Data Analysis with Python Programming for Early Career Ocean Professionals (ECOPs)

Online: 26th February – 19th March 2024

A training report by Xochitl Elias (coordinator ECOP Central America) & Gabriel Akoko Juma (coordinator- BlueCaD Project)

ECOP Central America

OceanTeacher Global Academy

I. Introduction

A) Course overview

A specialized Python programming course was conducted for early-career ocean professionals in Central America. This course responded to the increasing need for technical skills in the study of oceanographic data, equipping participants with essential programming abilities to manage, analyze, and visualize diverse ocean datasets.

The course introduced participants to the fundamentals of Python programming, with a focus on using the language for effective data processing, analysis, and visualization in the context of marine research. Tailored specifically for early-career professionals and oceanography students with little to no prior experience in Python, the training provided hands-on experience with real-world datasets. These datasets spanned multiple disciplines, including physical, chemical, and biological oceanography, as well as fisheries data, which are vital for understanding ocean health and sustainability.

By enhancing the technical capabilities of participants, the course contributed to several key Sustainable Development Goals (SDGs), particularly:

- SDG 14: Life Below Water, by empowering participants to analyze oceanographic data more effectively, thereby supporting efforts to conserve and sustainably use the oceans, seas, and marine resources.
- SDG 4: Quality Education, by offering high-quality, accessible programming education to early-career professionals, fostering the next generation of ocean researchers.
- SDG 13: Climate Action, by enabling participants to explore and analyze climate-relevant ocean data, contributing to better-informed decision-making in climate adaptation and mitigation efforts.

The use of oceanographic datasets in training allowed participants to better understand the challenges facing marine environments and how data analysis can support research, conservation, and sustainable management of marine ecosystems.

B) Participant Demographics

The course received a total of 114 applicants, 40 of whom were from Central America. Of these, 32 enrolled in the course, with 7 successfully completing it and receiving certificates. The majority of participants (56%) were female, while 43% were male (Figure 1).

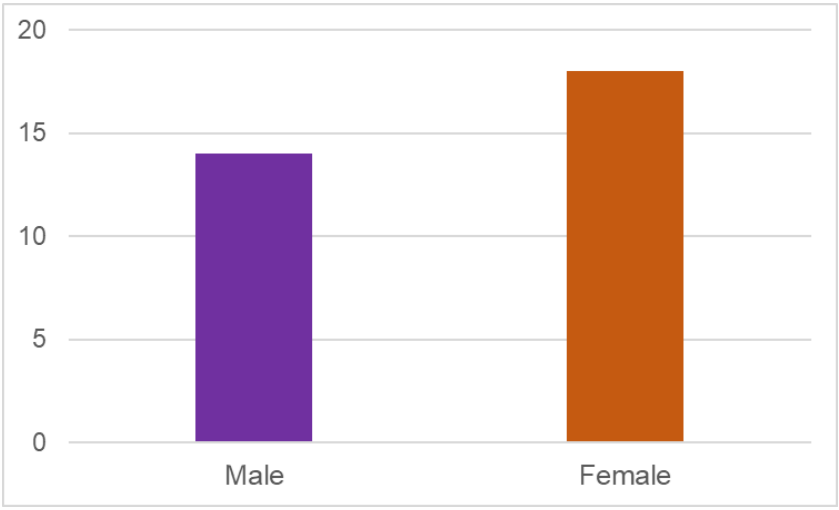

Figure 1. Gender of Central American participants

The nationalities of the participants were fairly balanced. The majority were from Costa Rica (28%), followed by Guatemala (16%) and Nicaragua (16%). However, the fewest participants were from El Salvador (9%). Nearly all participants resided in Central America, with the exception of one participant living in Spain (Figure 2).

