MDPI

Article

Autonomous Learning of New Environments With a Robotic Team Employing Hyper-Spectral Remote Sensing, Comprehensive In-Situ Sensing and Machine Learning

David J. Lary†*, David Schaefer †, John Waczak †, Adam Aker †, Aaron Barbosa†, Lakitha O.H. Wijeratne †, Shawhin Talebi†, Bharana Fernando†, John Sadler†, Tatiana Lary†, and Matthew D. Lary†

* Correspondence: David.Lary@utdallas.edu
† Hanson Center for Space Sciences, University of Texas at Dallas, Richardson TX 75080, USA

Abstract: This paper describes and demonstrates an autonomous robotic team that can rapidly learn the characteristics of environments that it has never seen before. The flexible paradigm is easily scalable to multi-robot, multi-sensor autonomous teams, and is relevant to satellite calibration/validation and the creation of new remote sensing data products. A case study is described for the rapid characterisation of the aquatic environment, over a period of just a few minutes we acquired thousands of training data points. This training data allowed our machine learning algorithms to rapidly learn by example and provide wide area maps of the composition of the environment. Along side these larger autonomous robots two smaller robots that can be deployed by a single individual were also deployed (a walking robot and a robotic hover-board), observing significant small scale spatial variability.

Keywords: Machine Learning; Hyper-spectral Imaging; Robot Team; Autonomous; UAV; Robotic Boat

Citation: Lary, D.J.; Schaefer, D.; Waczak, J.; Aker, A.; Barbosa, A.; Wijeratne, L.O.H.; Talebi, S.; Fernando, B.; Sadler, J.; Lary, T.; Lary, M.D. Autonomous Learning of New Environments With a Robotic Team Employing Hyper-Spectral Remote Sensing, Comprehensive In-Situ Sensing and Machine Learning. *Sensors* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

This paper describes a robotic team that can rapidly learn new environments. The system described here demonstrates a flexible paradigm that is easily scalable to multi-robot, multi-sensor autonomous teams. A case study is described for the rapid characterisation of the aquatic environment. Other authors have described in detail various configurations of autonomous robots, for example [1–6].

The aquatic environment was chosen, as it includes extra challenges with regards to ease of access, further demonstrating the value of the approach. When considering the usefulness of being able to conduct such rapid surveys, it is worth noting that, for just the oil spill response use case alone, the National Academy of Sciences estimates that the annual oil spill quantities range from 1.7 million tons to 8.8 million tons. Over 70% of this release is due to human activities. The result of these spills include dead wildlife, contaminated water and oil-covered marshlands [7–10]. So being able to rapidly survey such areas to guide clean-up operations is of considerable use. It is also of use in

Figure 1. Photographs of the robot team during a Fall 2020 deployment in North Texas.

a wide variety of contexts, from general environmental surveys, to studying harmful algal blooms, to the clean-up operations after natural disasters, such as hurricanes, etc.

In the example described in this paper, the fully autonomous team includes a robotic boat that carries a suite of sensors to measure water composition in real time as well as a sonar, and an autonomous UAV equipped with a down-welling irradiance spectrometer, hyper-spectral and thermal imagers, together with an onboard Machine Learning (ML) capability. Figure 1 shows photographs of the robot team during a December 2020 deployment in North Texas.

Besides this capability being useful by itself, there is a wider significance for earth observing satellite missions. A key component to each and every space agency earth observation mission is the delivery of a suite of data products and the calibration/validation of these products. The paradigm demonstrated can reduce the time and cost of producing new remote sensing data products, while increasing functionality and data quality and providing new real-time automated calibration/validation capabilities.

The approach also provides enhanced capabilities for real-time onboard data product creation, reducing product delivery latency. The end-to-end demonstration uses all off-the-shelf components, representing a reduction in costs and risk when prototyping new mission concepts. A key element is the use of embedded machine learning, so we will refer to the approach as Rapid Embedded Prototyping for Advanced Applications (REPAA).

1.1. Hyper-Spectral Imaging

The human eye perceives the color of visible light in three bands using the cones, the photoreceptor cells in the retina (Figure 2). These three broad bands are red (centered on 564 nm), green (centered on 534 nm), and blue (centered on 420 nm). By contrast, instead of using just three broad bands, hyper-spectral cameras divide the spectrum into a very large number of narrow bands, in our case 463 bands from 391-1,011 nm. A hyper-cube is a three-dimensional dataset consisting of a stack of two-dimensional image layers each for a different wavelength. So for each pixel in the image we have a multi-wavelength spectra (spectral signature). This is shown schematically in the lower left of Figure 2. On the right we see a conventional RGB color image with only three bands, images for red, green and blue wavelengths.

Chemicals absorb light in a characteristic way. Their absorption spectra is a function of their chemical structure. Figure 3a shows the structure of chlorophyll and the associated absorption spectra. So that we can accurately calculate the reflectivity at each wavelength our autonomous UAV measures both the incident downwelling irradiance of incident solar radiation and a hyper-spectral imager pointed directly down at the earth's surface below the UAV. For every pixel we measure an entire spectrum with a hyper-spectral camera so we can identify chemicals within the scene.

An example reflectivity hyper-spectral data cube collected during a robot team deployment in North Texas during November 2020 is shown in Figure 3b. This data cube includes the area where an inert dye was released to test the system. The dye used was Rhodamine WT, a fluorescent, xanthene dye, that has long been used as a hydrologic tracer in surface water systems. The spectral signature of the dye is clearly visible in the hyper-spectral data cube. The top layer of the hyper-spectral data cube shows the regular RGB image, the 463 stacked layers below show the reflectivity (on a log-scale) for each wavelength band between 391 and 1,011 nm.

2. Materials and Methods

All the data for the machine learning data product creation was collected in a coordinated automated manner using the autonomous robotic team. An overview of the robotic team members and their sensor payloads is as follows.

2.1. Robotic Vehicles

A Maritime Robotics Otter (https://www.maritimerobotics.com/otter) autonomous boat was used. With a footprint of only 200 x 108 x 81.5 cm, a weight of 55 kg, and dual electrical fixed thrusters, it is an easily deployable asset that can be transported in a van or even within normal airliners to a survey site. With a cruise speed of 2 knots it has a duration of 20 hours from one charge of the batteries. It can use WiFi, cellular and an optional AIS receiver for communication to the control station.

A Freefly Alta-X (https://freeflysystems.com/alta-x) autonomous professional quad-copter was used. It was specifically designed to carry cameras, with a payload capacity of up to 35 lb, a long range data link, and autonomy provided by the Open PX4 flight stack. The open source QGroundControl software was used to control the autonomous operations (https://freeflysystems.com/support/alta-pro-support). QGroundControl is available for Mac, Windows, iOS and Android.

