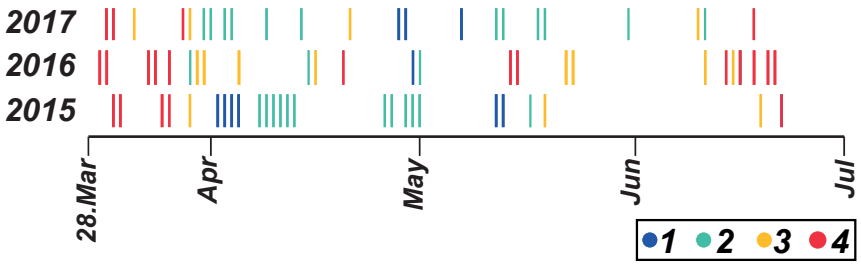



**Figure 2.** The dates of usable SST images from Monterey Bay for the period March to July 2015, 2016, and 2017. Each image is marked as a vertical line, colored according to the initial classification results (class 1, 2, 3, and 4) in Section 5.1.

Compared to terrestrial or seafloor applications, some additional challenges exist in the upper ocean domain, namely: i) the non-static boundaries and surface features present in the environment (Blondeau-Patissier et al. 2014), ii) the lower signal-to-noise ratio for signals arising from water masses, and iii) the presence of surface effects such as sun, and sky glint (Emberton et al. 2016). Studying ocean surface features for analysis of surface slicks, currents, fronts, waves, and wind interactions are discussed in Ryan et al. (2010) and Chen (2012), with an emphasis on synthetic aperture radar images. High frequency (HF) radar is also combined with remote sensing in Das et al. (2010), where the sampling and hotspot prediction of harmful algal blooms is examined. This is also studied in Bernstein et al. (2013), where a shore-based recognition pipeline is suggested, building on remote sensing data for event detection, feature localization, and trajectory prediction. Frolov et al. (2013) analyzes the spatial and temporal decorrelation scales seen in marine algal blooms using fluorescence line height imagery, to strategize monitoring of such episodic events. Frolov et al. (2012) investigates short-term prediction of surface currents using HF-radar observations to develop a linear autoregression model, where the climatology of conditional mean flow-fields for upwelling, downwelling, and relaxation in Monterey Bay is presented. Optical water types were found using fuzzy clustering analysis on spectral information in Eleveld et al. (2017), aimed towards identifying different types of lakes, e.g. clear versus turbid waters. At large spatial scales, Oliver and Irwin (2008) uses remote sensing data to automatically resolve different oceanic regions with certain spatiotemporal characteristics, to monitor the effect of El Niño events.

A number of approaches have been explored combining onboard models and remote sensing to guide sampling in the ocean. Smith et al. (2010) utilizes forecasts from a high resolution ocean model combined with remote sensing to pre-plan missions with multiple autonomous underwater vehicles (AUVs). Areas with high concentration of Chlorophyll *a* (a proxy for phytoplankton abundance) are identified from the satellite imagery and simulated forward in time using an ocean model. Presenting only simulated results, the paper shows the potential and also the challenges of leveraging prior data. Issues related to small scale discrepancy between model simulation (used for planning) and the actual conditions, aligns with the assumptions in our work, and the fact that we are focusing on predicting and planning based on large-scale (regional) features. Chao et al. (2017), provides a discussion and preliminary results in closing the loop between numerical ocean models, robotic platform sampling, and data assimilation. Multiple information streams are proposed to update and improve sampling strategies without human intervention. However, the robotic assets depend on human involvement and robust communications using shore-based assimilation and planning methods. Consequently, this drives the need towards elevated levels of autonomy and onboard sampling strategies for situational awareness, such as presented in our work.

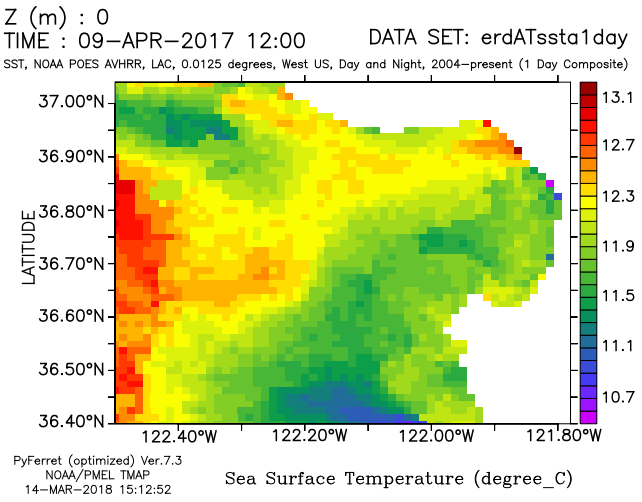
This paper describes a novel way of building compact ocean models using remote sensing products for use in robotic sampling. As our work uses pre-processed data (1 day average SST), some of the raw information (e.g. quality flags) is forsaken for practical purposes of obtaining and working with the data. The proposed method provides the ability to work directly with both temperature and spatial patterns, as this information is carried along throughout the analysis using a combined feature vector. Contrary to other work, the classification results are confirmed and verified by an independent marine data source (preceding wind history), allowing oceanographic processes to be tied to the subsequent analysis and use of the data. Wind drives horizontal and vertical circulations, which in turn determines magnitudes, gradients, and spatial patterns in SST (Rosenfeld et al. 1994). The method is unsupervised and can be used with a very small number of images ($\leq 100$).

## 3 Preliminaries

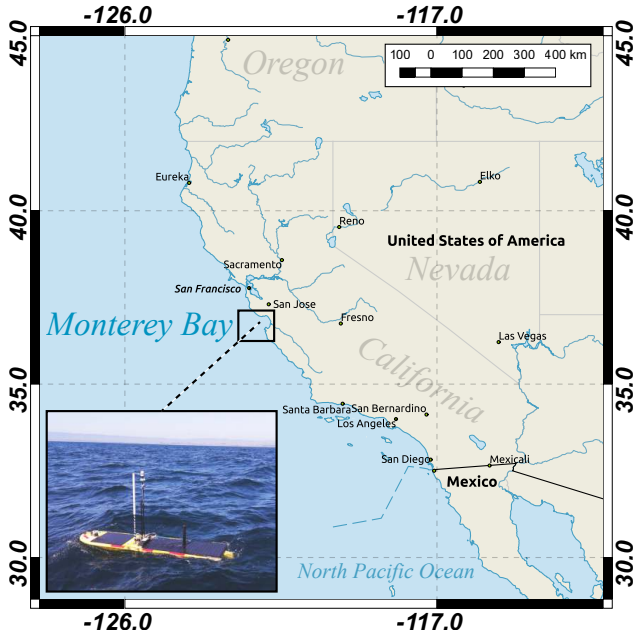
### 3.1 Satellite Data Sources

The SST remote sensing images used as a basis for this model are provided from a high resolution radiometer onboard the Polar-orbiting Operational Environmental spacecraft (POEs) NOAA-17 and NOAA-18. The data are processed and mapped to an equal angle grid (0.0125 degrees latitude by 0.0125 degrees longitude) using a simple arithmetic mean, producing both individual and composite images from 1 to 14 days duration. This may provide some averaging artifacts, see Fig. 15a. The nominal accuracy is about 0.7 degrees Celsius (C), covering the west coast of North America.*

*The data are publicly available through the National Oceanic and Atmospheric Administration (NOAA) NWS Monterey Regional Forecast Office and the CoastWatch program, from their ERDDAP server https://bit.ly/2ngyP6c.




**(a)** Example of a SST image that was used to make the compact model.




**(b)** Map of California and Monterey Bay with the SST region marked.

**Figure 3.** 3a A SST image from NOAA NWS Monterey Regional Forecast Office showing Monterey Bay (between 36.65 and 37°N). 3b Map of California and Monterey Bay with the SST region marked. The WaveGlider ASV is shown in the inset. *Image courtesy: Chris Wahl, MBARI.*

We used one-day average SST images, similar to Fig. 3a, from 2015, 2016, and 2017 in this work covering Monterey Bay, California (marked in Fig. 3b), for a 100 day period from March until the start of August, yielding 300 images in total. This period of the year was chosen in large part because wind-driven coastal upwelling and associated thermal signatures are strongest. Due to cloud cover, only about 25% of the images had a quality that was useful for our application, giving us a total of 74 images to work with. Local cloud and fog cover can limit the data availability; henceforth longer time periods should be considered for inclusion in the data sources if images are influenced by these factors. Equally, too much exposure can create data gaps that introduce bias into the final model, as images and their spatial patterns are left out of the analysis. Cloud and fog cover is further a motivating factor for actually deploying autonomous vehicles that can aid in estimating ocean conditions. Although the SST images in this analysis are limited to spring and summer, coastal upwelling circulation in this region continues into the fall season and is closely linked to coastal land features (Rosenfeld et al. 1994). Therefore, the interpretation of structural information by these methods should be similarly applicable during fall. However, the magnitude of SST shifts during fall due to seasonal warming throughout the region. This aspect of the environment motivates a seasonally dependent method, using images that are consistent seasonally.

