D6.1 – Report on Summer School and Its Material

Luciana Pedro

Work Package:	Work Package 4 Exploitability
Responsible Institution:	University of Reading (UREAD)
Contributing Institutions:	· Barcelona Supercomputing Center (BSC) · Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (Cerfacs) · Centro Euro-Mediterranean sui Cambiamenti Climatici (CMCC) · Federal Office of Meteorology and Climatology (MeteoSwiss) · DataDirect Networks (DDN) · Deutsches Klimarechenzentrum GmbH (DKRZ) · European Centre for Medium-Range Weather Forecasts (ECMWF) · National Centre for Atmospheric Science (NCAS) · Science and Technology Facilities Council (STFC) · Seagate Systems UK Limited
Date of Submission:	October 14, 2020

Contents

1 Introduction

1.1 Introduction

This is a report covering the teaching material and discussing the result of the conducted survey regarding the quality and effectiveness of the summer school and the individual OER materials.

Chapter 2 looks back to the task proposal as stated in the project **Excellence in Simulation of Weather and Climate in Europe, Phase 2 (ESiWACE2)**. The topics are described in Sections 2.2 to 2.7 and the summer school proposal can be found in Section 2.8.

From 24 to 28 August 2020, we realised the Summer School on Effective HPC for Climate and Weather 2020 (Chapter 3). The event was entirely online and provided five days of learning and debating HPC systems and climate and weather applications.

We received 162 applications, and 54 participants who attended more than 70% of the event were granted a certificate of participation. The countries with the most participants were: UK (19.8%), India (12.3%), and Germany (11.1%). In total, we had 38 countries represented in the event. All applicants were added to a mailing list to facilitate communication between them and the organisers, and to inform the participants about related events afterwards.

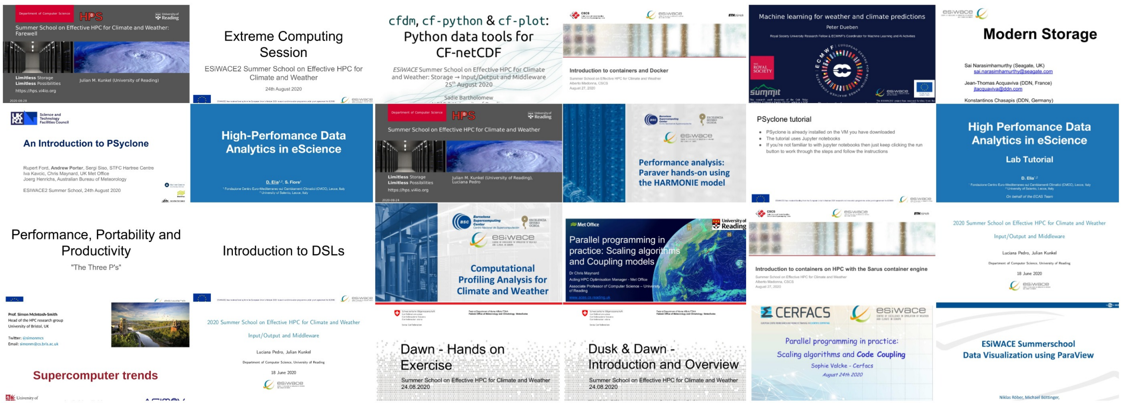


Table 1.1: Snapshot of first-page presentations.

The format of the summer school was slightly changed to accommodate the current situation in the world. From Monday to Thursday, a topical session consisting of an academic lecture and hands-on/lab practicals was delivered. This made in total 18 hours of lecture and 10 hours of hands-on training. There was also a virtual visit to ECMWF, covering tours to the Computer Hall and to the Weather Room. The sessions covered the following topics:

Monday Extreme-Scale Computation (STFC, MeteoSwiss) and Parallel Programming in Practice (CERFACS, UREAD).

Tuesday Modern Storage (Seagate, DDN) and Input/Output and Middleware (UREAD).

Wednesday Machine Learning (ECMWF) and High-Performance Data Analytics and Visualisation (CMCC, DKRZ).

Thursday Performance Analysis (BSC) and Containers (ETH Zürich, NCAS).

On Friday, a different format was administered. We organised extra sessions in a Q&A format for each lab that happened during the week, and we finished the program with a keynote talk. Daniel Klocke, from the German Meteorological Service (DWD), presented his work on global storm- and ocean eddy-resolving coupled climate simulations and the contributions of projects DYAMOND and DYAMOND2.

All sessions were recorded and the recordings, together with the presentations and material for each session, are available on the main summer school webpage. The links are also available throughout this document, in their respective sessions, and in Chapter 5.

After the event, we asked the participants to fill out a survey. Here are some of the main numbers and comments about the summer school:

- 38.5% said the summer school was better than they had expected and 46.2% said it was as good as they had expected;
- 92.3% said they had an excellent or good experience;
- Selected quotes from attendees about what they most liked:
  - "The combination of learning the theory, getting some hands-on it, and working with the learned things";
  - "The interdisciplinary spirit of the presentations and profound knowledge of people invited to talk";
  - "High-quality presentations and I can revisit it when I need on YouTube".

Chapter 4 is about the summer school 2021. To be done.

2 The Proposal

2.1 Introduction

The main targets of ESiWACE2 proposal concerning training are:

- In collaboration with PRACE, address the skills gap in computational science in the targeted domain by specialised training and capacity building measures to develop the human capital resources for increased adoption of advanced HPC in the industry (including SMEs) and academia.
- Crossing the chasm in education between computational science and weather and climate modelling concerning extreme-scale computing by supporting and organising specialised training, workshops and summer schools on HPC software engineering, I/O, HPDA, containerisation, DSLs, etc., with a focus on (pre-)exascale challenges for weather and climate applications (WP6).
- Synchronisation of training with PRACE via Coordination and Support Action (CSA) for Centre of Excellence (CoE) (WP6, WP7).
- Interactions with the European HPC ecosystem (WP6).
- Organise two summer schools to train scientists in the efficient usage of supercomputers for high-resolution Earth system modelling.