All of the robotic team members carry a high-accuracy GPS and INS so that every data point can be geo-located and time stamped. Each of the robots can also join the same network which connects the robots and their ground-control stations. Our robots use long-range Ubiquiti 5 GHz LiteBeam airMAX WiFi (https://www.ui.com). The airMAX Time Division Multiple Access (TDMA) protocol allows each client to send and receive data using pre-designated time slots managed by an intelligent AP controller. This

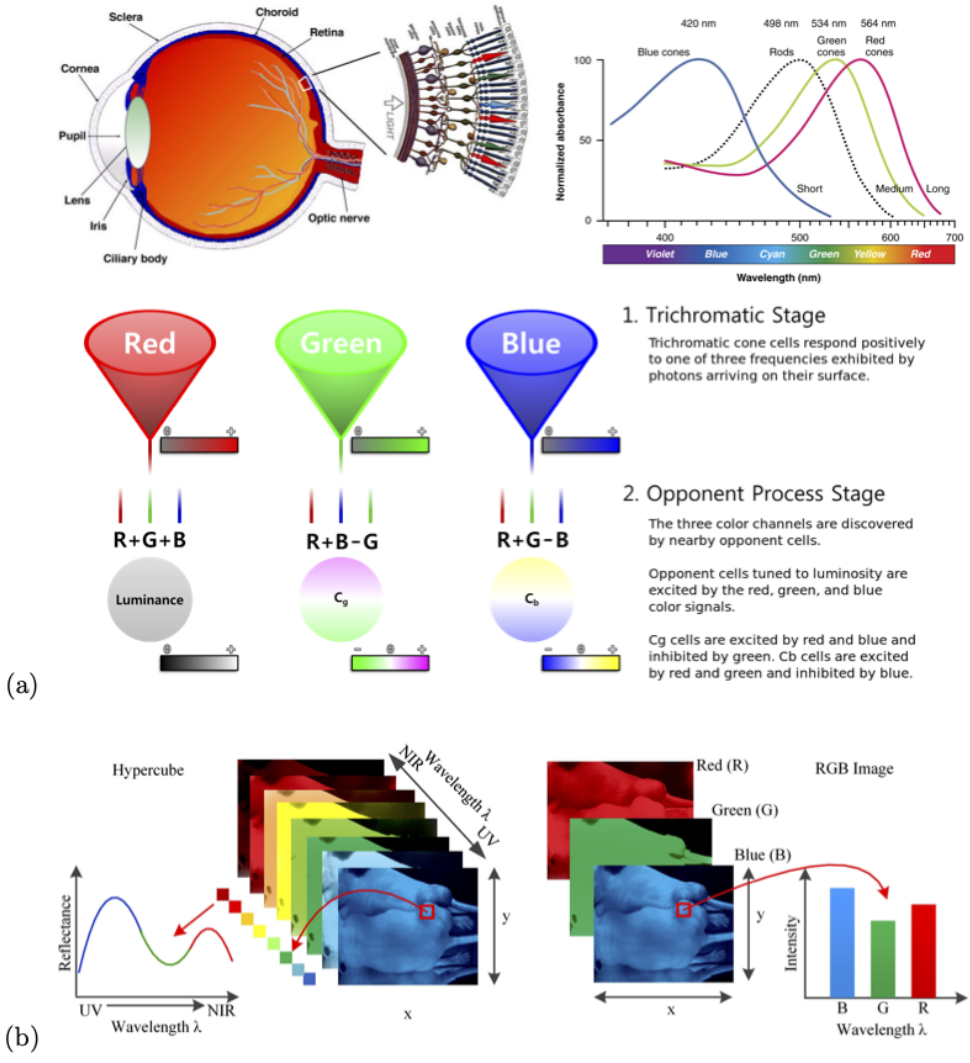

Figure 2. Panel (a) Trichromatic cone cells in the eye respond to one of three wavelength ranges (RGB). Panel (b) shows a comparison between a hyper-spectral data-cube and RGB images.

 doi:10.20944/preprints202102.0454.v1

time slot method eliminates hidden node collisions and maximizes airtime efficiency. This WiFi network is connected to the internet using a Cradlepoint cellular modem (https://cradlepoint.com).

This network also includes a local Synology network-attached storage (NAS) (https://www.synology.com) device in the robot team control trailer, which in real-time syncs the data collected to the NAS in our home laboratory in the university.

2.2. Boat Sensors

The robotic boat payload included a BioSonics MX Aquatic Habitat Echosounder sonar for rapid assessment and mapping of aquatic vegetation, substrate and bathymetry (https://www.biosonicsinc.com/products/mx-aquatic-habitat-echosounder/). Three Eureka Manta-40 multi-probes (https://www.waterprobes.com/multiprobes-and-sondes-for-monitori), a Sequoia Scientific LISST-ABS acoustic backscatter sediment sensor (https://www.sequoiasci.com/product/lisst-abs/), and an Airmar Technology Corporation 220WX ultra-sonic weather monitoring sensor (https://www.airmar.com/weather-description.html?id=153).

The first Manta-40 multi-probe included sensors for temperature and turbidity and Turner Designs Cyclops-7 submersible Titanium body fluorometers (https://www.turnerdesigns.com/cyclops-7f-submersible-fluorometer) for Chlorophyll A, Chlorophyll A with Red Excitation, Blue-Green Algae for fresh water (Phycocyanin), Blue-Green Algae for salt water (Phycoerythrin), and CDOM/FDOM. The second Manta-40 multi-

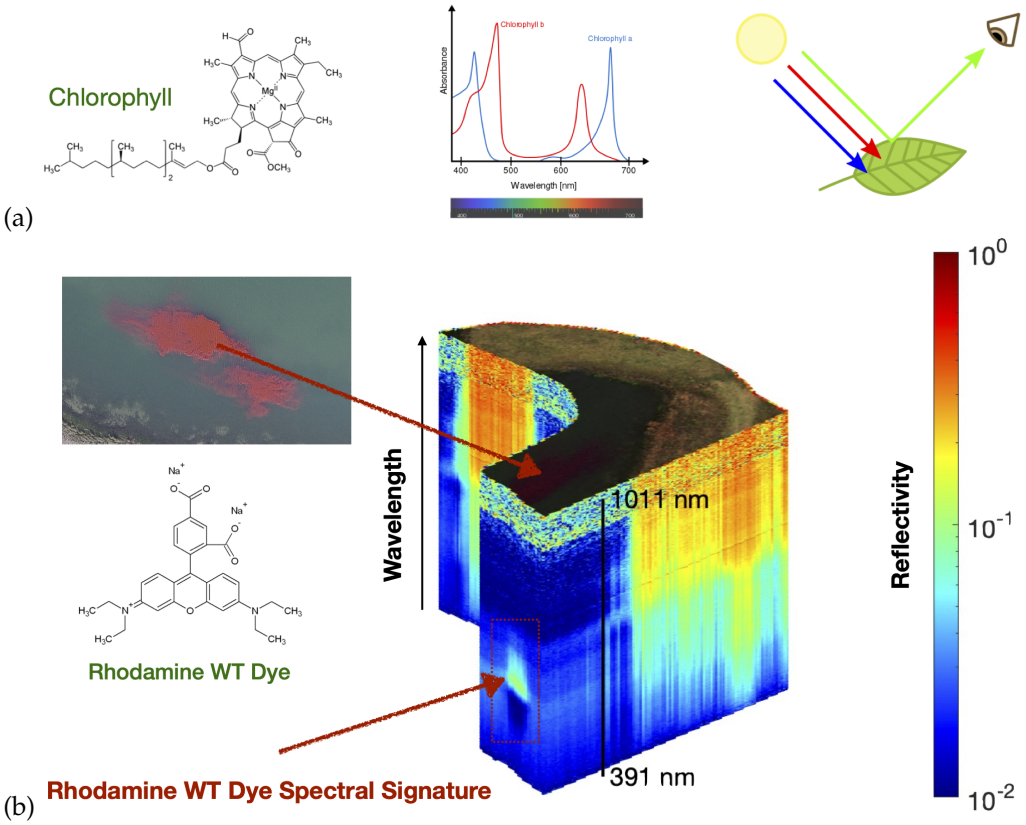

Figure 3. Panel (a) Chemicals absorb light in a characteristic way. Their absorption spectra is a function of their chemical structure. For every pixel we measure an entire spectrum with a hyper-spectral camera so we can identify chemicals within the scene. Panel (b) shows an example hyper-spectral data cube collected in North Texas on November 23, 2020. This particular data cube includes a simulant release, Rhodamine WT. The top layer of the hyper-spectral data cube shows the regular RGB image, the 463 stacked layers below show the reflectivity (on a log-scale) for each wavelength band between 391 and 1,011 nm.