The time stamp of images in Fig. 2, shows availability over sequential days, and therefore the likelihood of having a similar mean temperature. However, SST patterns can be vastly different in sequential days because energetic currents change water mass distributions rapidly, motivating the case for looking at spatial similarities.

### 3.2 Unsupervised learning and classification of SST images

In order to quantify different spatial patterns in SST images, we use an unsupervised classification method based on sparse image representation. The compressed representations are obtained from employing dictionary learning techniques, before hierarchical clustering (Everitt et al. 2011) is performed in several steps to classify the images. The approach aims at classifying images with the same SST pattern into dominant/archetypical classes, having distinct oceanographic significance. Clustering the raw image data (pixels) is not effective, due to their high dimensionality. Using bulk characteristics, such as the mean, min, and max temperature, is also possible. This approach can get good results, but leads to unnecessary smoothing/blur (in the conditional mean) as classes are combined without spatial information. To demonstrate this, a comparison of the classification variability is presented in Section 4.2. Consequently, clustering a compressed/sparse representation is more viable, as both temperature magnitude and spatial information can be combined together. Moreover, the dimensionality can be kept low, making it easier to cluster and hence identify and differentiate between the characteristics of each image.

### 3.3 Dictionary Learning

Sparse dictionary learning (Aharon et al. 2006; Mairal et al. 2009) is a method that can be used to build sparse representations of data. The resulting format is a set of coefficients, collected as a *sparse code*. Dictionary learning is similar to *Principal Component Analysis* (PCA) (Jolliffe 2011) (also known as *Empirical Orthogonal Functions*) in that the coefficients form a linear combination of certain basic elements, referred here as *atoms*. The atoms can be considered as instances of "characteristic patterns" that can be combined to reconstruct the input. The combined matrix of atoms is called a *dictionary*, usually denoted as $\boldsymbol{D}$, while the sparse codes are usually noted as $\boldsymbol{\alpha}$, having the relationship

$$\boldsymbol{x}_k = \boldsymbol{D}\boldsymbol{\alpha}_k, \tag{1}$$

where $\boldsymbol{x}_k$ is the raw image $k$, with $\boldsymbol{D}$ as the dictionary, and $\boldsymbol{\alpha}_k$ as the unique coding for that specific image. We use `MiniBatchDictionaryLearning` or `MBDL` from (Pedregosa et al. 2011) to find the dictionary using the least angle regression method (LARS) Mairal et al. (2009), while the sparse codes are found using an orthogonal matching pursuit (OMP) algorithm (Tropp and Gilbert 2007) that greedily selects the dictionary atoms sequentially, through computation of the inner products between the image and dictionary columns. The loss function used in this optimization can be formulated as

$$l(\boldsymbol{x}, \boldsymbol{D}) = \min_{\boldsymbol{\alpha}} \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \qquad (2)$$

where $\lambda$ is a regularization parameter (set to 1 in our implementation), $\|\cdot\|_2^2$ denotes the squared Euclidean norm, and $\|\cdot\|_1$ denotes the $\ell_1$ norm. This loss function, also known as *basis pursuit* (Chen et al. 2001), encodes two essential optimization criteria. The first term seeks to minimize the reconstruction error (depending on $\boldsymbol{D}$ and $\boldsymbol{\alpha}$), while the second seeks to attain a sparse solution for $\boldsymbol{\alpha}$. As we also have to find $\boldsymbol{D}$, to solve for the function above, this is usually rewritten as a joint optimization problem with respect to the dictionary $\boldsymbol{D}$ and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_n]$ as

$$\min_{\boldsymbol{D},\boldsymbol{\alpha}} \frac{1}{n}\sum_{k=1}^{n}(\frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{D}\boldsymbol{\alpha}_k\|_2^2 + \lambda\|\boldsymbol{\alpha}_k\|_1) \qquad (3)$$

where $n$ is the total number of images. To obtain the best dictionary $\boldsymbol{D}$ and sparse code $\boldsymbol{\alpha}$ the optimization alternates between optimizing one parameter while keeping one parameter fixed (see Mairal et al. (2009) for details). The images are decomposed into patches with a specific size, and fed into the algorithm in batches, to improve convergence performance. The number of patches is non-random and determined by the patch size, following a left to right, top to bottom extraction pattern. For the specific SST images used here, the patches also include a land/cloud mask. The size of these patches will influence the quality of the result and needs to be selected based on the small- and large-scale similarities in the input images; sensitivity analysis should therefore be used to identify this parameter (see Section 4.1). In the implementation used here, finding the dictionary $\boldsymbol{D}$ using the `MBDL` library involves an iterative process that uses a random state to initiate the model, hence reproducible results require using the same pre-defined random seed.

## 4 Methods

### 4.1 Proposed Classification Methodology

In separating and collecting dominant spatial patterns from SST images an important aspect that we emphasize in this work is that images sharing a common oceanographic evolution need to be identified. This implies finding images that share a mutual historic progression of, for example, wind and currents, that contribute to shaping a particular environmental condition. Involvement of local oceanographic expertise and knowledge is therefore essential in finding a separation scheme that is justifiable. This is especially true since the images are snapshots of a continuous process and separation into classes will imply some form of discretization.

The classification method builds off of the idea of classifying images based on a sparse representation (sparse code), instead of their high dimensional pixel space. The classification is also conducted in two steps, illustrated in Fig. 4 as a branching graph with a new dictionary and sparse coding generated after each step (i.e. "initial classification" and "secondary classification").

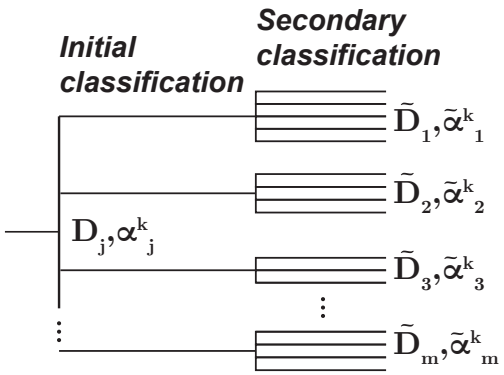



**Figure 4.** Based on the dictionary $\boldsymbol{D}$ a sparse code $\boldsymbol{\alpha}$ is found. These codes are used to classify images into distinct classes. After an initial classification this step is repeated, with a new dictionary and codes generated from the images within a class; branching into smaller sub-classes with higher levels of similarity.

This two-step classification increases the performance of the method, as the larger variability (magnitude) is handled in the first classification, followed by a more fine scale separation (spatial patterns) in the second. The detailed steps involved can be described as follows, for image $k$ and each class $j = 1, \ldots, m$:

1. Normalize all images with the global mean and standard deviation.
2. Extract a fixed number of patches from the SST image (e.g. 260 patches of size $40 \times 40$ pixels), and stack these in a 1 dimensional (1D) vector.
3. The 1D vectors are collected in the data matrix and fed into the `MBDL` subroutine to retrieve the initial dictionary $\boldsymbol{D}_j$.
4. Each SST image is then compressed using the dictionary $\boldsymbol{D}_j$ to form a unique sparse coding for the image $\boldsymbol{\alpha}_j^k$.
5. These first $\boldsymbol{\alpha}_j^k$ coefficients are then classified using hierarchical clustering.
6. Steps 2 & 3 are then repeated for the images within the derived classes to retrieve a new sub-class specific dictionary $\tilde{\boldsymbol{D}}_j$.
7. Step 4 is repeated using the sub-class specific dictionary $\tilde{\boldsymbol{D}}_j$ to obtain the final sparse codes $\tilde{\boldsymbol{\alpha}}_j^k$.
8. The final sparse coding $\tilde{\boldsymbol{\alpha}}_j^k$ is now further classified into sub-groups with a higher level of similarity using another step of hierarchical clustering. This is the secondary classification as shown in Fig. 4. This final classification is subject to a criteria where the mean number of images across the final classes should be at

| Parameter | Value | Comment |
|---|---|---|
| Input data set | 74 images, size: $52 \times 59$ pixels | Type: 1 km, 1 day average SST |
| Patch size $[x_{pa}]$ | $40 \times 40$ | The patch pixel size |
| Sample count | 260 | The number of sample patches from each image |
| Dict. size $[x_{ds}]$ (initial) | 4 | Number of atoms in the initial dictionary |
| Dict. size $[x_{ds}]$ (secondary) | 8 | Number of atoms in the secondary dictionary |
| Distance metric | Ward | The distance metric used by the hierarchical clustering |

**Table 1.** Overview of raw data, as well as parameters used in dictionary learning and hierarchical clustering.

least 3.0 (see below); if not the images are flagged as being too distinct.

