

Ocean Data Analysis with R Programming for Early Career Ocean Professionals (ECOPs) - Asia [ONLINE]

A training report by IOC/OTGA/ECOPs

Online 9-31/10/23



Contents	
Introduction	2
Overview	2
Participant Demographics	2
Outreach platforms	2
Duration	3
Tools	3
Course Outline and Learning Outcomes	3
Module 1: Fundamental concepts in programming in R.....	3
Module 2: Data exploration and analysis.....	4
Module 3: Multivariate analysis.....	4
Module 4: Data visualization and graphics in R	4
Assessment	5
Lessons learnt	Ошибка! Закладка не определена.
Feedback	6
Suggestions for improvement	7
Testimonials	Ошибка! Закладка не определена.
Class Photos	8
Live sessions.....	Ошибка! Закладка не определена.
List of Participants.....	8
Training Schedule	12
Course Organisers and Lecturers	14

Introduction

Overview

The online course on Ocean Data Analysis with R Programming was envisioned to be an introductory course for marine conservation for Asian Early Career Ocean Professionals (ECOPs) leveraging the use of R programming language and software environment (RStudio) for data exploration, cleaning, organizing, analysis, and visualization in a more powerful way. With this initiative our aim was to contribute to the call for Decade Actions from the United Nations Decade of Ocean Science for Sustainable Development by focusing on addressing the Ocean Challenges 7, 9 and 10. The course consisted of a series of webinars introducing topics related to analyze oceanographic data (physical, biological, chemical) and fisheries data with R. The webinars included lectures by invited experts and hands-on tutorial sessions on analysis of oceanographic data for the Asian ECOPs to gain useful skills and knowledge in utilizing programming R as a tool to provide robust analyses in coastal and marine environment. The objectives of the training were as follows:

- Be familiar with the R environment.
- Learn how to use R for the most common data tasks: loading, cleaning, transforming and visualizing data.
- Produce meaningful descriptive statistics and informative graphs.