probe included sensors for temperature, conductivity (with specific conductance, salinity, and total dissolved solids, TDS), pH (with separate reference electrode), optical dissolved-oxygen, turbidity and Ion Selective Electrodes by Analytical Sensors and Instruments (http://www.asi-sensors.com/) for ammonium (NH_4^+), bromide (Br^-), calcium (Ca^{++}), chloride (Cl^-), nitrate (NO_3^-), and sodium (Na^+). The third Manta-40 multi-probe included sensors for temperature, turbidity, a total dissolved gas sensor, and Turner Designs Cyclops-7 submersible Titanium body fluorometers for optical brighteners, crude oil, refined fuels, and tryptophan.

In addition, a portable Membrane Inlet Mass Spectrometer (MIMS) designed and built by Prof. Verbeck of the University of North Texas is available (but not used in these deployments) to switch every 3 seconds between sampling the water composition and the air composition.

2.3. Aerial Sensors

The aerial vehicle used a Gremsy H16 gimbal (https://gremsy.com/gremsy-h16) made with aircraft grade aluminum and carbon fiber to carry a Resonon Visible+Near-Infrared (VNIR) Pika XC2 (https://resonon.com/Pika-XC2) hyper-spectral camera (391–1,011 nm) with a Schneider Xenoplan 1.4/17 mm lens, and a FLIR Duo Pro R, (640x512, 25 mm, 30 Hz) combining a high resolution, radiometric thermal imager, 4K color camera, and a full suite of onboard sensors (https://www.flir.com/products/duo-pro-r/). On the top of the quad copter there is a sky facing Ocean Optics UV-Vis-NIR spectrometers measuring the incident down-welling irradiance allowing us to calculate reflectance.

2.4. Geo-rectification

The hyper-spectral data cubes collected are very large and are written in real time to the solid-state disk (SSD) attached to the Resonon Pika XC2. To facilitate the real-time processing of these files the Camera SSD is exported as a Network File System (NFS) mount so that a second onboard computer can geo-rectify the hyper-spectral data cubes as they are created. These hyper-spectral data cubes provide a visible and near infrared spectrum (391–1,011 nm) for each pixel. Once these data cubes are geo-rectified in real-time they are available for onboard machine learning using edge computing onboard the aerial vehicle.

2.5. Machine Learning

The accurate geo-tagging and time stamping of all data from all members of the robot team allows automation of the machine learning data product creation. For every location at which the robotic boat sampled the in-situ water composition we associate a VNIR remotely sensed spectrum (391–1,011 nm) provided by the hyper-spectral data cubes collected by the aerial-vehicle. This data is then be used for multi-variate non-linear non-parametric machine learning, where the inputs are the spectrum, in this case 462 values from the 391–1,011 nm spectra, and the outputs are each of the values measured in-situ by the robotic boat. A variety of machine learning approaches were used. These approaches included, shallow neural networks with hyper-parameter optimization, ensembles of hyper-parameter optimized decision trees, gaussian process regression with hyper-parameter optimization, and a super-learner including all of the previously mentioned approaches. Each empirical non-linear non-parametric fit is evaluated by constructing both a scatter diagram and a quantile-quantile plot of the values estimated by the machine learning model plotted against the actual values in the independent validation dataset.

The use of machine learning in this study builds on our heritage of using machine learning for sensing applications over the last two decades [11–28].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 February 2021 doi:10.20944/preprints202102.0454.v1