It is important to ensure that the final classification result is not too distinct, e.g. having only one image in a class, resulting in loss of effectiveness towards building a compact model. It is therefore important to consider the *separation distance* (SD) in hierarchical clustering as a measure of similarity of classified images. A practical aspect of using hierarchical clustering is that compared to other algorithms such as *k-means* (MacQueen 1967), where the number of clusters must be pre-specified, hierarchical clustering can use the dendrogram to generate the clustering structure. Consider therefore the dendrogram in Fig. 5 used for finding SD in the initial step. Lowering the SD yields more classes with fewer images and vice-versa. The distance metric used for clustering is the Ward sum-of-squares minimization metric (Ward Jr 1963), chosen since it is more permissive of cluster shape/size assumptions (Anderberg 1973), and which performed better on the sparse codes. In using this distance measure it is important that the data vectors we are operating on are normalized by the mean and standard deviation (Wilks 2011), hence *Step 1* above. The SD therefore, for the initial clustering is found based on the dendrogram and evaluation of the separation results. As noted in Wilks (2011), setting the SD usually requires a subjective choice that depends on the goal of the classification. There exist various statistical machine learning tools that consider the bias versus variance trade-offs (e.g. James et al. (2013)), but there is no single "best" approach. Since this initial grouping is followed by a secondary classification, this choice is not as decisive a factor, as the secondary clustering, as only the major features are to be identified. Nevertheless, local oceanographic expertise and knowledge should be involved to find a justifiable separation scheme. For the first step of the classification, a general rule of thumb would be to avoid a SD that yields too sparse a set of classes.

For the secondary clustering step, finding a SD that does not result in an imbalance in variance (i.e. too big/small classes), becomes important. The goal here is to find similar images that can be used in a conditional mean, while also excluding images that are too distinct. Similarity is, as noted, controlled by the SD. To achieve an adequate final result we iteratively change the SD until we reach a classification result where the mean number of images across the sub-classes is above or equal to a certain threshold, set to 3.0. This threshold is chosen such that the generated classes must span several images and avoid creating classes with dissimilar magnitude and spatial structure. For example, if a secondary classification with 12 images starts with a SD$=500$ and yields the sub-classes $[1, 2, 3, 4, 5]$ with associated image counts of $[5, 4, 1, 1, 1]$, the mean number of images in each class equals 2.4, and three classes have only one image. This does not satisfy our criterion, and the SD can therefore be increased, relaxing the measure of how similar the images need to be. The SD is then relaxed (increased) until one of the classes containing only one image is combined into one of the other five, as this reaches our criterion.

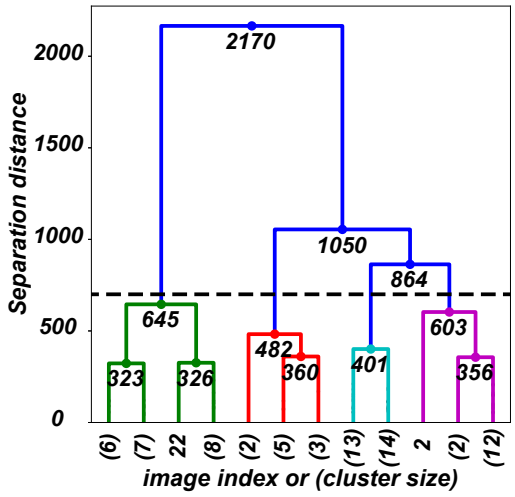
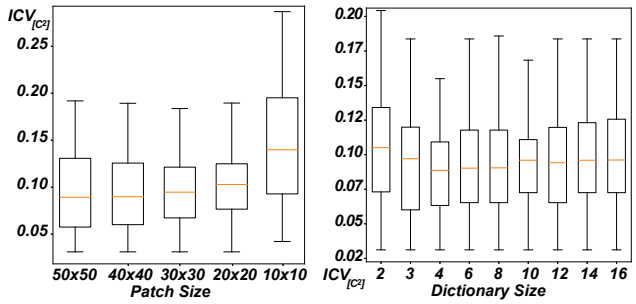



**Figure 5.** The dendrogram for the initial clustering step. Four classes are being separated using a distance measure set to 700. The numbers on the x-axis are either image index or cluster size (in parentheses).

Parameters that are associated with the above procedure, are shown in Table 1. The patch and dictionary size can be adjusted to yield different results. The relationship between the patch size $x_{pa}$, dictionary size $x_{ds}$, and code length $x_{cl}$ is given, using the fixed dimensions $52 \times 59$ for the images, as $x_{cl} = ((52 - x_{pa} + 1) \times (59 - x_{pa} + 1)) \times x_{ds}$ and therefore, the code length depends on the square of the patch size, while increasing linearly with dictionary size. This code length is important as clustering becomes more complex and may lead to degraded classification accuracy for higher dimensional vectors, as can be seen in Fig. 6a. Parameter selection was done by comparing the intra-class variability (ICV) for each class $j = 1, \ldots, m$ measuring how different the images within a class are by comparing the temperature variability at each location across the images, given as:

$$\begin{aligned} \mathrm{ICV}_j &= \sum_i \sigma_{ij}^2, \\ \sigma_{ij}^2 &= \frac{1}{n_j} \sum_{k \in \otimes_j} (x_{ik} - \mu_{ij})^2 \\ \mu_{ij} &= \frac{1}{n_j} \sum_{k \in \otimes_j} x_{ik}, \end{aligned} \tag{4}$$

where $x_{ik}$ is the temperature at location $i$ in image $k$, $\otimes_j$ is the set of images belonging to class $j$, and $n_j$ is the number of images in this set. Moreover, $\mu_{ij}$ and $\sigma_{ij}^2$ are the sample mean and variance in temperature at location $i$ for class $j$. Using ICV as a metric, the patch and dictionary size is chosen based on the settings that give the lowest ICV after several iterations for each setting with a non-constant random seed. The patch size sensitivity (Fig. 6a), shows a drop in ICV with increasing patch size, that gradually flattens as the patch size increases. The variability is slightly increasing for patch sizes above $30 \times 30$, without any gain in accuracy. Based on this analysis we use a patch size of $40 \times 40$ pixels. Fig. 6b shows that a dictionary size of $4$ is sufficient to achieve both low ICV spread and value. For secondary classification, the dictionary size is increased to $8$ as the number of images being classified are fewer and their similarity is higher (filtered by the initial classification), hence a more descriptive code can be applied.




**(a)** ICV sensitivity to patch size. **(b)** ICV sensitivity to dictionary size.

**Figure 6.** Intra-Class Variability (ICV) sensitivity using different patch size $x_{pa}$ and dictionary size $x_{ds}$. Fig. 6a shows a box plot of the ICV spread for different patch sizes; larger $x_{pa}$ tend to yield lower ICV values. Fig. 6b shows the ICV variation for different dictionary sizes. A dictionary size of 4 is a good choice, with low ICV spread and value.

### 4.2 Comparative Analysis

As the proposed method provides enhancement of classification by including structural information, the performance of the methodology is compared by using i) hierarchical clustering of temperature metrics only (*minimum, maximum, min-max range,* and *mean* temperature [i.e. no sparse codes]), and ii) directly classifying the raw images using a *k-means* approach. To deal with an unknown number of clusters, we use *X-means* (Pelleg and Moore 2000), while for the hierarchical approach we use the dendrogram. As the temperature magnitudes from the SST can contain several strong separating factors some initial categorization can be achieved; however, neglecting spatial information leads to unnecessary variability. By comparing the ICV given in Eq. (4) between the proposed, temperature-only, and X-means approach derived classes, the effect of using spatial structure can be demonstrated. This is shown in Fig. 7 where a histogram presents the ICV for the derived classes, where each unit block is a class with a corresponding ICV.

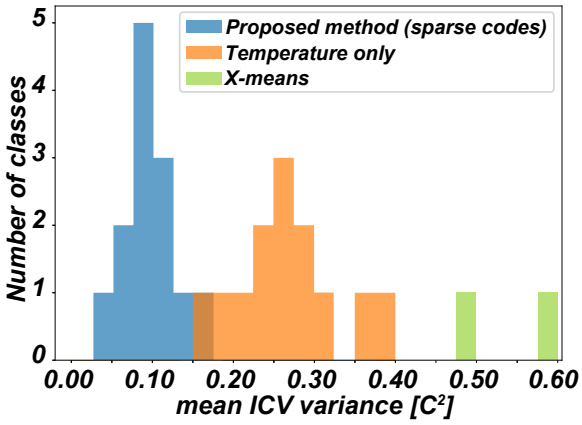



**Figure 7.** Histogram of the ICV for the derived classes using the proposed-, temperature only-, and X-means approaches.

The figure shows that classifying using only temperature metrics or raw data leads to higher variance. As the sparse codes are able to compress information about temperature and spatial patterns into one vector (with lower dimension), the resulting evaluation in the downstream classification task is simplified; this, in turn, leads to a more accurate classification with lower variance. Not surprisingly, X-means clustering has the worst performance, as the method only splits the data into two classes (cold and warm). Classifying the data further using X-means results in no further separation.

## 5 Classification Results

The final classification groups are used to make different conditional means that together constitute a compact model. An important aspect of the approach is to verify that images which are classified together share a common formation history, i.e. having been influenced by the same sequence of physical processes. Having a shared evolution of oceanographic conditions also implies that there exists a common and distinct spatial pattern that can be clustered together to make what we choose to call a *condition*. Wind observations are effective for this purpose as it is the dominant driver of circulation, SST variability, and environmental structure in this region of study. The use of other sources of data, rather than wind, is also possible and will vary depending on the study region. The comparison and verification is shown in Fig. 8 together with the initial clustering result.