Participant Demographics

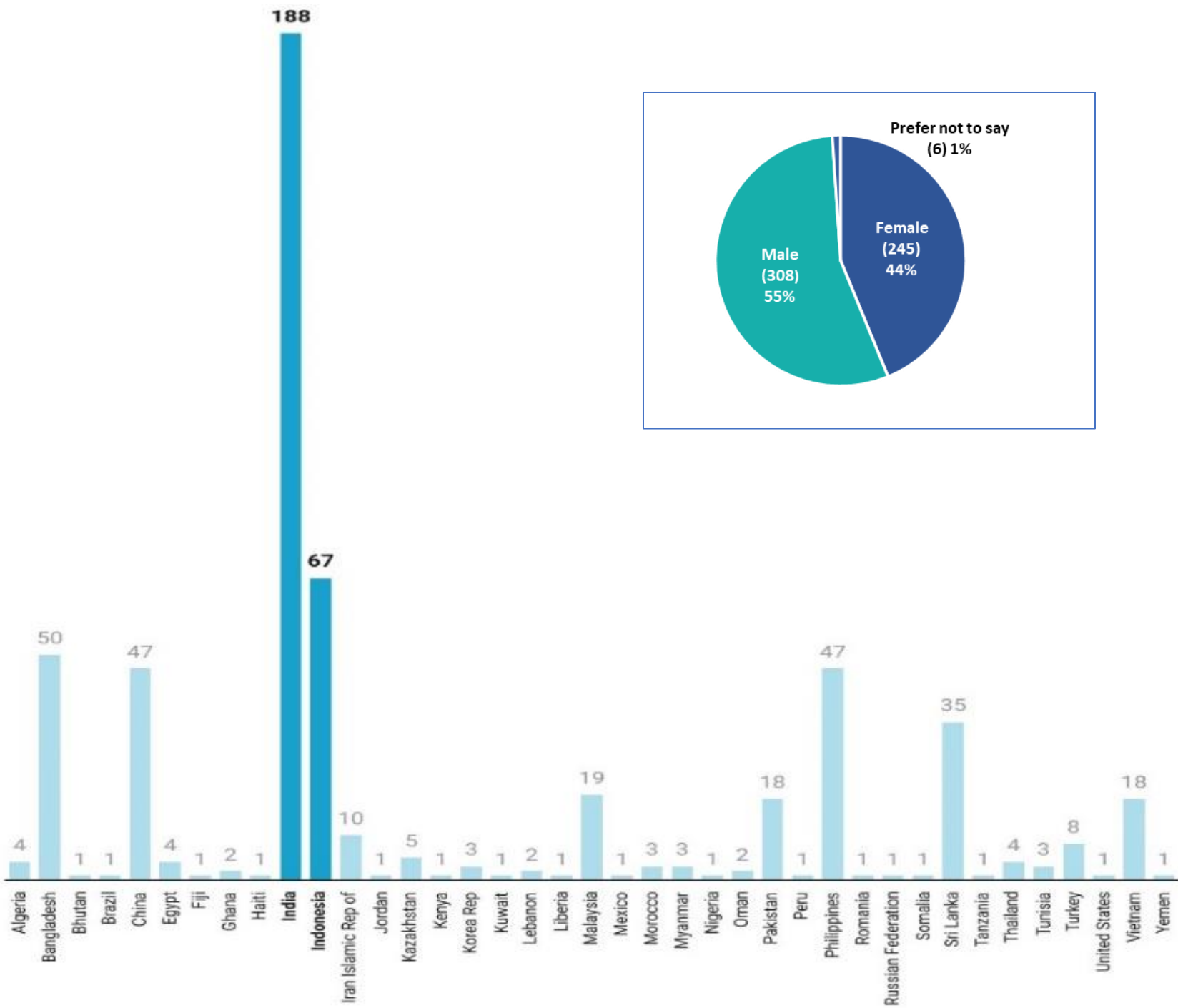
Applications from around the world were received, totaling 559 submissions, with the majority originating from Asia and a few from other regions. This widespread interest underscores the significance and appeal of the training courses offered. Given the training's emphasis on hands-on learning and active engagement throughout both live and asynchronous sessions, a rigorous selection process led to the acceptance of 62 participants. An additional 10 applicants were placed on a waiting list to fill any potential vacancies resulting from dropouts. The geographical distribution of the applicants' nationalities is illustrated below.