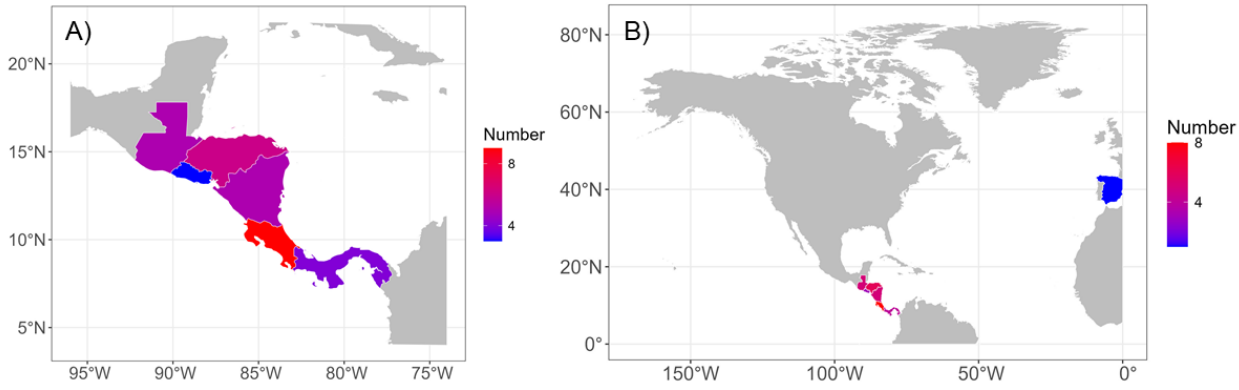

Figure 2. Nationality (A) and country of residency (B) of participants

II. Course outline and learning outcomes

The course was divided into four modules, each designed to progressively build participants' skills. We started with the installation of Python on various platforms, followed by basic data analysis techniques. The third module focused on analyzing geospatial data, and the final module covered plotting geospatial results. Each module will be explained in more detail in the following sections.

A) Module 1: Introduction to programming with Python

Overview:

In this module, participants learned the basics of Python programming, including how to install and use Python for data analysis. The installation of Anaconda was covered for Linux, Mac, and Windows users, with an explanation of its benefits, such as access to more up-to-date packages via conda-forge compared to the default Anaconda channel.

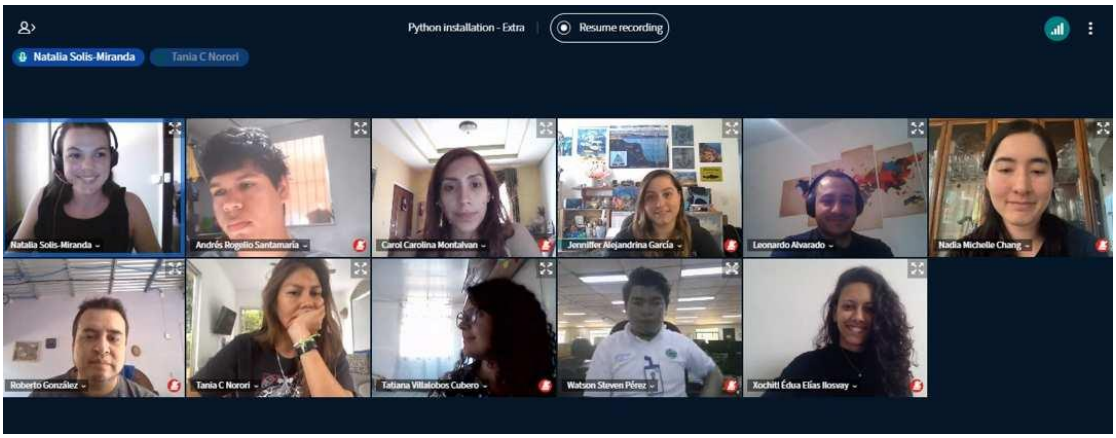

Figure 3. Participants during the life session in Module 1

Participants were introduced to Python fundamentals, including array operations, key libraries like NumPy, and methods for reading and analyzing basic tabular data. In the exercise, they created a plot showing the standard deviation and median of pigments data for each year using numpy.std() and numpy.median() functions.

Learning outcomes:

- Environment Setup with Mamba or Conda: Participants were able to set up a Python programming environment using Mamba or Conda for efficient package management.

- Python Installation: Participants learned to install and configure Python across different operating systems, ensuring they could run Python scripts on their own machines.

- Understanding Python Basics: Participants developed foundational Python skills, including working with variables, data structures, and basic functions.

- Introduction to Google Colab: Participants were introduced to Google Colab, a cloud-based platform, enabling them to run Python code without local setup.

- Working with Programming Notebooks: Participants learned to use Jupyter Notebooks and similar programming notebooks for interactive coding, analysis, and documentation.