### 5.1 Initial Classification

The initial classification is dominated by the temperature magnitude, as this is a stronger separating factor than spatial patterns. The `SD` used in hierarchical clustering was found using the dendrogram in Fig. 5, producing four classes. Fig.

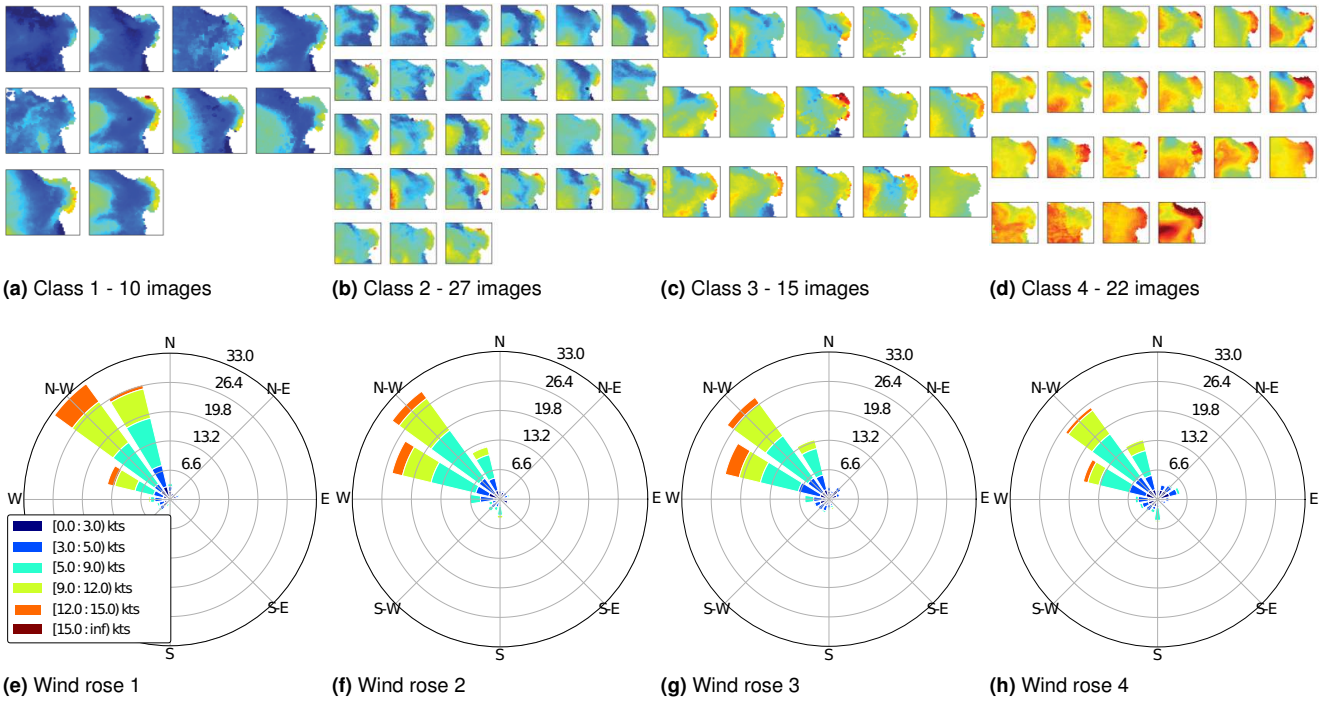



**Figure 8.** The four classes derived from initial classification results shown with the associated wind history during the 4 days preceding each image within a class. The wind rose plots represent a summary of wind speeds and the directions from which it was blowing. Wind speeds are defined by color (legend) and percentage of speeds in a given range are shown by occupancy of each color. Cold upwelled class 1 has stronger north-westerly winds, while warm relaxation class 4 has weaker magnitude and greater spread in direction. The temperature images are all using the same temperature scale from $10^{\circ}$C to $16^{\circ}$C.

8 shows these together with the aggregated 4 day preceding wind history. The images appear in arrangement going from *cold* to *warm*, or in the oceanographic context, from *upwelled* to *relaxation* dominated waters. The associated winds confirm this by showing a correspondence between the upwelled (class 1 – cold) SST images with strong north-westerly winds, and weaker more spread wind pattern for the relaxation (class 4 – warm) SST images. The intermediate classes comprise images covering the transition between these two oceanographic conditions. The initial classification has two classes of mainly cold water ([class 1 and 2] 10 + 27 = 37 images), and two classes of warmer waters ([class 3 and 4] 15 + 22 = 37 images), with the same number of images. This suggests a balance between *upwelling* and *relaxation* events.

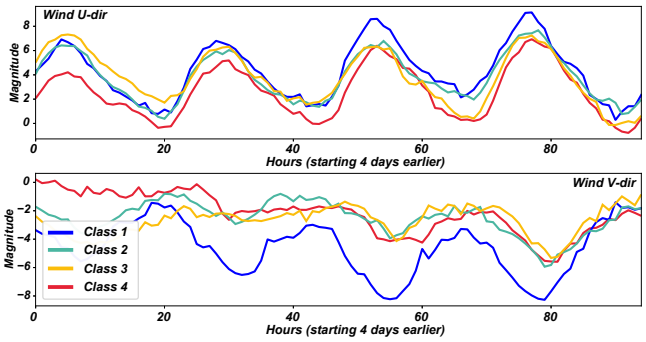


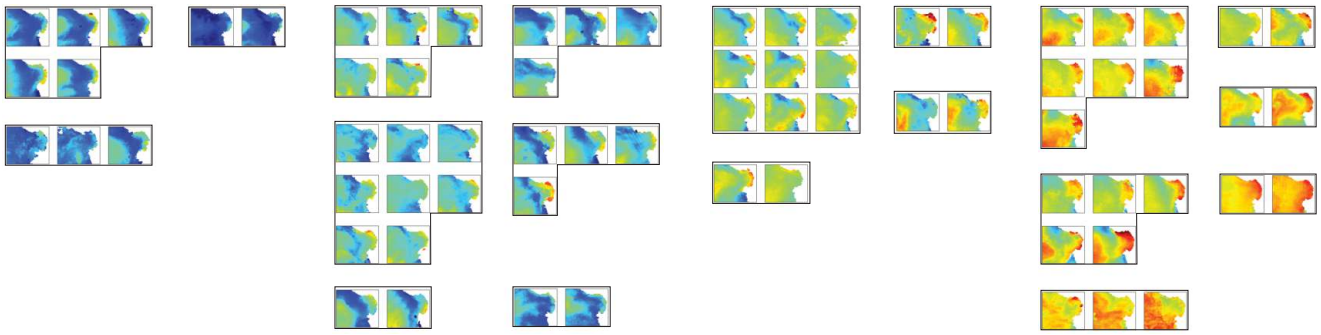

**Figure 9.** The average 4 day wind history for each class given in terms of u- and v-components also show a significant difference between the classes. Note the strong 2 day amplitude for class 1.

Fig. 9 shows the average u- and v-components[†] of the wind history for each class. Similar to Fig. 8, the average trend shows class 1 has the strongest average winds, and class 4 the weakest. A strongly negative v component of the wind is equivalent to north-westerly, i.e. blowing from the northwest and upwelling favorable. There is also a strong diurnal signal in the winds, evident in all classes in the u component. This predominantly east-west "sea-breeze" is driven by differential heating of land and ocean through day/night.

The initial classification shows that some structure follows from the temperature, but as seen in Section 4.2, using temperature alone will not achieve the best clustering result. Magnitude information helps us avoid a pitfall where images that have a similar pattern but different temperature magnitude are clustered together, disrupting the conditional means by averaging images that represent different oceanographic states.

To understand this initial classification further Fig. 11 shows the projection of the sparse codes using principal component analysis (PCA) projected into a 2D plane, visualizing the information contained in the vectors. As previously noted, as the SST images are covering a continuous process some images are overlapping and could potentially be associated with more than one class. It is also possible to identify certain images that can be deemed as outliers. The colors in Fig. 11a show where the hierarchical clustering distance, SD, is making the distinction.

[†]The u signifies the zonal velocity component of the wind, while v the meridional component.