Nationality

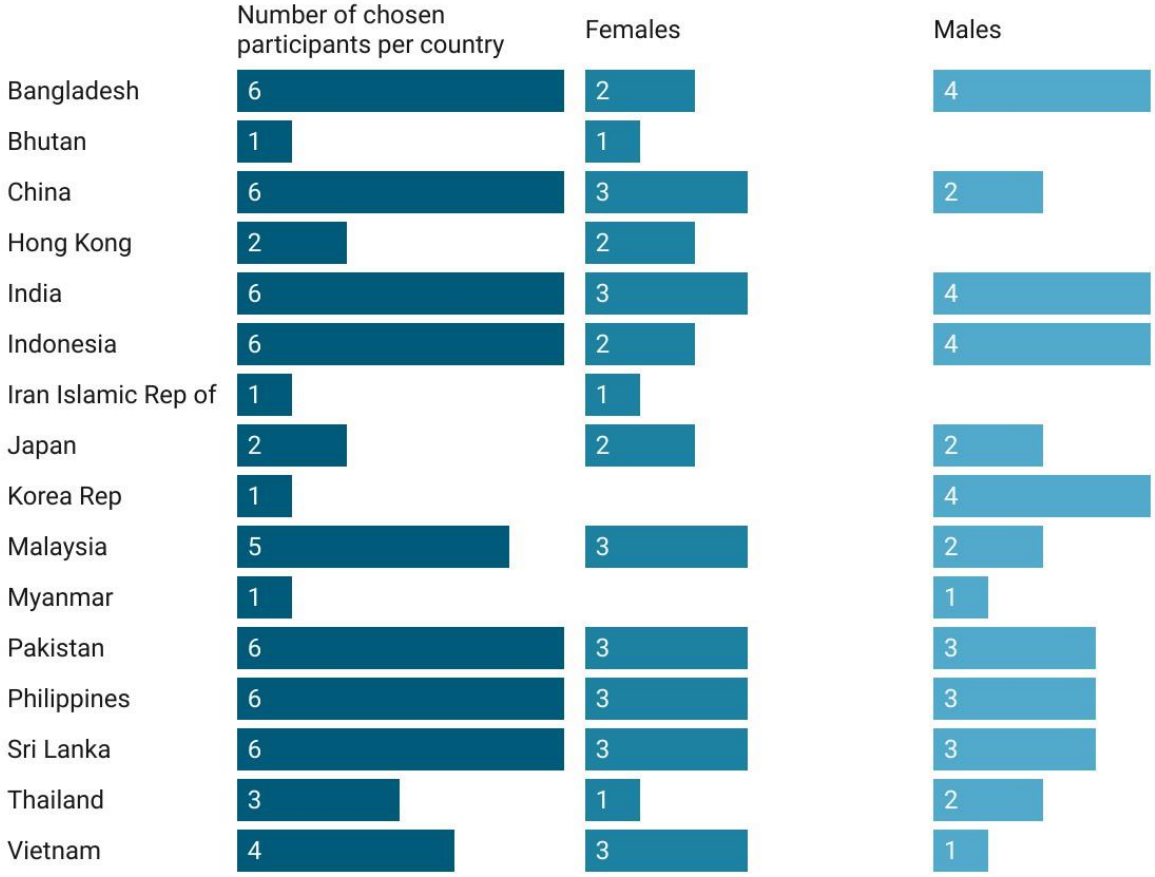
Participants from Asia

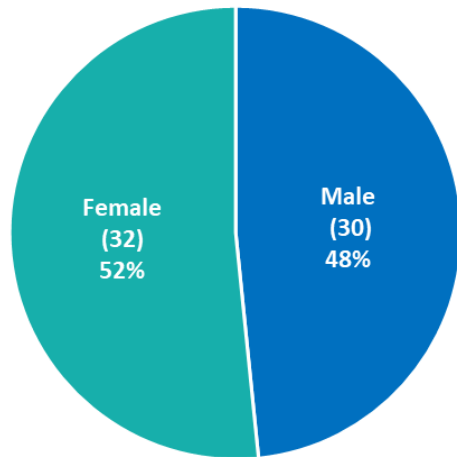


Moreover, the distribution of applications per nationality and the gender ratio of all applicants are depicted below. Predominantly, a substantial number of applications were submitted from India and Indonesia. Additionally, 44% of the participants are female and whereas 55% were male.



From the list of 62 selected participants, individuals were chosen from 16 different countries. These countries include Bangladesh, Bhutan, China, Hong Kong, India, Indonesia, Iran (Islamic Republic), Japan, the Republic of Korea, Malaysia, Myanmar, Pakistan, Philippines, Sri Lanka, Thailand, and Vietnam. The distribution among countries and genders is presented below, where 52% of the chosen participants were female. For a detailed listing of participants per nationality, please refer to the "List of Participants."