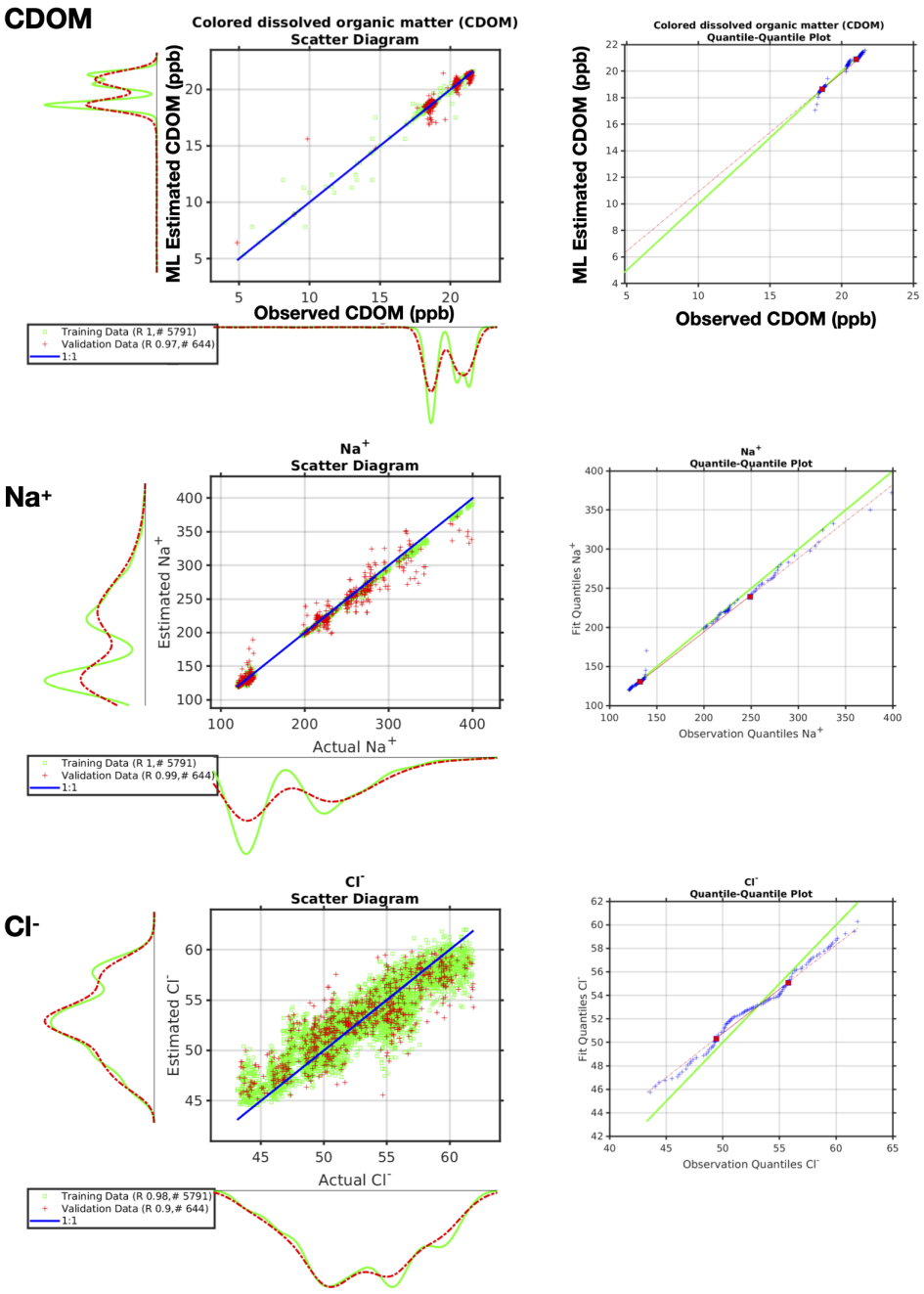

Figure 4. Machine learning performance quantified by both scatter diagrams and quantile-quantile plots utilizing data collected autonomously by the robot team during three exercises during November and December 2020 in North Texas. The three examples shown here are for CDOM, Na^+ and Cl^-. The scatter diagrams show the actual observations (mg/l) on the x-axis and the machine learning estimate on the y-axis. The green curves are for the training data, the red for the independent validation. The legend shows the number of points in the training and validation datasets and their associated correlation coefficients. The quantile-quantile plots show the observation quantiles on the x-axis and the machine learning estimate quantiles on the y-axis.

3. Learning Modes

We designed each component of our system to be flexible for different scenarios and deployment configurations. The entire system is called a Cyber Physical Observatory. To appreciate the benefits of this, a few basic definitions/descriptions are helpful. The

 doi:10.20944/preprints202102.0454.v1

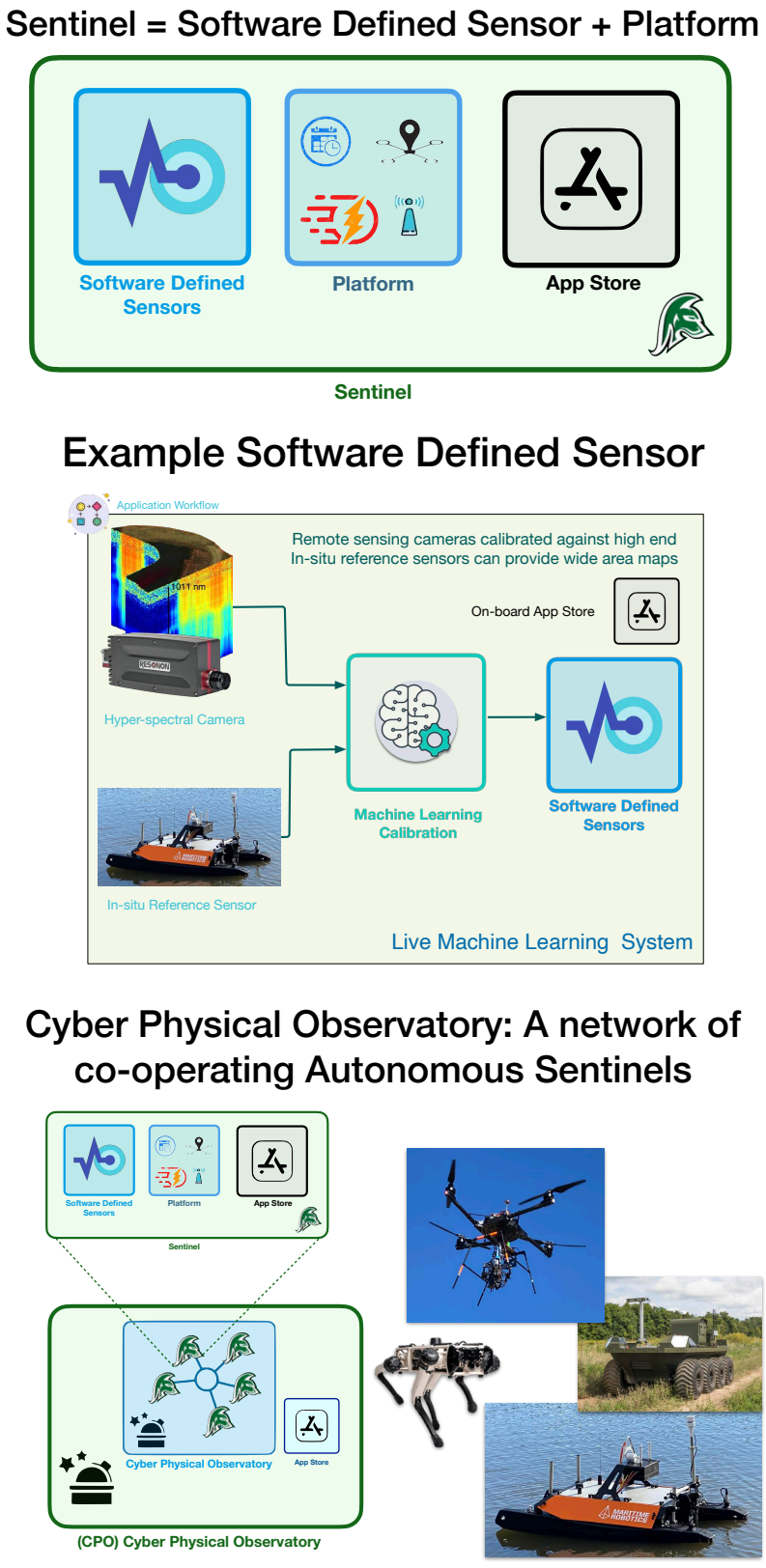

Figure 5. The Cyber Physical Observatory is a collection of sentinels that provide real-time data. A Sentinel is a Software Defined Sensor mounted on a Platform. A Platform supplies the Software Defined Sensor with power, timestamps for all observations, communication, and mobility where applicable. A Software Defined Sensor is a smart sensor package which combines a physical sensing system with machine learning providing a variety of calibrated data products which can be updated via an app store.

Cyber Physical Observatory is a collection of sentinels and/or robot teams that provide real-time data and actionable insights and whose capabilities can be updated via an app store. The Robot Team is a collection of co-operative autonomous sentinels. A Sentinel is a Software Defined Sensor mounted on a Platform. A Platform supplies the Software Defined Sensor with power, timestamps for all observations, communication, and mobility where applicable. In some of our other applications these even include wearable sensors. A Software Defined Sensor is a smart sensor package which combines a physical sensing system with software/machine learning providing a variety of calibrated data products which can be updated via an app store.

Two distinct machine learning modalities are useful when trying to rapidly learn new environments (Figure 6). Mode 1: Coordinated robots using onboard Machine Learning for specific data products. Mode 2: Unsupervised classification.