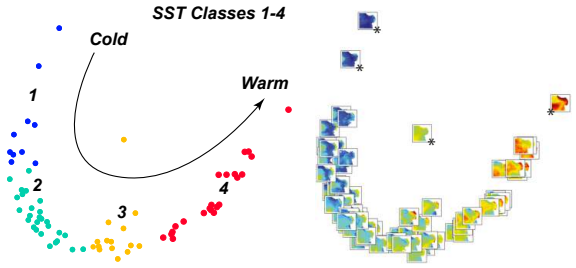


**(a)** Sub-Class 1 **(b)** Sub-Class 2 **(c)** Sub-Class 3 **(d)** Sub-Class 4

**Figure 10.** The results from the secondary classification with 19 sub-classes. Four initial classes have been further sub-divided into sub-classes based on their class $j$ dependent $\tilde{\boldsymbol{D}}_j$ and $\tilde{\boldsymbol{\alpha}}_j^k$ dictionary and code. Note the number of groups represent only valid sub-classes, as some of the images are left out from the compact model which are deemed to be distinct.

| Classification | Classes | Image Count | Mean Class Size | ICV |
|---|---|---|---|---|
| 1st | [1 2 3 4] | [22 10 27 15] | 18.5 | 0.30 |
| 2nd - Class 1 | [1 2 3] | [3 2 5] | 3.33 | 0.12 |
| 2nd - Class 2 | [1 2 3 4 5 6 7*] | [2 3 5 2 2 7 1*] | 3.14 | 0.09 |
| 2nd - Class 3 | [1 2 3 4] | [2 2 2 9] | 3.75 | 0.13 |
| 2nd - Class 4 | [1 2* 3* 4 5 6 7 8] | [2 1* 1* 4 4 2 8 5] | 3.37 | 0.13 |

**Table 2.** Classification Results. The $ICV$, given in Eq. (4), measures the mean sub-class variability. (*) marks groups with only 1 image which are considered too distinct and hence removed.




**(a)** PCA projection. **(b)** PCA projection - image scatter

**Figure 11.** The initial classification shown as projected into 2D using PCA, shown with the classified label color and class membership. It is evident that the images represent a continuous process which can be expressed as progress from cold to warm conditions. Note the outlier images (marked *).

## 5.2 Secondary Classification Refinement

The images in each class are to be further distinguished in the secondary classification. The images in this step are now already sorted according to matching temperature, which we have shown, by comparing the wind history, can be traced to distinct evolution of oceanographic conditions. Within each class, a new dictionary $\tilde{\boldsymbol{D}}_j$ can now be found, that can specialize in finding a final sparse code $\tilde{\boldsymbol{\alpha}}_j^k$, that factors in more spatial information, as shown in Fig. 7. The parameters for the secondary classification, found in Table 1, are similar to the initial classification apart from a larger dictionary size. As noted, the dictionary size has increased because the number of images being classified are fewer and their similarity is higher in this step, hence a more descriptive code can be applied.

Fig. 10 shows the resulting secondary classification, with a total of 19 sub-classes (see Table 2). The classes containing only one image are deemed too distinct for further inclusion into the compact model. As expected the mean class size is above 3.0 as specified from the criteria in *Step 8*, with $ICV$ values from $0.09 - 0.13$. It is also apparent that images are now sorted both by temperature as well as spatial features (see e.g. the upper left corner of Fig. 10c). However, cloud cover can pose challenges. In some cases, this "false" pattern match can be such a dominating feature that images with a similar false feature are classified together, (e.g. the lower group in Fig. 10a) as shown in Fig. 12.

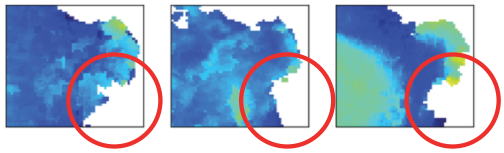
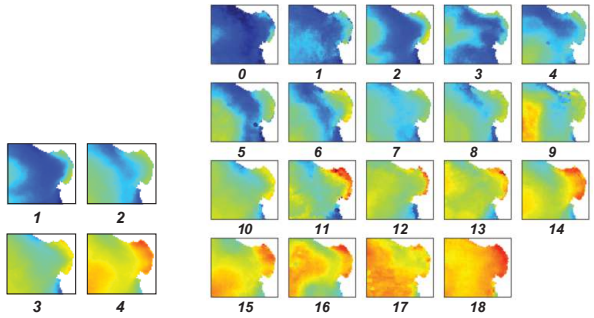


**Figure 12.** The misclassification of images due to cloud cover. The last image should be in the first group in Fig. 10a, rather than the lower group.

Taking the average across each sub-class in Fig. 10, yields the different conditional means and the final compact model, shown in Fig. 13b. Rather than relying on the means from the initial classification in Fig. 13a, these representations hold more spatial information that is advantageous when trying to predict the current state of the environment. From these conditional means one can observe not only the transition from cold to warm conditions, but also the varying spatial structure that develops within and outside Monterey Bay.




**(a)** Initial conditional means

**(b)** Secondary conditional means

**Figure 13.** The conditional means derived from the initial and secondary classification.

## 6 Forecast and Sampling Policies

We now use these conditional means as a compact model from which a prediction of the environment, specifically the SST field, can be made. Having the capability to compare *in-situ* data with the model provides a way to determine which types of historical conditions that best fit these data. On this basis an estimate of the current state can be formed, usually taken as a weighted combination of the "best" class candidates. A common method for verifying the reliability of the prediction is to compare the root-mean-square error (RMSE) of the class mean to the average class spread, as suggested in Fortin et al. (2014); the idea is that the standard deviation of the class spread should be approximately equal to the RMSE. We further study how the prediction probabilities of different states depend on the data, and how this can be used in designing sampling strategies.

### 6.1 Predicting the Environment

From Fig. 1, an intuitive way to predicting the environment is by combining one or several classes using a weighted average. During robotic deployments, as with a WaveGlider ASV, the prediction is done conditional on data $\boldsymbol{y} = (y_1, \ldots, y_{n_p})$, where $n_p$ is the number of measurements made during the survey. Based on satellite data, the probability $P(\text{class} = j)$ is estimated as the fraction of images in each class over the total number of images, $P(\text{class} = j) = n_j/n$. Conditional on data $\boldsymbol{y}$ the probability of class $j$ becomes:

$$P(\text{class} = j|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|j)P(\text{class} = j)}{p(\boldsymbol{y})} \qquad (5)$$
$$p(\boldsymbol{y}) = \sum_j p(\boldsymbol{y}|j)P(\text{class} = j).$$

The class likelihood, in Eq. (5) above, is approximated as a *multivariate normal distribution*:

$$p(\boldsymbol{y}|j) = \frac{1}{\sqrt{(2\pi)^{n_p}|\boldsymbol{\Sigma}_j|}} e^{\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_j)\right)}, \qquad (6)$$

where $\boldsymbol{\mu}_j = (\mu_{1j}, \ldots, \mu_{n_p j})$ is the class-$j$ vector of mean values along the survey trajectory, and $\boldsymbol{\Sigma}_j$ is the associated covariance matrix at these $n_p$ sampling locations. This covariance is defined by elements $\boldsymbol{\Sigma}_j(i,i') = \text{diag}(\sigma_{ij})\,\boldsymbol{R}\,\text{diag}(\sigma_{i'j})$, where $\boldsymbol{R}$ is a distance based correlation matrix (Matérn 2013), i.e. $R(i,i') = (1+\phi h_{ii'})e^{-\phi h_{ii'}}$, where $h_{ii'}$ is the Euclidean distance between sampling locations $i$ and $i'$, and $\phi$ is indicative of the correlation range. In this paper we use ∼15 km, based on Frolov et al. (2014, Fig. 4b) showing decorrelation scales for Monterey Bay.

Prediction is done using class probabilities as a function of the data gathering window. This means that we integrate one more observation at every step, and re-calculate the probability over the classes, given this growing subset of data. Because of the spatial correlation in the model, induced via $\boldsymbol{R}$, the assimilation of one more observation will not have the same effect as it would for independent data. The final prediction at location $i$, based on available data $\boldsymbol{y}$, is a weighted average according to the conditional distribution in Eq. (5),

$$\hat{SST}_i = \sum_{j=1}^{m} P(\text{class} = j|\boldsymbol{y})\mu_{ij}. \qquad (7)$$

where $\mu_{ij}$ is the conditional mean in class $j$. A brief example using this approach is presented using data from a WaveGlider ASV. This vehicle records the temperature at 0.4 m depth using a Seabird CTD (conductivity, temperature, and depth) sensor, which we will use to predict the SST.