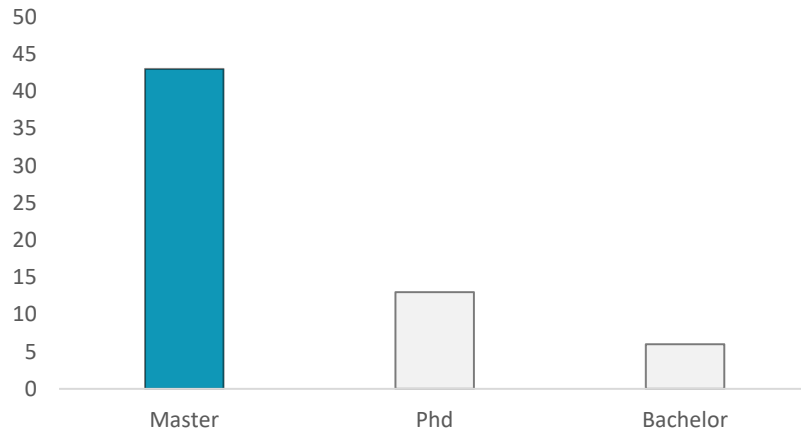




Our selected participants, ranging in age from 22 to 38, had an average of three years of professional experience.

Average & Ranges	Years
Average age among participants	29
Age range	22-38
Average years of experience	3
Experience range	0-11

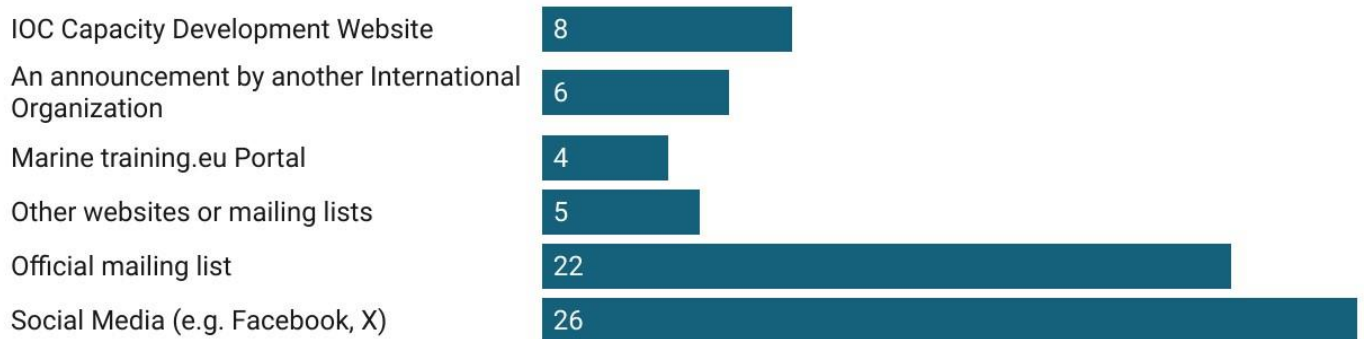
It's noteworthy that 68% of our selected participants hold a master's degree as their highest academic qualification.



Outreach platforms

The selected participants expressed that they learned about the course through the mailing list; however, others discovered it through a different website rather than the official one. It is noteworthy that the mailing list is creating a greater impact than other means of outreach.

Outreach platforms



Among the selected to join the training, 18 of them completed the hands-on training. The training required significant commitment from the trainees.

Duration

The three-week online course combined synchronous and asynchronous learning elements. It included webinars, hands-on exercises, Q&A, and presentations, with active sessions lasting a maximum of 2 hours. Participants engaged in asynchronous work, covering reading assignments, tutorials, assignments, and group work. Synchronous sessions were held three times a week, focusing on troubleshooting, Q&A, and discussions to enhance the overall learning experience.

Tools

The online training, facilitated through the OTGA platform, incorporated hands-on sessions using RStudio and the R programming language for processing and analyzing satellite data. RStudio's user-friendly interface allowed participants to seamlessly integrate live code with narrative text, equations, and visualizations. Participants needed a laptop or desktop, the latest version of web browsers (Chrome, Edge, Firefox, or Safari), and access to the internet. A Linux system, preferably Ubuntu 16.04 or above, was recommended for an optimal learning experience. These specifications ensured a smooth and effective online training experience for participants in oceanography, marine sciences, and related fields.