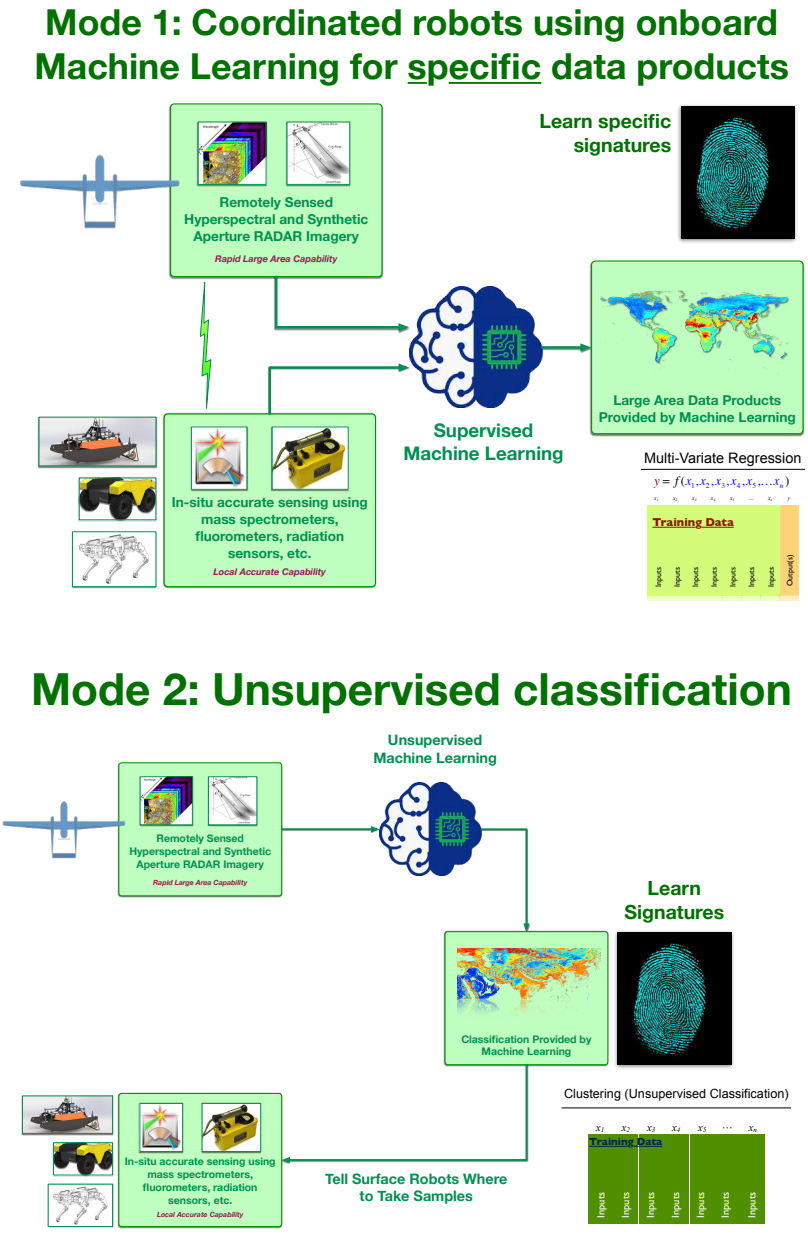

Figure 6. The autonomous robotic team operates in two modes. Mode 1: Coordinated robots using onboard Machine Learning for specific data products. Mode 2: Unsupervised classification.

 doi:10.20944/preprints202102.0454.v1

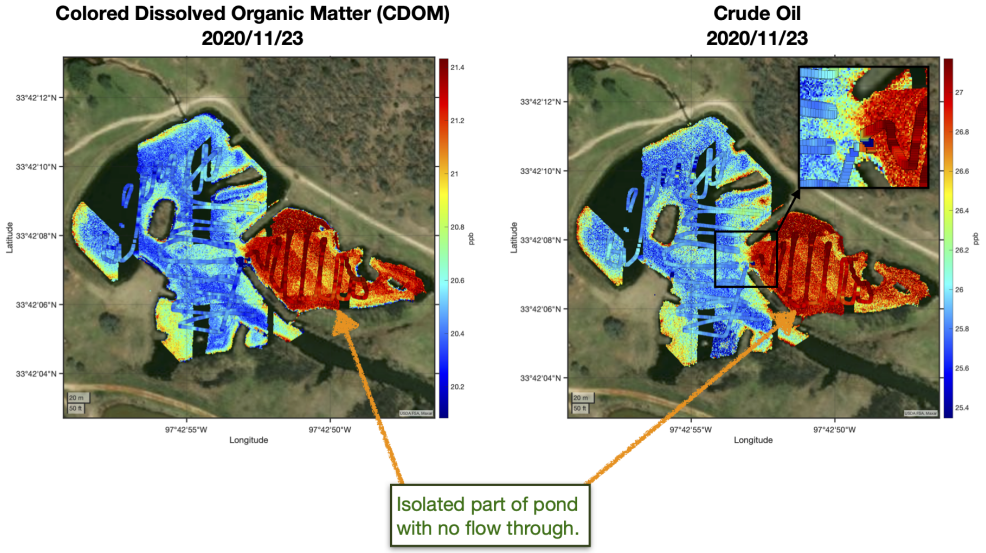

Figure 7. Example crude oil and colored dissolved organic mater (CDOM) data collected autonomously by the robot team on November 23, 2020 in North Texas. The maps show the CDOM and crude oil estimated from the hyper-spectral imager using machine learning as the background colors and the actual in-situ boat observations as the overlaid color filled squares. Note that the isolated part of the pond which has now fresh water in-flux has higher levels of CDOM and crude oil with a sharp gradient across the inlet in both the estimates using the hyper-spectral image and the boat observations.

In Mode 1 the robot team members rapidly collect the machine learning training data in a carefully coordinated. For our example deployment in North Texas during the Fall of 2020, over a period of about fifteen minutes thousands of precisely collocated measurements were made by the robotic team. The robotic boat autonomously measuring in-situ ground truth of a large array of parameters using the sensors described above, while the robotic aerial vehicle gathered remotely sensed observations of exactly the same locations using hyper-spectral and thermal imaging. These remotely sensed observations could be readily extended to cover a wider wavelength range and to include Synthetic Aperture Radar (SAR). Once the training data is acquired the machine learning algorithms can rapidly learn the mapping from the remotely sensed observations to the in-situ ground truth. Figure 4 shows three different examples of the validation of these autonomously acquired machine learning data products being independently verified using scatter diagrams and quantile-quantile plots.

Once the machine learning algorithm(s) have been trained they can then be used to rapidly provide wide-area maps with just the remotely sensed observations. Two examples of this are shown in Figure 7. These can be processed onboard the aerial vehicle and the results streamed in real-time to the ground control station. The robotic boat can then be autonomously tasked to verify the wide area maps by collecting independent validation data.

In Mode 2 we would like to perform a fine-grained multi-class surface classification of the entire domain. This is done by providing the remotely sensed data (in this case the hyper-spectral and thermal imagery) to an unsupervised classification. The unsupervised machine learning characterizes the distinct regions and zones in the area of interest. This can be particularly useful when trying to identify the location of particular contaminants, suggesting the optimum sampling patterns required beyond the usual clover leaf, star or box patterns used for contaminant searches.

4. Results

Over a period of just a few minutes we acquire thousands of training data points. This training data allows our machine learning algorithms to rapidly learn by example. The machine learning fit used here is an optimized ensemble of regression trees [29–31] with hyper-parameter optimization [32] implemented in Matlab version 2021a (https://www.mathworks.com) using the function `fitrensemble` with all hyper-paramter optimization selcted and parallel processing enabled. A loop is executed over all the variables measured by the robotic that we would like to estimate using the hyper-spectral imagery.