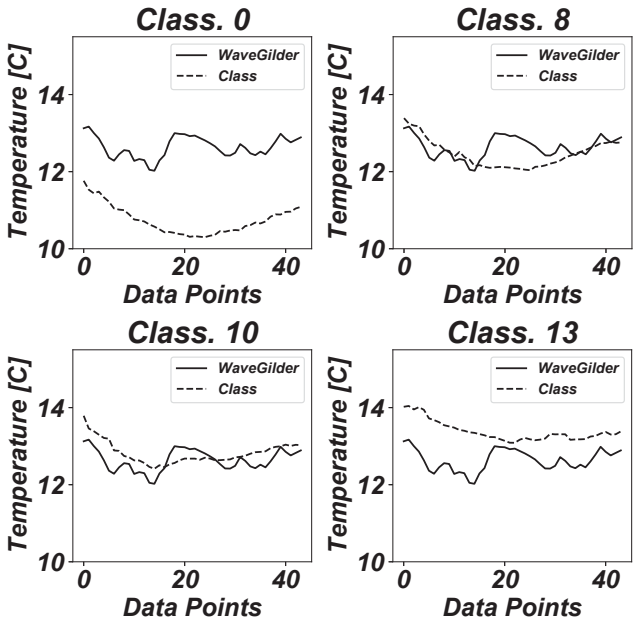
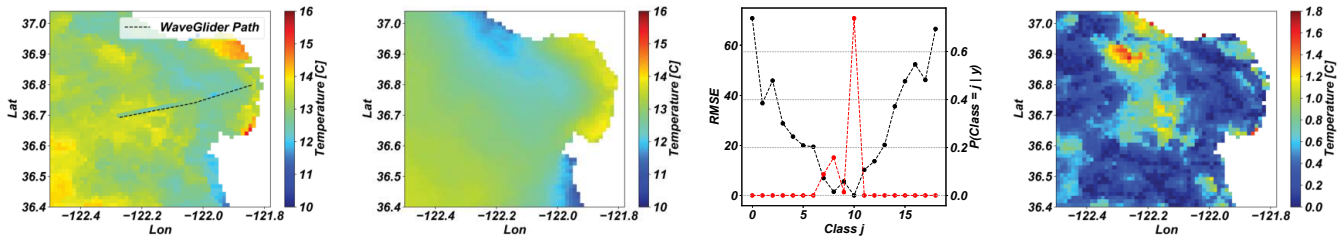



**Figure 14.** The *in-situ* temperature profile superimposed on the temperature profile for class 0, 8, 10, and 13.

We use data from the $17^{th}$ of May, 2018 in Monterey Bay, during the **CANON**[‡] field experiment. The prediction, using Eq. (7), is compared to the actual 1 day average SST for that day. Fig. 14 shows the recorded temperature profile across the survey locations (the survey path is shown in Fig. 15a) together with the corresponding profile from class 0, 8, 10, and 13. Clearly, some classes match the data better

[‡] https://bit.ly/2KHpfCH




**(a)** True SST 2018-05-17. **(b)** Estimated SST. **(c)** Likelihood and RMSE. **(d)** Estimation error.

**Figure 15.** The results from evaluating the classes by using the likelihood and the RMSE. The true SST (15a), estimated SST (15b), the likelihood (in red) and RMSE (in black) for each class (15c), and the spatial temperature difference (15d) are shown. The track of the WaveGlider is shown in 15a with the observed temperatures overlaid. There are some averaging artifacts in the true SST that show up in the estimation error (15d).

than others. Class 8 and 10 follow closely throughout, while class 0 and 13 have poor correspondence. We use this data to predict the environment using $P(\text{class} = j|\boldsymbol{y})$ as the weights. Thus, the likelihood is therefore expected to be high around class 8 and 10, and low for the others. The prediction is shown in Fig. 15b, together with the actual SST image in Fig. 15a. The accuracy of the prediction is, as expected, better on a larger scale, with some smaller spatial features that are not captured by the model. Fig. 15d shows this temperature difference spatially. Note that the the nominal accuracy of the SST images is about 0.7 degrees Celsius. As the original daily composite SST image is not without error, e.g. noise from daily averaging, some regions will show exaggerated difference. It is also reasonable to assume that there is a discrepancy between the WaveGlider data and the daily average SST, which also contributes towards estimation error. This can be seen with closer inspection of Fig. 15a, as the track and overlaid temperature are colder than the average SST. Fig. 15c shows the RMSE and together with the likelihood $P(\text{class} = j|\boldsymbol{y})$. The RMSE is calculated as:

$$RMSE(j, \boldsymbol{y}) = \sqrt{\frac{\sum_{i=1}^{n_p}(y_i - \mu_{ij})^2}{n_p}}. \tag{8}$$

There is a correspondence between the likelihood $P(\text{class} = j|\boldsymbol{y})$ and the RMSE. As expected the maximum likelihood and the lowest RMSE occur around class 10. As the likelihood adjusts for spatial correlation and prior probability $P(\text{class} = j)$, there is some difference between the two measures of similarity (e.g. class 9). Comparing the mean RMSE of the estimated SST against the class spread (0.45 vs. 0.77), indicate that we are in accordance with the reliability measure discussed in the beginning of this section.

## 6.2 Evaluation of Model Prediction Error

For discrete models, such as the one developed here or Lilleborge et al. (2016), the intent of data collection is to pull the predictive probabilities closer to 0 or 1 (Eidsvik et al. 2015). The *a priori* prediction error (before any data are recorded) is given via the most likely ocean state class as $j^* = \text{argmax}_j\{P(\text{class} = j)\}$ as:

$$PE = 1 - P(\text{class} = j^*). \tag{9}$$

Conditional on data $\boldsymbol{y}$ the prediction error is $PE(\boldsymbol{y}) = 1 - P(\text{class} = j^*(\boldsymbol{y})|\boldsymbol{y})$, where now the most likely ocean state is given as $j^*(\boldsymbol{y}) = \text{argmax}_j\{P(\text{class} = j|\boldsymbol{y})\}$, with probabilities defined in Eqn. (5).

To evaluate the model and different sampling strategies before any data $\boldsymbol{y}$ is collected, we can look at the average posterior prediction error obtained by integrating over all possible data as:

$$\begin{aligned} PE(\boldsymbol{y}) &= E\{1 - P(\text{class} = j|\boldsymbol{y})\} \\ &= \int (1 - P(\text{class} = j^*(\boldsymbol{y})|\boldsymbol{y})p(\boldsymbol{y}))dy, \end{aligned} \tag{10}$$

where $j^*(\boldsymbol{y})$ is the class with the largest probabilities conditional on $\boldsymbol{y}$. The improvement made by data collection can now be compared over various experimental designs (i.e. survey paths) using the Monte Carlo approach outlined in Eidsvik et al. (2015, Ch. 5.6). To analyze the effect of data gathering we start by generating synthetic data from the model. This entails drawing a random class $j^b$ from probabilities $P(\text{class} = j)$, $j = 1, \ldots, m$, and drawing data $\boldsymbol{y}^b$ conditional on this class based on the likelihood model in Eq. (6). We set:

$$\boldsymbol{y}^b = \boldsymbol{\mu}_{j^b} + \boldsymbol{L}_{j^b}\boldsymbol{z}. \tag{11}$$

This approach uses the Cholesky factorization (Nash 1990) of the covariance matrix $\boldsymbol{\Sigma}_j = \boldsymbol{L}_j\boldsymbol{L}_j^T$ along with a length $n_p$ vector $\boldsymbol{z}$ of independent $N(0, 1)$ variables. Examples of generated data can be seen in Fig. 16. The final error prediction can now be computed using the generated data in a Monte Carlo approximation.

$$PE(\boldsymbol{y}) \sim \frac{1}{B}\sum_{b=1}^{B}(1 - p(j^*(\boldsymbol{y}^b)|\boldsymbol{y}^b)), \tag{12}$$

where $B$ is the number of iterations (we used $B = 500$).

It is now possible for a practical comparison of the reduction of prediction error for different survey times (the effect of gathering more data) and survey locations (survey trajectories). The results are shown in Fig. 17, as a function over the survey duration, and for three different survey paths, using the WaveGlider and assuming a platform speed of approximately 2.2 kts. Using this, different survey lengths can be correlated to mission time.

Fig. 17a shows the prediction error for a given mission duration. Naturally, longer missions result in more

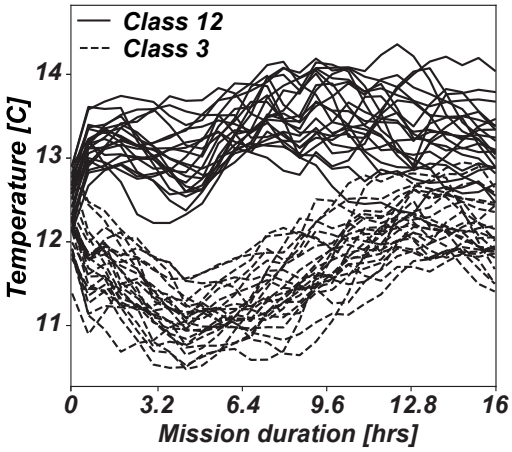
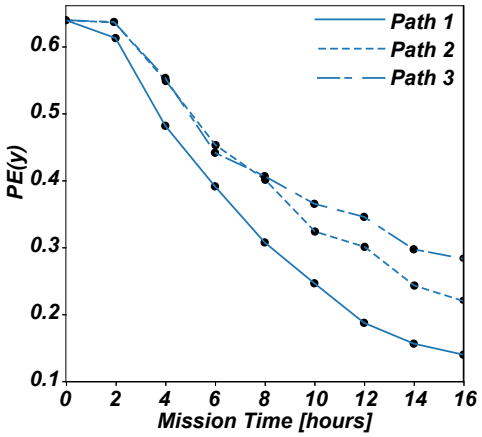
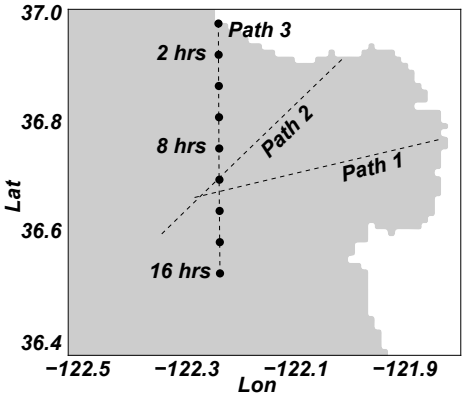



**Figure 16.** An example of 50 synthetic survey lines data for class $j = 12$ and $j = 3$, with missions lasting 16 hours.

observations and less error, reflecting the amount of data that are available. The effect of different survey paths is shown by using three different routes, shown together in Fig. 17b, with their evolution and final $PE(\boldsymbol{y})$, as shown in Fig. 17a. Path 1, which crosses both the inner and outer bay, produces the lowest prediction error ($PE(\boldsymbol{y}) \sim 0.15$).