Course Outline and Learning Outcomes

The online course provided an introductory experience to the R programming language and RStudio software for Early Career Ocean Professionals with no prior R experience. Focused on data exploration, cleaning, organization, analysis, and visualization, the program utilized relevant oceanographic and fisheries datasets to enhance practical application. By the end of the course, participants were expected to be familiar with the R environment, proficient in common data tasks, such as loading and transforming data, and capable of producing meaningful descriptive statistics and informative graphs. The curriculum aimed to equip learners with essential skills for effective data manipulation and analysis in oceanography, marine sciences, and related fields.

Module 1: Fundamental concepts in programming in R

The course covered several key topics, starting with the installation and setup of R and RStudio, along with the installation of packages and an overview of the environment. Participants delved into exploring essential R packages, particularly focusing on Tidyverse and dplyr packages.

Fundamental concepts such as vectors, matrices, factors, lists, data frames, and sorting were thoroughly examined, emphasizing their practical applications within the R programming language.

This module aimed to provide participants with a strong foundation in navigating and utilizing crucial elements of R for effective data manipulation and analysis in oceanography, marine sciences, and related disciplines.

Module 2: Data exploration and analysis

The course progressed to advanced topics, encompassing the testing of assumptions for statistical analysis, data transformation techniques, and univariate statistical tests. Participants also delved into the complexities of multiple regression analysis, with a focus on applying these statistical methods within the R programming language. By exploring assumption testing, data transformation, and various statistical tests, participants gained the ability to produce scatter plots, boxplots, and time series plots using ggplot, apply universal plot settings, describe and apply faceting in ggplot, modify aesthetics in existing ggplot plots, and demonstrate the creation of complex and customized plots from data in a data frame.

Module 3: Multivariate analysis

In Module 3, the course progressed to multivariate analysis, covering Principal Component Analysis (PCA), Redundancy Analysis (RDA), and clustering techniques within the R programming language. Participants engaged in hands-on exploration of these advanced statistical methods, aiming to enhance their proficiency in applying multivariate analysis to diverse datasets relevant to oceanography, marine sciences, and related disciplines. Practical examples and interactive sessions provided valuable insights into the application of PCA, RDA, and clustering for effective data interpretation and pattern recognition.

Additionally, the module included the testing of assumptions for statistical analysis, univariate statistical tests, and linear models (simple and multiple regression). Upon successful completion of this module, participants were expected to explain common assumptions in statistical analysis, perform common univariate statistical tests in R, and demonstrate how to fit and interpret linear models.

Module 4: Data visualization and graphics in R

In the final module, Module 4, the course centered on data visualization and graphics in R. Participants not only learned various plotting techniques using the base plotting system in R but also delved into the powerful `ggplot2` package for crafting sophisticated and customized visualizations. Through practical exercises, participants gained hands-on experience in creating visually compelling data representations. The module's goal was to equip learners with the skills to effectively communicate complex information through graphical displays. Additionally, participants learned how to perform unconstrained ordination, co inertia analysis, and clustering in R, enhancing their ability to reduce the dimensionality of multivariate data, compare ordination techniques, and employ unsupervised machine learning for data grouping. After successfully completing this module, participants should be proficient in performing unconstrained ordination, co inertia analysis, and clustering in R.

Assessment

The assessment structure included asynchronous homework assignments designed to provide students with practical problem-solving opportunities, reinforcing their understanding of the course material. Additionally, participants were engaged in an individual or team project, where they applied their acquired skills to conduct a data analysis project. The project involved using the OBIS platform or their dataset, ensuring a real-world application of the course concepts. The final assessment component was a presentation, wherein students showcased their data analysis using R. This comprehensive evaluation framework aimed to gauge participants' proficiency in applying R programming to practical scenarios, fostering a deeper understanding of the material, and encouraging effective communication of analytical insights.