B) Module 2: Data analysis with Pandas

Overview:

In this module, participants learned the basics of data analysis using the Pandas library, with a particular emphasis on working with dates and times using the datetime library. The module provided an overview of how to efficiently handle and analyze temporal data, starting with the creation of datetime objects. Participants were introduced to various methods for working with dates and times, including how to convert strings into datetime objects by specifying the correct format.

Additionally, participants explored how to use datetime within a Pandas DataFrame to perform calculations, such as comparing, filtering, or aggregating data based on time. This enabled them to manipulate and analyze datasets with temporal components more effectively.

For the exercise, participants were required to apply these concepts by creating their own scatter plot using their own data and submitting the resulting Python script.

Learning outcomes:

- Use Pandas: Participants learned to utilize the Pandas library for data manipulation and analysis, gaining skills in handling and processing tabular data.
- Create and Differentiate Plot vs. Scatterplot: Participants acquired the ability to create various types of plots, including line plots and scatter plots, and understood the differences between them and their appropriate use cases.
- Use Datetime: Participants became proficient in using the datetime library to manage and analyze time-based data, including creating and working with datetime objects.
- Process Data: Participants learned to process real-world datasets, such as those from the National Data Buoy Centre (NDBC), applying their skills to clean, analyze, and interpret data.

C) Module 3: Data analysis

Overview:

In this module, participants worked with the Xarray Python package to open and process spatial and temporal datasets, such as satellite measurements. The focus of the lesson was on exploring a global ocean geochemistry dataset. Participants learned how to define the path to the dataset, load it, and examine its basic structure.

The Xarray package was introduced as a powerful tool for handling multidimensional data, offering labels in the form of dimensions, coordinates, and attributes on top of raw NumPy-like arrays. This functionality provided a more intuitive and concise way to work with complex datasets, reducing the likelihood of errors during analysis.

For the exercise, participants selected a dataset from the same publication used in the lesson. They were tasked with choosing a specific region, analyzing the changes over time, and calculating the average of the selected data. To present their findings, participants created at least two different types of plots to illustrate the results and prepared a brief presentation of their analysis.

Learning outcomes:

- Basic Usage of Xarray: Participants developed the skills to open files, process data, create plots, and save data using Xarray.
- Comparison of Xarray with NumPy and Pandas: Participants gained insight into the differences and similarities between Xarray, NumPy, and Pandas, and learned how to choose the appropriate tool for various tasks.
- Advantages of Xarray: Participants explored the benefits of Xarray, including its support for labeled dimensions and more intuitive data manipulation compared to NumPy and Pandas.
- Work with Datetime: Participants acquired the ability to handle and analyze time-based data using Xarray's datetime functionalities.

D) Module 4: Geospatial data analysis

Overview:

The module emphasized several key functionalities of Xarray that enhance data manipulation. Participants discovered how Xarray allows for the assignment of dimension names in a DataArray, which makes it easier to manage and interpret multidimensional data. They also learned to subset and select data by coordinate values, a process that is similar to NumPy-like array operations but with added convenience. Xarray supports approximate selection and interpolation, enabling more sophisticated data queries and analysis.