**(a)** Predicted error vs. mission time.




**(b)** The different mission paths.

**Figure 17.** (17a) The effect of gathering more data (expressed as mission duration) on the prediction error, calculated for each of the three paths. The error drops (from close to the prior probability $PE \sim 0.64$) as more information is obtained. (17b) By comparing different survey paths, one can observe that some locations are more informative than others, showing difference in $PE$-curves.

A possible reason for this is that path 1 crosses both gradients inside the bay, as well as gradients that are prominent further offshore, covering the usual band where up welling fronts occur. The curve starts at the prior probability $PE \sim 0.64$, which arises from Eq. (9), estimated as the fraction of images in each scenario over the total number of images. This type of investigation is useful, since it provides an estimate of both the value of mission duration and location, which is often an unknown when planning survey campaigns. Such an analysis can also be conducted across multiple platforms; each platform can be evaluated by simulating a different coverage (survey speed), spatial correlation ($\phi$), measurement noise, etc. Optimization of coverage versus cost is also possible, finding effective solutions that maximize the cost per observation. Such methods of reasoning ("What is the value of the data and how much data is enough?") are often referred to as *value of information*-analysis (Eidsvik et al. 2015).

## 7 Discussion

In the absence of remote sensing, description of regional *high resolution* data may be unavailable. A compact model, as the one developed in this work, and the supporting statistical tools, can help provide contextual *low resolution* information, by using *in-situ* observations. Reducing the global uncertainty is necessary for enabling efficient planning of vehicle surveys, that rely on evaluating the conditions at unexplored locations, as well as variability and associated correlation structures. With this in mind, some aspects of the presented approach are discussed, in order to shed light on potential benefits and pitfalls.

Assimilation using an onboard numerical ocean model that accounts for time is currently not possible or practical due to time and computational limitations, hence compact or reduced order models are needed. The current compact model is static, i.e. the classes themselves are not modified during the mission. In practice, this means that small scale features will not be well resolved. One could use *Gaussian Process Regression* (Rasmussen and Williams 2006) to assimilate the observations and correct this locally. However, the primary capability here is to predict regional features, hence updating each class locally is of limited value, as it is the prediction at *unobserved locations* that are most interesting towards future sampling. Prediction works by taking a weighted average, using a likelihood function. There exist several strategies for finding an alternative weighting scheme. The current approach can be improved by including co-variates (e.g. wind measurements) to further determine some of the global conditions and find the weights conditioned on this. The weightage can also be found using optimization such as Sequential Least SQuares Programming (SLSQP) (Nocedal and Wright 2000) to minimize the error between the observations and a weighted combination of the classes.

To create sparse feature vectors of the images that is suited for subsequent classification we make use of dictionary learning techniques, as discussed in Section 3.3. Alternative methods using state-of-the-art deep learning approaches such as Convolutional Auto-encoders (Song et al. 2013; Aljalbout et al. 2018) can also be used to

create similar sparse representations without labeled data. However, given the limited size of the data set (74 images) these approaches are not feasible to use, even with data augmentation, as they require larger datasets, (e.g. LeCun et al. (2010)) in order to effectively tune the large number of parameters in the network; this is well-known issue for neural network architectures (Liu et al. 2017). In remote sensing applications, increasing the size of the data set is also not straightforward (Mountrakis et al. 2011), as discussed below in more detail.

There are inherent limitations in using SST satellite images due to cloud cover, such as the lack of uniformity in data acquisition (gaps in data), and the introduction of artifacts (e.g. interpolation around missing pixels). Thus, images have to be quality controlled before inclusion into the data-set, which may end up becoming smaller than expected. One question, in this regard, is whether numerical ocean models could be used instead. Ocean models can offer repeated and synoptic fields at different scales and resolutions, including surface and subsurface patterns. However, at the current skill level, coastal ocean variability and structure cannot be estimated on scales and accuracies sufficient for definite representation (Lermusiaux 2006), i.e. we cannot expect the models to tell us right where a filament, eddy, or bloom will be located, as an example. Assimilation of SST and high-frequency radar current data (see e.g. Frolov et al. (2012)) can certainly help the models move closer to ground-truth, but the nudging of 3-D fields with spotty 2-D data can create its own issues. For instance, spotty coverage in observations strongly limits how effectively a model can be corrected toward observed patterns, and this data assimilation can create artifacts that can be misinterpreted as "features of interest". Ultimately, effective observation of the ocean will require a joint effort between a range of data sources operating at different scales, where assimilation of *in-situ* data collected from autonomous robotic platforms will be crucial. Methods, such as the one suggested in this paper, will be an important tool to strategize this sampling.

An interesting application of this work, once a model is built, is to calculate a vehicle trajectory that results in the greatest level of class separation, that is, a path that best reduces the uncertainty in the model by fast convergence to a ranking (a probable scenario); similar to the analysis in Fig. 17a. Finding such a route can improve prediction of the oceanographic state by gathering temperature data in areas that will enhance discrimination, and can locate areas where multidisciplinary sensing from platforms will be most informative. Criteria such as expected reduction of variance, mutual information, or entropy can be used to find these locations.

Regional factors such as wind, bathymetry, and currents contribute to shaping the spatial patterns used for classification, which contribute to shaping SD and other parameters. Local oceanographic expertise is therefore necessary for validation. The method also involves some work related to tuning of hyperparameters, that will need supervision by oceanographers to ensure that a physically sound separation is used for building the model. Once configured, the method can operate on its own, and be automatically set to ingest new SST images. We use different settings for the dictionary size in the initial and secondary classification. This increases the length of the sparse codes, but allows for more spatial/pattern based information to be used in classification; initially this might not be necessary as the temperature dominates.

## 8 Conclusion

We have developed a new methodology for classifying remote sensing products such as SST towards building compact models that can be utilized by autonomous vehicles to provide environmental estimates. The method allows continuous processes to be segmented and compressed onto a basis of classes/scenarios that can be used as a framework for informing trajectory planning and sampling design. This approach can enhance the effectiveness of ocean observing campaigns and in the end help scientists understand the influence of regional oceanographic variables. We show examples using real data from Monterey Bay, California, where the compact model is combined with *in-situ* data to predict regional oceanographic states and bulk features. The results show that local observations can be used to yield information on a synoptic scale, but are limited to only resolving regional details. We have also evaluated the prediction error of the model and demonstrated the sensitivity to data, from the perspective of an ASV.

## References


Aharon M, Elad M and Bruckstein A (2006) k-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing* 54(11): 4311–4322.

Aljalbout E, Golkov V, Siddiqui Y, Strobel M and Cremers D (2018) Clustering with deep learning: Taxonomy and new methods. *arXiv preprint arXiv:1801.07648* .

Anderberg MR (1973) Cluster analysis for applications. *Probability and Mathematical Statistics, New York Academic Press, 1973* .

Bernstein M, Graham R, Cline D, Dolan JM and Rajan K (2013) Learning-based event response for marine robotics. *IEEE International Conference on Intelligent Robots and Systems* : 3362–3367DOI:10.1109/IROS.2013.6696835.

Blondeau-Patissier D, Gower JF, Dekker AG, Phinn SR and Brando VE (2014) A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in Oceanography* 123: 123 – 144. DOI:https://doi.org/10.1016/j.pocean.2013.12.008. URL http://www.sciencedirect.com/science/article/pii/S0079661114000020.

Chao Y, Chien S, Kinsey J, Flexas MM, Erickson ZK, Farrara J, Fratantoni D, Branch A, Chu S, Troesch M, Claus B and Society TO (2017) Satellites to Seafloor: Towards fully autonomous ocean sampling. *Oceanography* 30(2): 160–168.

Chen C (2012) *Signal and Image Processing for Remote Sensing, Second Edition*. Electrical engineering / remote sensing. Taylor & Francis. ISBN 9781439855966. URL https://books.google.no/books?id=QQDHl3L867QC.

Chen SS, Donoho DL and Saunders MA (2001) Atomic decomposition by basis pursuit. *SIAM review* 43(1): 129–159.

Chen Y, Nasrabadi NM and Tran TD (2011) Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing* 49(10): 3973–3985. DOI:10.1109/TGRS.2011.2129595.

Das J, Rajan K, Frolov S, Ryan JP, Py F, Caron DA, Sukhatme GS, Ryan JP, Caron DA and Sukhatme GS (2010) Towards marine bloom trajectory prediction for AUV mission planning. *Proceedings - IEEE International Conference on Robotics and Automation* : 4784–4790DOI:10.1109/ROBOT.2010.5509930.