Course feedback

Feedback

Through course evaluation, the learners expressed that their expectations were met and that the course content and delivery was in line with the call. They also provided areas for improvement. A summary of a select feedback is illustrated in the following tables.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The learning outcomes are clearly stated, measurable and are appropriate to the level of the target audience	9	5	0	1	0
Expectations for student learning were clearly defined	6	6	1	2	0
The course covered the content announced/expected	12	2	1	0	0
I achieved the course stated aims and objectives	6	5	3	1	0
The course platform is well-organized and easy to navigate.	6	5	3	1	0
All course pages are readable and visually consistent.	7	6	2	0	0
All course pages are functionally consistent and communicate course information clearly and in sequential order throughout the course.	7	5	2	0	1
Multimedia files are clear, adequate, compatible with multiple operating systems and requires only a free, standard, and easily downloadable plug-in.	7	6	2	0	0
The course content is appropriate to the goals of the course.	10	4	0	1	0
The learning activities helped me achieve the stated learning outcomes.	7	6	2	0	0

The relationship between completing learning activities and meeting the learning outcomes was clearly explained	6	7	1	1	0
During this course, I was often engaged in learning activities such as discussion forums, wikis, chat, projects, group work, etc.	4	6	3	2	0
The course facilitators provided effective guidance and feedback.	8	5	2	0	0
During this course, I had the opportunity to interact with the instructor as often as needed	6	7	2	0	0
During this course, I had the opportunity to interact with other learners	3	6	4	2	0

	Very good	Good	Neutral	Poor	Very Poor	Does not apply (self-paced course)
How do you rate the response from the course facilitators to your questions or issues raised?	8	4	3	0	0	0
How do you rate the response from the OTGA Secretariat to your questions or issues raised regarding the use of the OceanTeacher e-Learning Platform?	7	5	2	0	0	1

Suggestions for improvement

According to the students' feedback here are the main areas of improvement and their reasoning:

Course Structure and Materials:

- Enhance clarity in instructions to improve the overall course-following experience.
- Ensure alignment between session content and written materials.
- Create a comprehensive PDF with course content and materials for easy reference.

Technological Infrastructure:

- Improve the course website for a more convenient and user-friendly experience.
- Address missing content issues, such as grey rectangles on the course website.
- Explore alternative communication platforms, like a Facebook learning group or WhatsApp group, to facilitate student interaction.

Learning Engagement and Assessment:

- Integrate quizzes for better understanding assessment and interactive learning.
- Encourage post-class discussions among students to foster collaboration.
- Provide additional support and resources for new learners to ease the learning curve.

Communication and Support:

- Implement timely reminder emails for class disruptions or rescheduled sessions.
- Ensure consistent updates on the course website to avoid confusion.
- Enhance accessibility by adding captions to live session recordings.
- Consider extending the course duration for a more comprehensive exploration of topics.
- Introduce more active learning components, including exercises, to reinforce theoretical concepts.
- Develop a comprehensive course workbook and incorporate more ocean-specific data into assignments.
- Increase examples and interaction time with facilitators to address individual queries.
- Strengthen communication channels between participants and facilitators to build a supportive learning community.

[Class Photos](#)

[List of Participants](#)

Below is the list of 62 approved trainees out of which 18 completed the training. Please see the pie graph of successful trainees.