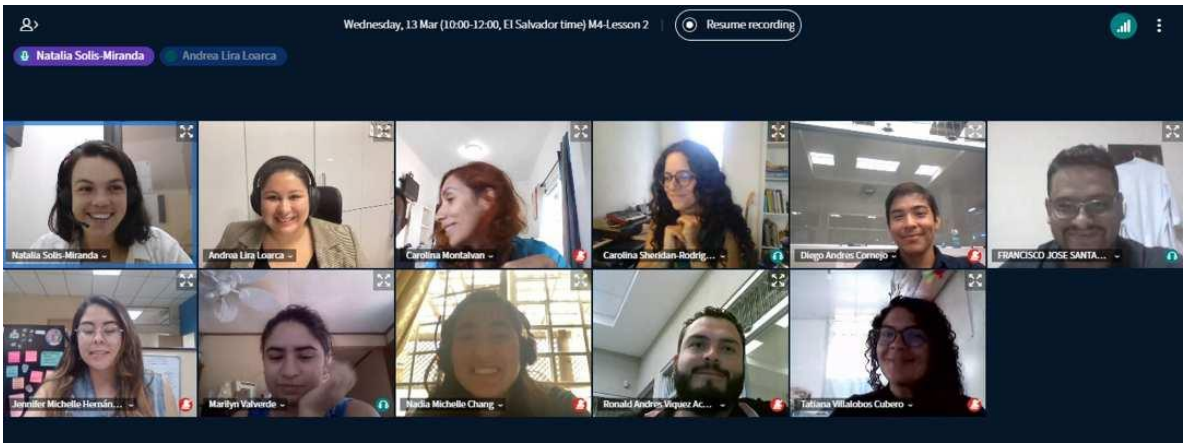

Figure 4. Participants during the life session in Module 4

Another significant benefit of using Xarray was its automated plotting capabilities. Participants experienced how Xarray simplifies the plotting process by providing sensible axis labels, which streamlines the visualization of complex datasets. They also engaged in simple plotting tasks and learned how to save their processed data for future use. Through these activities, participants gained a deeper understanding of how Xarray's labeled data structures facilitate efficient and intuitive geospatial data analysis.

Learning outcomes:

- Geospatial Data: Participants acquired skills to work with geospatial data, including longitude, latitude, and time dimensions.
- Simple Plotting with Xarray: Participants learned to create basic plots using Xarray, enabling effective visualization of data.
- Save Data: Participants gained the ability to save processed data for future use and analysis.

III. Course Feedback

Fifteen participants provided feedback on the course, focusing on four key areas: the clarity of objectives and schedule, the information available on the OTGA platform, the quality of the course content, and the performance of the facilitators and instructors.

A) Course objectives and time:

Participants suggested extending the course duration and improving the planning of the initial setup phase. They recommended conducting the installation asynchronously, with a dedicated session to address any issues. Some found time zone differences challenging and felt that deadlines were unnecessary, as participants should be able to progress at their own pace. Others mentioned the need to rewatch videos for better understanding and expressed difficulty collaborating with team members due to tight submission deadlines. More detailed feedback on course objectives and timing is available in Figure 5.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The learning outcomes are clearly stated, measurable and are appropriate to the level of the target audience	3	10	1	0	1
Expectations for student learning were clearly defined	5	9	0	1	0
The course covered the content announced/expected	8	7	0	0	0
I achieved the course stated aims and objectives	7	6	1	0	1

Figure 5. Participants' feedback on the course objectives and timeline.

B) Platform resources and materials

Participants were generally satisfied with the communication, but noted some issues with resource availability. One module's PDF was missing initially, and files were sometimes unavailable or not updated in a timely manner, leading to requests for alternative links. The quality of explanations in the materials varied, with some appreciating the detailed approach in certain PDFs while others felt that more explanation was needed in others. Further feedback on the resources and material provided in the platform is provided in Figure 6.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The course platform is well-organized and easy to navigate	8	7	0	0	0
All course pages are readable and visually consistent	6	6	3	0	0
All course pages are functionally consistent and communicate course information clearly and in sequential order throughout the course	7	5	3	0	0
Multimedia files are clear, adequate, compatible with multiple operating systems and requires only a free, standard, and easily downloadable plug-in	8	5	2	0	0

Figure 6. Participants' feedback on the course platform and resources.

C) Course content

Participants suggested incorporating review sessions with instructors after submitting assignments to address questions and improve understanding. While some felt the initial practice time was too short, the time allocated for the final report was considered sufficient. More detailed feedback is provided in Figure 7.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Multimedia files are clear, adequate, compatible with multiple operating systems and requires only a free, standard, and easily downloadable plug-in	8	5	2	0	0
The course content is appropriate to the goals of the course	5	8	2	0	0
The learning activities helped me achieve the stated learning outcomes	8	6	1	0	0
The relationship between completing learning activities and meeting the learning outcomes was clearly explained	6	7	0	2	0
During this course, I was often engaged in learning activities such as discussion forums, wikis, chat, projects, group work, etc	6	5	3	1	0

Figure 7. Participants' feedback on the course content

D) Course facilitators and instructors

Participants appreciated the facilitators' ability to simplify the process and their patience. However, some felt there was limited time for personalized support. They also found the WhatsApp group highly useful for additional assistance. More detailed feedback from the participants is provided in Figure 8.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
	Very good	Good	Neutral	Poor	Very Poor
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	4	9	0	2	0
During this course, I had the opportunity to interact with other learners	6	6	2	1	0
How do you rate the response from the course facilitators to your questions or issues raised?	7	7	1	0	0
How do you rate the response from the OTGA Secretariat to your questions or issues raised regarding the use of the Ocean?	8	5	0	0	1