Eidsvik J, Mukerji T and Bhattacharjya D (2015) *Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis*. Cambridge: Cambridge University Press. ISBN 9781139628785. DOI:10.1017/CBO9781139628785. URL https://www.cambridge.org/core/books/value-of-information-in-the-earth-sciences/61119AB2F707D557E49E00BF9FD6FE39.

Eleveld MA, Ruescas AB, Hommersom A, Moore TS, Peters SWM and Brockmann C (2017) An optical classification tool for global lake waters. *Remote Sensing* 9(5). DOI:10.3390/rs9050420. URL http://www.mdpi.com/2072-4292/9/5/420.

Emberton S, Chittka L, Cavallaro A and Wang M (2016) Sensor capability and atmospheric correction in ocean colour remote sensing. *Remote Sensing* 8(1). DOI:10.3390/rs8010001. URL http://www.mdpi.com/2072-4292/8/1/1.

Everitt B, Landau S, Leese M and Stahl D (2011) *Cluster Analysis*. Wiley Series in Probability and Statistics. Wiley. ISBN 9780470978443. URL https://books.google.no/books?id=w3bE1kqd-48C.

Fortin V, Abaza M, Anctil F and Turcotte R (2014) Why should ensemble spread match the rmse of the ensemble mean? *Journal of Hydrometeorology* 15(4): 1708–1713.

Frolov S, Garau B and Bellingham J (2014) Can we do better than the grid survey: Optimal synoptic surveys in presence of variable uncertainty and decorrelation scales. *Journal of Geophysical Research: Oceans* 119: 5071–5090. DOI:10.1002/2013JC009521.Received.

Frolov S, Kudela RM and Bellingham JG (2013) Monitoring of harmful algal blooms in the era of diminishing resources: A case study of the u.s. west coast. *Harmful Algae* 21-22: 1 – 12. DOI:https://doi.org/10.1016/j.hal.2012.11.001. URL http://www.sciencedirect.com/science/article/pii/S1568988312001503.

Frolov S, Paduan J, Cook M and Bellingham J (2012) Improved statistical prediction of surface currents based on historic HF-radar observations. *Ocean Dynamics* 62(7): 1111–1122. DOI:10.1007/s10236-012-0553-5.

Gonalves ML, Netto MLA, Costa JAF and Jnior JZ (2008) An unsupervised method of classifying remotely sensed images using kohonen selforganizing maps and agglomerative hierarchical clustering methods. *International Journal of Remote Sensing* 29(11): 3171–3207. DOI:10.1080/01431160701442146. URL https://doi.org/10.1080/01431160701442146.

James G, Witten D, Hastie T and Tibshirani R (2013) *An introduction to statistical learning*, volume 112. Springer.

Johannessen OM, Sandven S, Jenkins AD, Durand D, Pettersson LH, Espedal H, Evensen G and Hamre T (2000) Satellite earth observation in operational oceanography. *Coastal Engineering* 41(1-3): 155–176. DOI:10.1016/S0378-3839(00)00030-2.

Jolliffe I (2011) Principal component analysis. In: *International encyclopedia of statistical science*. Springer, pp. 1094–1096.

LeCun Y, Cortes C and Burges C (2010) Mnist handwritten digit database. *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist* 2: 18.

Lermusiaux PFJ (2006) Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *Journal of Computational Physics* 217(1): 176–199. DOI:10.1016/j.jcp.2006.02.010. URL https://doi.org/10.1016/j.jcp.2006.02.010.

Lilleborge M, Hauge R and Eidsvik J (2016) Information gathering in bayesian networks applied to petroleum prospecting. *Mathematical Geosciences* 48(3): 233–257.

Liu W, Wang Z, Liu X, Zeng N, Liu Y and Alsaadi FE (2017) A survey of deep neural network architectures and their applications. *Neurocomputing* 234: 11 – 26. DOI:https://doi.org/10.1016/j.neucom.2016.12.038. URL http://www.sciencedirect.com/science/article/pii/S0925231216315533.

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1. pp. 281–297.

Mairal J, Bach F, Ponce J and Sapiro G (2009) Online dictionary learning for sparse coding. In: *Proceedings of the 26th annual international conference on machine learning*. ACM, pp. 689–696.

Matérn B (2013) Spatial variation. *Meddelanden från Statens Skogsforskningsinstitut* 36(5): 1–144.

McKay J, Monga V and Raj R (2016) Robust sonar atr with pose corrected sparse reconstruction-based classification. In: *OCEANS 2016 MTS/IEEE Monterey*. pp. 1–5. DOI:10.1109/OCEANS.2016.7761189.

Mountrakis G, Im J and Ogole C (2011) Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3): 247 – 259. DOI:https://doi.org/10.1016/j.isprsjprs.2010.11.001. URL http://www.sciencedirect.com/science/article/pii/S0924271610001140.

Mulla DJ (2013) Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering* 114(4): 358 – 371. DOI:https://doi.org/10.1016/j.biosystemseng.2012.08.009. URL http://www.sciencedirect.com/science/article/pii/S1537511012001419. Special Issue: Sensing Technologies for Sustainable Agriculture.

Munk W (2002) Testimony to the U.S. Commission on Ocean Policy. URL http://govinfo.library.unt.edu/oceancommission/meetings/apr18_19_02/munk_statement.pdf.

Nash J (1990) The cholesky decomposition. *Compact numerical methods for computers: Linear algebra and function minimisation* 2.

Nocedal J and Wright S (2000) *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York. ISBN 9780387987934. URL https://books.google.no/books?id=epc5fX0lqRIC.

Oliver MJ and Irwin AJ (2008) Objective global ocean biogeographic provinces. *Geophysical Research Letters* 35(15). DOI:10.1029/2008GL034238. URL

https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GL034238.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12: 2825–2830.

Pelleg D and Moore A (2000) X-means: Extending k-means with efficient estimation of the number of clusters. In: *In Proceedings of the 17th International Conf. on Machine Learning*. Morgan Kaufmann, pp. 727–734.

Rasmussen EC and Williams CKI (2006) *Gaussian Processes for Machine Learning*. 1 edition. MIT Press. ISBN 026218253X. DOI:10.1142/S0129065704001899.

Rosenfeld LK, Schwing FB, Garfield N and Tracy DE (1994) Bifurcated flow from an upwelling center: a cold water source for monterey bay. *Continental Shelf Research* 14(9): 931–964.

Ryan JP, Fischer AM, Kudela RM, McManus MA, Myers JS, Paduan JD, Ruhsam CM, Woodson CB and Zhang Y (2010) Recurrent frontal slicks of a coastal ocean upwelling shadow. *Journal of Geophysical Research: Oceans* 115(12): 1–15. DOI: 10.1029/2010JC006398.

Saritha S and Kumar GS (2017) Inter-spectral and intra-spectral features for effective classification of remotely sensed images. *Procedia Computer Science* 115: 549 – 555. DOI:https://doi.org/10.1016/j.procs.2017.09.113. URL http://www.sciencedirect.com/science/article/pii/S1877050917319415. 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India.

Smith RN, Das J, Yi C, Caron DA, Jones BH and Sukhatme GS (2010) Cooperative multi-AUV tracking of phytoplankton blooms based on ocean model predictions. *OCEANS'10 IEEE Sydney, OCEANSSYD 2010* DOI:10.1109/OCEANSSYD.2010.5603594.

Song C, Liu F, Huang Y, Wang L and Tan T (2013) Auto-encoder based data clustering. In: *Iberoamerican Congress on Pattern Recognition*. Springer, pp. 117–124.

Tropp JA and Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory* 53(12): 4655–4666.

Walter V (2004) Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 58: 225–238.

Ward Jr JH (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301): 236–244.

Wilkinson GG (2005) Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing* 43(3): 433–440. DOI:10.1109/TGRS.2004.837325.

Wilks DS (2011) Cluster analysis. In: *International geophysics*, volume 100. Elsevier, pp. 603–616.

Xie Y, Sha Z and Yu M (2008) Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology* 1(1): 9–23. DOI:10.1093/jpe/rtm005. URL http://dx.doi.org/10.1093/jpe/rtm005.

## Acknowledgements

This work was part of the ENTiCE project [§] funded by the Research Council of Norway Project # 255303/E40, the *Nansen Legacy Program*, project number # 27272, and AMOS [¶], Center of Excellence, project number # 223254. TM acknowledges support from the Stanford Center for Earth Resources Forecasting and from the Dean of the Stanford School of Earth, Energy, and Environmental Sciences. MBARI authors are funded by a block grant from the David and Lucile Packard Foundation. The WaveGlider data was provided from the MBARI **CANON** campaign with support from Christopher Wahl, Devon Northcott, and Francisco Chavez. Finally, the authors are grateful for the support from MBARI engineering, science, and marine operations. The work was conducted when TOF was a visiting scholar at MBARI.

[§] http://sintef.no/en/projects/entice/

[¶] https://www.ntnu.edu/amos