Name	Last	Genero	Nationality
------	------	--------	-------------

1	Abdul Wahab	Khan	Male	Pakistan
2	Abu Bokkar	Siddique	Male	Bangladesh
3	Adhikari Arachchiralage	Ishan Kavinda Adhikari	Male	Sri Lanka
4	Ahalya	Suresh	Female	Sri Lanka
5	Albertina	Dias	Female	India
6	Amalia Irma	Nurwidya	Female	Indonesia
7	Amalina	Ibrahim	Female	Malaysia
8	Ananya	Bhuiya	Female	Bangladesh
9	Ariana Isabel Santos	Agustines	Female	Philippines
10	Bamunavita Gamage	Pasindu Bimsara	Male	Sri Lanka
11	Bassam	Nasir	Male	Pakistan
12	Bimali Madeesha Keerthisinghe	Kosgoda Pathiranage	Female	Sri Lanka
13	Cailian	Liu	Female	China
14	Chenyi	Zou	Female	China
15	Damith	Pasindu	Male	Sri Lanka
16	Dayana	Mathew	Female	India
17	Dimas Didik	Prayogo	Male	Indonesia
18	Elnaz	Aghaali	Female	Iran Islamic Rep Of
19	Gaurav	Anand	Male	India
20	Hendry	Siagian	Male	Indonesia
21	Imtiaj Ahmed	Easty	Male	Bangladesh
22	Jatdilok	Titioatchasai	Male	Thailand
23	Jing	Yang	Female	China
24	Ke	Xu	Male	China
25	Mai Anh	Nguyen	Female	Vietnam

26	Manh Linh	Nguyen	Male	Vietnam
27	Maria Jordana	Olano	Female	Philippines
28	Mark Paulo	Tolentino	Male	Philippines
29	Maryam	Khan	Female	Pakistan
30	Md. Merajul Khan	Jewel	Male	Bangladesh
31	Michelle Glory	G Jonik	Female	Malaysia
32	Mohamad Lukman Aidid	Bin Mohd Yusoff	Male	Malaysia
33	Mohammad Saifur	Rahman	Male	Bangladesh
34	Muhammad	Halim	Male	Indonesia
35	Muhammad Khalid	Ahmad Kamal	Male	Malaysia
36	Nabila	Nusrat	Female	Bangladesh
37	Pavithra	Balamurugan	Female	India
38	Peng	Lian	Male	China
39	Pranita	Rimal	Female	Bhutan
40	Putri	Utari	Female	Indonesia
41	Raffi	Isah	Male	Philippines
42	Reyland	Alegroso	Male	Philippines
43	Sahini	Perera	Female	Sri Lanka
44	Sathish	Chennuri	Male	India
45	Satrio	Hani Samudra	Male	Indonesia
46	Shahrukh	Zohaib	Male	Pakistan
47	Shenghui	Li	Female	China
48	Shubha	Singh	Female	Fiji
49	Sumyat	Naing	Male	Myanmar
50	Syeda Urooj	Fatima	Female	Pakistan

51	Tania	Sultana	Female	Bangladesh
52	Tengku Nur Alia	Tengku Kamarol Aznam	Female	Malaysia
53	Thi Ngoc Oanh	Bui	Female	Vietnam
54	Tushar	Lad	Male	India
55	Tuyen	Phan Thi Ha	Female	Vietnam
56	Ushna	Bint E Ishfaq	Female	Pakistan
57	Valerie	Culis	Female	Philippines
58	Veronica	Lam	Female	China
59	Winnie	Awuor	Female	Kenya
60	Wonkeun	Choi	Male	Korea Rep
61	Yangyang	Zhao	Male	China
62	Yitno	Suprpto	Male	Indonesia

Training Schedule

Module	Date	Lesson	Activity	Time
Module 1			Fundamental concepts in programming in R	
	9th October	Lesson 1	- Introduction and housekeeping	20 min
			- Installing and setting R and RStudio, installing packages and understanding the environment	90 min
			- Break	10 min
	10th Oct	Lesson 2	- Exploring essential packages including Tidyverse, dplyr etc	55 min
			- Understanding vectors, matrices, factors, lists, dataframes, sorting and their use in R	55 min
			- Break	10 min
	11th Oct		- Students carry out own practice with the OBIS or own data	
Module 2			Data visualization and graphics in R	
	12th Oct	Lesson 1	- Plotting with ggplot2 (introduction)	25 min
			- Building your plots iteratively	
			- Box plots	25 min
		- Break	10 min	
		Lesson 2	- Plotting time series data	60 min