Figure 8. Participants' feedback on the course facilitators and instructors

E) Overall comments from the participants about the course:

In addition to feedback on specific aspects of the course, participants also shared open comments on their overall experience and the key benefits they gained. Below are some of the opinions expressed by participants:

Participant 1:

"The step by step of phyton, the first lesson where crucial for a good development of the rest of the exercises and I have to recognize that Leonardo was very patience with all the students doubts with the help of Natalia and Xochilt that were always pendant of our questions too."

Participant 2:

"The hands-on practical exercises were incredibly beneficial. They provided a real-world application of the concepts taught in the course, allowing me to understanding. The

comprehensive explanations provided by the instructors were clear and easy to follow, making complex topics more accessible."

Participant 3:

"We would share the screen to solve any doubt gave us more confidence to use python and actually try to code."

F) Suggestions for improvement from the participants:

In addition to feedback on specific aspects of the course, participants also shared open comments on their overall experience and the key benefits they gained. Below are some of the opinions expressed by participants.

Participant 1:

"I suggest that for this kind of courses, the platform can have pre-recorded videos so for things like installing python, we could have had watch a video of how to do it, and then in the synchronic video we asked our doubts or solve any problem. I suggest this because at the beginning of the course we lost valuable time that could have been used better."

Participant 2:

"For all that phyton implies I think it was really good for the beginning I would recommend a second training course with focus on specific topics for example how to use the insitu data in Central America with the global modeling on climate change."

Participant 3:

"Incorporating more real-life case studies or projects would further enhance the practical application of the skills learned. Providing additional resources or recommended readings for further exploration of specific topics covered in the course would be beneficial for those who want to delve deeper into certain areas."

G) Comparing online learning and in-person courses: participant opinions

Participants shared their preferences regarding course format, with some expressing a clear preference for online courses. They appreciated the flexibility and accessibility of online learning, suggesting that it better suited their needs compared to in-person formats.

Participant 1:

"Online learning offers flexibility and accessibility, allowing individuals to learn at their own pace and from anywhere with an internet connection. It also provides access to a wide range of courses and resources that may not be available locally. However, effective online learning requires self-discipline and motivation to stay engaged and actively participate in the learning process."

Participant 2:

"Good sometimes it can be frustrating especially with programming but you always find a way, it is good for you to solve things, the recordings are always a good help and resources

I appreciate them, and online training expands the opportunity for people from different latitudes to learn and interact”

Participant 3:

“Personally, I would prefer to take this course online. The flexibility of online learning allows me to balance my other commitments while still having access to high-quality instruction and resources. Additionally, online courses often offer opportunities for asynchronous learning, enabling me to review materials at my own pace and revisit topics as needed.”

Others preferred in-person courses, though they acknowledged the potential challenges associated with this format:

Participant 1:

“[I would prefer to take this course] in a classroom definitely, because you have to watch the things that the profesors is doing in his or her computer and then do it in your computer, so you need at least two monitors to acomplish the work and to follow the rythm of the class and also to leverage the time that you are investing in the class”

Participant 2:

“[I would prefer to take this course] in the classroom for all the basic thing you need to learn but always with some help of videos that remain you the steps to follow”

Participant 3:

“Could be good [in the class], but usually due to costs it would be even shorter. Also, long distance pushes us students to find our solutions to problems encountered.

IV. Training schedule

Module	Date	Lesson	Activity	Time
Module 1			**Fundamental concepts in Python programming**	
	26th Feb	**Lesson 1**	- Introduction and housekeeping	10 min
			- Setting up environment with Anaconda and Python installation	45 min
			- Break	10 min
			- **Python basics** - Library and uses of library - Array operations	45 min
	27th Feb		- Students carry out own practice with the module's or own data	
	28th Feb	**Lesson 2**	- Analysing pigment data - Visualization	55 min
			- Break	10 min
			- Tuple, list, and dictionaries - Exercise	55 min
	29th Feb		- Students carry out own practice with the module's or own data	
Module 2			**Data visualization with Pandas**	
	1st March	**Lesson 1**	- Pandas - Objectives. - Storing pandas datasets - Selecting and plotting a column - Performing operations in pandas	60 min
			- Break	10 min
		Lesson 2	- Selecting and filtering data - Use resampling and rolling mean - Scatter plot	50 min
	4th March	**Lesson 3**	- Datetime	110 min
			- Break	10 min
Module 3			**Data analysis**	