For each of these variables a balanced training dataset is constructed. This is done by considering each input and output variable in the training dataset in turn and calculating n percentiles, from each of these n percentile ranges covering the entire PDF, from each percentile range we select m random values (where $m < n$) for the training and a different set of random values for independent validation.

Figure 4 shows an example of the colored dissolved organic mater (CDOM) data collected autonomously by the robot team on November 23, 2020 in North Texas, along with some of the aqueous ion data. The panel shows a scatter diagram of the actual observations on the x-axis and the machine learning estimate on the y-axis. The green curves are for the training data, the red for the independent validation. On each axis we also show the associated PDFs. The ideal result is shown in blue (a slope of 1 and an intercept of zero for the scatter diagram).

Figure 7 shows maps of the CDOM and crude oil concentration estimated using the machine learning as the background colors and the actual in-situ boat observations as the overlaid color filled squares. Note that the isolated part of the pond which has now fresh water in-flux has higher levels of CDOM and crude oil with a sharp gradient across the inlet in both the estimates using the hyper-spectral image and the boat observations. We note that there is good agreement between the machine learning estimate and the actual in-situ boat observations.

5. Discussion

5.1. Limitations

The fidelity of the data products provided by the autonomous robotic team are limited by the training data it is able to acquire. For example, our remote sensing hyper-spectral camera in the demonstration use case presented here observes the spectral region 391-1,011 nm. It would be useful to extend this spectral region so we can see more chromophores, and to extend the type of remote sensing imaging, e.g. to include Synthetic Aperture RADAR (SAR).

It would also be useful for the boat to have larger pontoons so that it can carry our mass-spectrometer that can sample both the air and water, switching between the two inlets every three seconds.

We would also like to extend the machine learning approaches to include Physics Based machine learning such that the machine learning is constrained by known physical principles.

5.2. Automating Data Product Creation

A key factor in providing remotely sensed water composition products is providing a comprehensive database of water composition (e.g. SeaBASS, the publicly shared archive of in-situ oceanographic and atmospheric data maintained by the NASA Ocean Biology Processing Group https://seabass.gsfc.nasa.gov). The cost of making the measurements of ocean composition can be substantial because it involves a significant ship time as well as a large support team. Secondly, since the satellites are in a fixed orbit with a fixed viewing geometry, the number of coincidences between the shipboard water observations and the orbiting satellite observations are, by definition, limited. Typically several thousand coincident observations are used in the tuning and creation of a NASA

ocean data product. In the REPAA approach, the entire system can be automated and objectively optimized. Thus, with a data rate of one observation every second, in a matter of hours we can gather tens of thousands of observations in a totally automated, fully coordinated manner, as was demonstrated in North Texas during November and December 2020 (Figure 1). There is explicit coordination between the water observations taken from the robotic boat and the continuous aerial observations made by the robotic aerial vehicle carrying a hyper-spectral imager. The system can be deployed to very diverse environments across a matter of just weeks to months, so over a matter of just weeks to months, millions of coordinated, precisely coincident records can be made. Furthermore, we have previously demonstrated, the data can be randomly partitioned into training and independent validation sets, and using the onboard machine learning, transformed into optimal water composition data products, using many orders of magnitude more observations than before at a fraction of the cost and in a fraction of the time.

Aurin et al. [33] provides one of the most comprehensive training datasets to date for Chromophoric Dissolved Organic Matter (CDOM). Their Global Ocean Carbon Algorithm Database (GOCAD) for Chromophoric Dissolved Organic Matter (CDOM) encompasses 20,000–100,000+ records (depending on the variable considered) and it is based on oceanographic campaigns conducted across the world over the past 30 years at great expense. In contrast, the autonomous robotic team can collect around 20,000+ precisely coordinated training records per hour. By design, the robotic team makes precisely coordinated overpasses of exactly the same locations, this leads to providing a training dataset with a high data rate. By deploying the team on multiple occasions at a diversity of locations one can rapidly build a comprehensive training dataset.

The traditional approach for creating remote sensing data products, as shown on the left of Figure 8, is compared with the approach used in this study, shown on the right. Using the REPAA approach, data collection and the creation of derivative data products can be carried out on the same day, for example in the December 2020 exercises in North Texas (Figure 1).

5.3. Improving Product Quality & Automating Cal/Val

Critical in improving product quality is the comprehensive training data set, which spans as much parameter space and variability that is actually found in the real world. This necessitates making observations in a large number of diverse contexts. Being able to make these observations with such a highly automated platform is a tremendous step forward and costs less. In summary, our robotic platform can address the issue of small scale variability encountered across a satellite pixel. These capabilities assist continuing validation/quality control and can help optimize the waveband selection for future satellite instruments and missions.

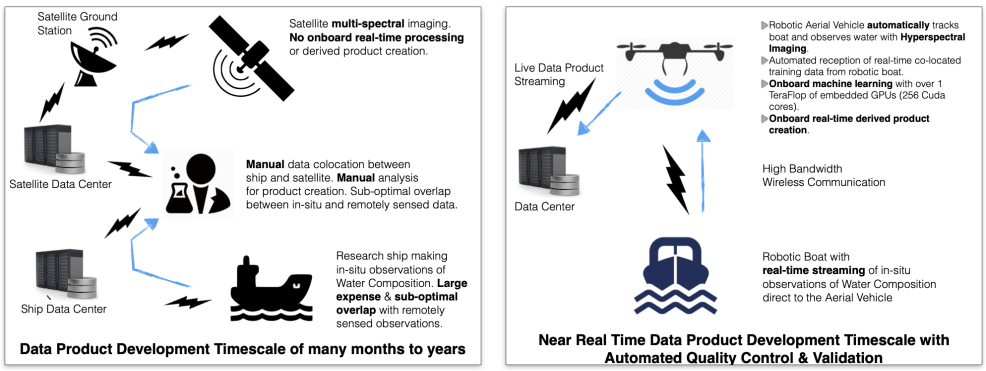

Figure 8. Schematics illustrating the traditional approach to creating remote sensing data products (left) and that used in this study (right).