	13th Oct		- Plotting time series data continues	20 min
		Lesson 3	- Integrating the pipe operator with ggplot2 - Faceting - ggplot2 themes	90 min
			- Break	10 min
	16th Oct		- Customization - Arranging plots - Exporting plots	60 min
			- Students choose their topics/research questions for projects (Individually or in groups)	60 min
Module 3			Data exploration and analysis	
	17th Oct	Lesson 1	- Testing assumptions for statistical analysis	55 min
			- Data transformation	55 min
			- Break	10 min
	18th Oct		- Students carry out own practice with the OBIS or own data/ work on projects	
	19th Oct	Lesson 2	- Univariate statistical tests	110 min
	20th Oct	Lesson 3	- Multiple regression analysis	55 min
			- Break	10 min
			- Multiple regression analysis continued	55 min
Module 4			Multivariate analysis	



	23rd Oct	Lesson 1	- PCA and RDA	110 min
			- Break	10 min
	24th Oct	Lesson 2	- Clustering	60 min
			- Overall Q & A session	60 min
	25th Oct		- Students carry out own practice with the OBIS or own data/ work on projects	
	26th Oct		- Students carry out own practice with the OBIS or own data/ work on projects	
	27th Oct		- Students hand in project reports	
	30th Oct		- Working on slides for presentations	
	31st Oct		- Group presentations and course wrap up	150 min

Course Organisers and Lecturers

The training course was devised by Gabriel Juma and Chunhua Jiang, and organized with the support of the Ocean Teacher Global Academy Secretariat. The training was delivered with contributions from Tan Qiao-guo, Margot Deschamps, Paula Nieto

See below the short bio of all contributors.

	<p>Bio</p> <ul style="list-style-type: none"> ● PhD Candidate in Coastal Ecology at Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research – Germany ● Holds a Postgraduate Diploma in Oceanography from Alfred Wegener Institute and MSc in Environmental Science from Chuka University-Kenya
<p>Gabriel Akoko Juma Coordinator, BlueCaD Project</p>	

	<ul style="list-style-type: none"> • Interests in Coastal Ecology with focus on Role of terrestrial runoff on nearshore environment, as well as Ocean Literacy <p>Role</p> <ul style="list-style-type: none"> • Development of course concept and workplan • Review of training modules on OTGA platform • Course moderation and co-hosting • Evaluation of assignments • Feedback and report writing • Course coordinator
 <p>OTGA Secretariat</p>	<p>Ocean Teacher Global Academy (OTGA) delivers customized training for ocean experts and professionals to increase national and regional capacity in coastal and marine sciences, services, and management.</p> <p>Role</p> <ul style="list-style-type: none"> • Hosting of e-learning platform • Provision of pedagogic support • Provision of administrative support
 <p>Chunhua JIANG</p> <p>State Key Laboratory for Estuarine and Coastal Research East China Normal University, State Key Laboratory for Estuarine and Coastal Research</p>	<p>No info available in the Ocean Expert platform</p>
<p>Tan Qiao-guo</p>	

	<p>Role</p> <p>Led the lecture on Fundamental Concepts in R Programming, covering topics from installing R and RStudio to exploring essential packages and understanding data structures. Also delivered the presentation on Multivariate Analysis, specifically covering PCA and RDA.</p>
<p>Paula Nieto PhD Student National Center for Genomic Analysis, Barcelona, Spain</p>	<p>Role</p> <p>In charge of the session on Data visualization in R graphics and provided guidance building plots iteratively, and customize, arrange and export the plots.</p>
<p>Margot Deschamps PhD Student Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research:</p>	<p>PhD Student Sea Shelf System Ecology at the Alfred Wegener Institute, Biological Institute Helgoland</p> <p>Role</p> <p>In charge of the Data exploration and analysis where it was taught testing assumption for statistical analysis, data transformation, univariate statistical tests, multiple regression analysis, and clustering multivariate analysis.</p>