	5th Mar	**Lesson 1**	- Libraries and datasets	110 min
			- Break	10 min
	6th Mar		- Students carry out own practice with the module's or own data	
	7th Mar	**Lesson 2**	- xarray basics - Opening a netcdf file - Preparing data - Processing and plotting data	110 min
			- Break	
	8th Mar		- Editing metadata - Saving as netcdf	55 min
		Lesson 3	- Break	10 min
			- numpy, pandas and xarray	55 min
Module 4			**Geospatial data analysis**	
	11th Mar	**Lesson 1**	- Introducing DataArray and Dataset - DataArray	55 min
			- Break	10 min
			- Sub setting and selection by coordinate values - Plotting with xarray	55 min
	12th Mar		- Students carry out own practice with the module's or own data	
	13th Mar	**Lesson 2**	- Arithmetic operations and aggregation methods	55 min
			- Break	10 min
			- GroupBy:Split, Apply, Combine	55 min
	14th Mar	**Lesson 3**	- Other high-level computation functionality and masking data - Masking Data	110 min
			- Break	10 min
	15th Mar		- Students carry out own practice with the module's or own data	

	17th Mar		- Students hand in project reports	
	19th Mar		- Students' presentation and course wrap up	

V. Course organizers and lecturers

The training course was designed by Gabriel Akoko Juma and Xochitl Édua Elías Ilosvay. It was delivered by Leonardo Mauricio Alvarado Bonilla, Rieke Schaefer, and Andrea Lira Loarca. Natalia Solís-Miranda provided additional support throughout the course. Below are brief bios of all contributors.

 Gabriel Akoko Juma *Coordinator*: Blue CaD Project	Bio: • Ph.D. 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) Role: • Development of course concept and work plan • Review of training modules on OGTA Platform • Course co-hosting • Evaluation of assignments • Feedback writing • Course coordinator
 Xochitl Édua Elías Ilosvay *Coordinator*: ECOP Central America	Bio: • PhD student in Marine Science, Technology, and Management at Future Oceans Lab, University of Vigo - Spain • Holds a MSc in Marine Biology from University of Bremen (Germany) Role: • Development of course concept and work plan

	• Review of training modules on OGTA Platform • Course moderation and co-hosting • Evaluation of assignments • Report writing • Course coordinator
 Natalia Solís-Miranda *Course support:* ECOP Programme – Training and Mentoring	Bio: • Ocean Sustainability Consultant • MSc Erasmus Mundus on Maritime Spatial Planning and BSc in Marine Biology by the National University of Costa Rica Role: • Course moderation and co-hosting • Feedback writing
OTG@ OceanTeacher Global Academy OTGA Secretariat	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. Role: • Hosting of e-learning platform • Provision of pedagogic support • Provision of administrative support
 Leonardo Mauricio Alvarado Bonilla *Post-doctoral Researcher* Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (Germany)	Bio: • PhD in Environmental Physics at the Institute of Environmental Physics, University of Bremen (Germany) • Licenciate in Physics at the University of El Salvador Role: • Lecturer Module 1 • Module evaluation

 Rieke Schaefer *Researcher,* Physikalisch-Technische Bundesanstalt (PTB) - Germany	Bio: • PhD student at GEOMAR (Kiel, Germany) and Physikalisch-Technische Bundesanstalt (Braunschweig, Germany) working on uncertainties in marine pH measurements • Volunteer for Climatematch Academy Role: • Lecturer Module 2 and 3 • Module evaluation
 Andrea Lira Loarca University of Genova - Italy	Bio: • Holds a joint PhD in Biochemical Fluid Dynamics by the University of Granada (Spain), and in Engineering and Architecture by the University of Parma (Italy) • Holds a MSc in Environmental Hydraulics from the University of Granada (Spain) • Co-coordinator of ECOP Guatemala Role: • Lecturer Module 4