 doi:10.20944/preprints202102.0454.v1

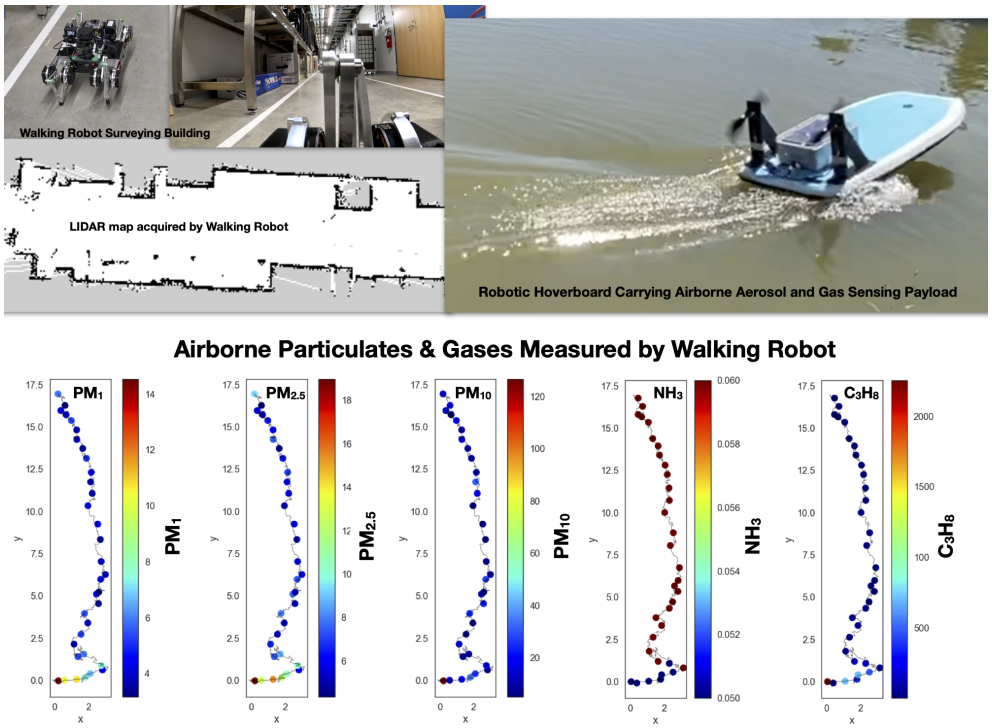

Figure 9. Photographs of the smaller walking robot (from Ghost Robotics) and a robotic hoverboard (conceived and built by Aaron Barbosa). For illustrative purposes both of these small robots carried exactly the same payload of sensors measuring the size spectrum of airborne particulates in the size range 0.3–43 microns and the abundance of a selection of gases. The laser scanner onboad the walking robot acquired a map of the vicinity while also measuring in-situ the atmospheric composition, finding very localized changes in the abundance of the airborne particulates of various sizes.

5.4. Reducing Latency for Product Delivery as well as Mission Risk, Cost, Weight and Size

Utilizing new embedded *onboard processing* (1 TeraFlop weighing just 88 g with a size of only 87 mm x 50 mm) for real-time onboard processing leads to reducing the latency in product delivery from hours/days to just the downlink time. The product delivery latency can be critical for decision support applications, such as oil spills, or other disaster response applications, and for routine forecasting and data assimilation applications. A risk reduction is also realized, by the ability to first deploy an end to end demonstrator, using entirely commercial off the shelf components and low cost aerial vehicles, with all software made Open Source.

5.5. Onboard App Store

There is currently a rapid enhancement in both observing capabilities and the embedded computing power from miniaturized low power devices. As these enhanced observing capabilities become routinely available on small cubesats (like hyperspectral imaging), the number of possible uses and applications for societal benefit grows. However, so does the bandwidth required for the downlink of the hyperspectral datacubes. So the possibility of onboard processing, for example using embedded machine learning, means that product creation can occur directly onboard the cubesats and then streamed live via the downlink. This reduces the latency of product creation and the bandwidth needed for the downlink. The next logical step, then, of a rapid prototyping and agile workflow, is an onboard app store, where new data products can be deployed to the remote sensing platform for seamless use onboard. A formalized development, testing, and deployment workflow with an app store facilitates an Earth-observing system that responds to the rapidly changing societal needs while maintaining a rigorous approach

to validation. This onboard app store can leverage the smart automated code generation that already exists off the shelf and is now routinely used for automobiles and aircraft across the world. The time has come for this to be the standard paradigm for earth observation as well.

5.6. Smaller Robots

There is also value in smaller robots that are easy to transport by a single individual. Figure 9 shows photographs of the smaller walking robot (from Ghost Robotics) and a robotic hover-board (conceived and built by Aaron Barbosa) that we deployed along size the larger autonomous robotic team for illustrative purposes. Both the walking robot and the robotic hover-board carried exactly the same payload of sensors that could be rapidly switched between the robots. The sensing payload measured every few seconds the full size spectrum of airborne particulates in the size range 0.3–43 microns and the abundance of a selection of gases. The laser scanner onboad the walking robot acquired a map of the vicinity while also measuring in-situ the atmospheric composition, finding very localized changes in the abundance of the airborne particulates of various sizes.

6. Conclusions

This paper described and demonstrated an autonomous robotic team that can rapidly learn the characteristics of environments that it has never seen before. The flexible paradigm is easily scalable to multi-robot, multi-sensor autonomous teams, and is relevant to satellite calibration/validation and the creation of new remote sensing data products. A case study was described for the rapid characterisation of the aquatic environment, over a period of just a few minutes we acquired thousands of training data points. This training data allowed our machine learning algorithms to rapidly learn by example and provide wide area maps of the composition of the environment. Along side these larger autonomous robots two smaller robots that can be deployed by a single individual were also deployed, a walking robot and a robotic hover-board, each measuring the full size spectrum of airborne particulates in the size range 0.3–43 microns and the abundance of a selection of gases, significant small scale spatial variability with evident in these hyper-localized observations.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, D.J.L.; methodology, D.J.L.; software, D.J.L and J.W.; field deployment and preparation D.J.L., D.S. J.W., A.A., A.B., L.O. H. W., S.T., B.F., J.S., M.D.L., and T.L.; validation, D.J.L.; formal analysis, D.J.L.; investigation, D.J.L.; resources, D.J.L.; data curation, D.J.L., J.W., A.A. and L.O.H.W.; writing—original draft preparation, D.J.L.; writing—review and editing, D.J.L.; visualization, D.J.L.; supervision, D.J.L.; project administration, D.J.L.; funding acquisition, D.J.L.

Funding: This research was funded by the following grants: The Texas National Security Network Excellence Fund award for Environmental Sensing Security Sentinels. SOFWERX award for Machine Learning for Robotic Team. Support from the University of Texas at Dallas Office of Sponsored Programs, Dean of Natural Sciences and Mathematics, and Chair of the Physics Department are gratefully acknowledged. The authors acknowledge the OIT-Cyberinfrastructure Research Computing group at the University of Texas at Dallas and the TRECIS CC* Cyberteam (NSF 2019135) for providing HPC resources that contributed to this research. URL: https://utdallas.edu/oit/departments/circ/

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Acknowledgments: Mr. Don MacLaughlin, Mr. Scotty MacLaughlin and the City of Plano, TX, are gratefully acknowledged for allowing us to deploy the autonomous robot team on their land. Without their assistance this work would not have been possible. Prof. Christopher Simmons is gratefully acknowledged for his computational support. Annette Rogers is gratefully acknowledged for support with arranging insurance coverage. Steven Lyles is gratefully acknowl-

edged for support arranging a secure place for the robot team. The authors acknowledge the OIT-Cyberinfrastructure Research Computing group at the University of Texas at Dallas and the TRECIS CC* Cyberteam (NSF 2019135) for providing HPC resources that contributed to this research. URL: https://utdallas.edu/oit/departments/circ/

Conflicts of Interest: The authors declare no conflict of interest

Abbreviations

The following abbreviations are used in this manuscript:

CDOM	Chromophoric Dissolved Organic Matter
GOCAD	Global Ocean Carbon Algorithm Database
GPS	Global Positioning System
INS	Inertial Navigation System
MIMS	Membrane Inlet Mass Spectrometer
ML	Machine Learning
NASA	The National Aeronautics and Space Administration
NFS	Network File System
REPAA	Rapid Embedded Prototyping for Advanced Applications
SeaBASS	SeaWiFS Bio-optical Archive and Storage System
SSD	Solid State Disk
UAV	Unmanned Aerial Vehicle
VNIR	Visible and Near-Infrared

References

1. Ingrand, F.; Ghallab, M. Deliberation for autonomous robots: A survey. *Artif. Intell.* **2017**, *247*, 10–44.
2. Islam, M.J.; Hong, J.; Sattar, J. Person-following by autonomous robots: A categorical overview. *The International Journal of Robotics Research* **2019**, *38*, 1581 – 1618.
3. Strom, D.P.; Bogoslavskyi, I.; Stachniss, C. Robust exploration and homing for autonomous robots. *Robotics Auton. Syst.* **2017**, *90*, 125–135.
4. Levent, H. Managing an Autonomous Robot Team : The Cerberus Team Case Study. 2005.
5. Davis, J.D.; Sevimli, Y.; Ackerman, M.; Chirikjian, G. A Robot Capable of Autonomous Robotic Team Repair: The Hex-DMR II System. 2016.
6. Husain, A.; Jones, H.; Kannan, B.; Wong, U.; Pimentel, T.; Tang, S.; Daftry, S.; Huber, S.; Whittaker, W. Mapping planetary caves with an autonomous, heterogeneous robot team. *2013 IEEE Aerospace Conference* **2013**, pp. 1–13.
7. Fingas, M.F.; Brown, C.E. Review of oil spill remote sensing. *Spill Science & Technology Bulletin* **1997**, *4*, 199 – 208. The Second International Symposium on Oil Spills, doi:http://dx.doi.org/10.1016/S1353-2561(98)00023-1.
8. Fingas, M. *Oil Spill Science and Technology*; Gulf Professional Publishing, 2010.
9. Liu, Y.; MacFadyen, A.; Ji, Z.; Weisberg, R. *Monitoring and Modeling the Deepwater Horizon Oil Spill: A Record Breaking Enterprise*; Geophysical Monograph Series, Wiley, 2013.
10. Cornwall, W. Deepwater Horizon: After the oil. *Science* **2015**, *348*, 22–29, [http://www.sciencemag.org/content/348/6230/22.full.pdf]. doi:10.1126/science.348.6230.22.
11. Brown, M.E.; Lary, D.J.; Vrieling, A.; Stathakis, D.; Mussa, H. Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS. *International Journal of Remote Sensing* **2008**, *29*, 7141–7158.
12. Lary, D.; Aulov, O. Space-based measurements of HCl: Intercomparison and historical context. *Journal of Geophysical Research: Atmospheres* **2008**, *113*.
13. Lary, D.J.; Remer, L.; MacNeill, D.; Roscoe, B.; Paradise, S. Machine learning and bias correction of MODIS aerosol optical depth. *IEEE Geoscience and Remote Sensing Letters* **2009**, *6*, 694–698.
14. Lary, D.; Waugh, D.; Douglass, A.; Stolarski, R.; Newman, P.; Mussa, H. Variations in stratospheric inorganic chlorine between 1991 and 2006. *Geophysical Research Letters* **2007**, *34*.
15. Lary, D.; Müller, M.; Mussa, H. Using neural networks to describe tracer correlations **2004**.
16. Lary, D.J. *Artificial intelligence in geoscience and remote sensing*; INTECH Open Access Publisher, 2010.
17. Malakar, N.K.; Knuth, K.H.; Lary, D.J., Maximum Joint Entropy and Information-Based Collaboration of Automated Learning Machines. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*; Goyal, P.; Giffin, A.; Knuth, K.H.; Vrscay, E., Eds.; 2012; Vol. 1443, *AIP Conference Proceedings*, pp. 230–237. doi:10.1063/1.3703640.
18. Lary, D.J.; Faruque, F.S.; Malakar, N.; Moore, A.; Roscoe, B.; Adams, Z.L.; Eggelston, Y. Estimating the global abundance of ground level presence of particulate matter (PM2.5). *Geospatial health* **2014**, *8*, 611–630.
19. Lary, D.; Lary, T.; Sattler, B. Using Machine Learning to Estimate Global PM2. 5 for Environmental Health Studies. *Environmental health insights* **2015**, *9*, 41.

20. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* **2016**, *7*, 3–10.
21. Kneen, M.A.; Lary, D.J.; Harrison, W.A.; Annegarn, H.J.; Brikowski, T.H. Interpretation of satellite retrievals of PM2.5 over the Southern African Interior. *Atmospheric Environment* **2016**, *128*, 53–64.
22. Liu, X.; Wu, D.; Zewdie, G.K.; Wijerante, L.; Timms, C.I.; Riley, A.; Levetin, E.; Lary, D.J. Using machine learning to estimate atmospheric Ambrosia pollen concentrations in Tulsa, OK. *Environmental Health Insights* **2017**, *11*, 1–10.
23. Nathan, B.J.; Lary, D.J. Combining Domain Filling with a Self-Organizing Map to Analyze Multi-Species Hydrocarbon Signatures on a Regional Scale. *Environmental Modeling and Assessment* **2019**, *191*.
24. Lary, D.J.; Zewdie, G.K.; Liu, X.; Wu, D.; Levetin, E.; Allee, R.J.; Malakar, N.; Walker, A.; Mussa, H.; Mannino, A.; others. Machine Learning Applications for Earth Observation. In *Earth Observation Open Science and Innovation. ISSI Scientific Report Series*; Springer, 2018; Vol. 15, pp. 165–218.
25. Lary, M.A.; Allsop, L.; Lary, D.J. Using Machine Learning to Examine the Relationship Between Asthma and Absenteeism. *Environmental Modeling and Assessment* **2019**, *191*.
26. Zewdie, G.K.; Lary, D.J.; Levetin, E.; Garuma, G.F. Applying deep neural networks and ensemble machine learning methods to forecast airborne ambrosia pollen. *International journal of environmental research and public health* **2019**, *16*, 1992.
27. Zewdie, G.K.; Lary, D.J.; Liu, X.; Wu, D.; Levetin, E. Estimating the daily pollen concentration in the atmosphere using machine learning and NEXRAD weather radar data. *Environmental Monitoring and Assessment* **2019**, *191*, 418.
28. Wijeratne, L.O.; Kiv, D.R.; Aker, A.R.; Talebi, S.; Lary, D.J. Using Machine Learning for the Calibration of Airborne Particulate Sensors. *Sensors* **2020**, *20*, 99.
29. Breiman, L. Random Forests. *Machine Learning* **2004**, *45*, 5–32.
30. Belgiu, M.; Dragut, L. Random forest in remote sensing: A review of applications and future directions. *Isprs Journal of Photogrammetry and Remote Sensing* **2016**, *114*, 24–31.
31. Probst, P.; Wright, M.N.; Boulesteix, A. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2019**, *9*.
32. Nocedal, J.; Wright, S. *Numerical optimization*; Springer Science & Business Media, 2006.
33. Aurin, D.; Maninno, A.; Lary, D.J. Remote Sensing of CDOM, CDOM Spectral Slope, and Dissolved Organic Carbon in the Global Ocean. *Applied Sciences* **2018**, *8*, 2434.