To achieve these goals, ESiWACE2 has designed Task 6.2:

Task 6.2: Training and schools on HPC software engineering, methods and tools

[Lead: CERFACS. Partners: ECMWF, BSC, METO, CMCC, UREAD, STFC, SEAGATE, ETH Zurich, MeteoSwiss, DDN, CNRS-IPSL]

This task organises the training offered in different HPC areas of ESiWACE2 pre-exascale expertise, i.e., I/O, computation (DSLs, C++ and coupling software), data analytics, and containerisation. It also organises two summer schools to train scientists in these matters.

For all training, we embrace the creation of online digital media as Open Educational Resources (OER) material that will be available under a permissive CC-by licence and can be reused by teachers and researchers outside the consortium. The resulting material will be presented during the summer schools.

All training also propose online sessions or face-to-face sessions at the partner's institution (Deliverable D6.2). Additionally, we will explore the collaboration with the international effort of the Universität Hamburg to establish an HPC certificate programme for scientists.

These training will result in a larger number of scientists and engineers with higher qualifications in the use and optimisation of climate and weather applications on tier-0 machines, increasing HPC awareness in that community. Ultimately, training will allow climate and weather research to be more productive, favouring European scientific excellence in that field.

In the next sections, we introduce the specific topics that will be covered in training with their description of how the training will happen individually and all together in two summer schools.

2.2 Training on I/O and HPC Awareness

This task covers training on the I/O hardware and software stack. Delivered material will target scientists accessing data on storage and developers that aim to utilise storage APIs efficiently.

We will create descriptive course material covering the co-designed software stack of WP4 as a standardised and efficient platform for I/O (Cylc, XIOS, ESDM, NetCDF, HDF5). We will explore how the analysis stack fits into it (in-depth analysis workflows are part of training on high-performance data analytics, see below). Furthermore, we will cover training on the underlying storage APIs and file systems (e.g., Mero/Clovis, Lustre), efficient creation of I/O dominated workflows, performance and efficiency considerations when dealing with file systems, NVM, tape or object storage, and cost-considerations in storage architectures.

2.3 Training on DSL

The rapid changes in the multiple supercomputing architectures used to run weather and climate codes and the different programming models used seriously affect the development productivity and the ability to retain a single source code running efficiently everywhere. Domain-specific languages provide a solution to portability of these codes.

In this training, we provide insights on DSLs considered in ESiWACE2 (PSyclone, CLAW and GridTools ecosystem) and demonstrate how to apply them to weather and climate models. Participants will theoretically and practically learn how to use the DSL languages to implement PDE operators.

During a hands-on session, participants will be encouraged to implement some of the benchmark models defined by WP2 using DSLs and to build their own toy models, followed by an in-depth evaluation of generated optimised implementation and performance benefits.

2.4 Training on C++ for HPC

ETH Zurich organises an advanced course for C++ in HPC as part of its regular training. This course will present advanced tools for effective C++ programming in the context of HPC, such as generic programming techniques, API development, specific C++-11/14 constructs, rather than treating parallel programming primitives, such as OpenMP or MPI, a subject of widely available courses.

2.5 Online Training Course for Code Coupling with OASIS3-MCT

The objectives of this task are to create an online course that teaches the participants basic concepts in code coupling, focussing on the ocean-atmosphere context, and help them learn how to use OASIS3-MCT.

This training course is for engineers, physicists, and computer scientists wishing to use this code coupling software in their own coupled model. The material will cover theoretical concepts about code coupling, instructions on how to download, install and compile OASIS3-MCT and finally implementation of the coupling between two toy models in a hands-on tutorial session.

2.6 Training on High-Performance Data Analytics

The objective of this task is to increase scientists' expertise on scientific data analysis at scale applied to climate and weather domains, using high-performance data analytics tools available from the open-source market (e.g., Ophidia). It will address several vertical training levels (e.g., intermediate, expert) from different horizontal perspectives (e.g., end-user, developer, administrator), covering from simple analytics tasks to workflows and applications (e.g., Python-based), and providing best practices and guidelines on dealing with massive scientific datasets on HPC architectures.

The delivered material, using the material mentioned above on I/O and HPC awareness as background, will include theoretical and practical aspects of big data (both introductory and advanced topics), high-performance data management (including performance and optimisation aspects), scientific data analytics at scale, and analytics workflows.

2.7 Training on Docker Containerisation

Docker is an open software technology for developers to build, ship, and run distributed applications in containers. This aim is achieved by describing the software environment and build instructions allowing the creation of a container that can be executed on different environments and does not depend on system-wide software packages.

Consisting of a rich ecosystem, Docker enables applications to be quickly assembled from components and eliminates the friction between development, QA, and production environments. As a result, the software can ship faster and run the same unchanged stack on laptops, data centres, and virtual machines. With Docker, developers can build any application in any language using any toolchain; *dockerised* applications are entirely portable and can run anywhere.

A recording of the material presented at the summer schools will be made available online to reach a wider audience.

2.8 Summer School in HPC for Weather and Climate

In this task, we will organise two summer schools for Earth system scientists, including PhD students and postdocs, covering the HPC aspects to run scientific workflows on a large scale HPC environments efficiently. A five-day summer school will take place in years two and three of the project and support about 30 students each year and fund their attendance. The selection procedure for subsidising students will be established six months before the summer school taking into account the experience gained with summer schools held within the IS-ENES projects. It will prioritise diversity and internationality and fund students depending on the financial support possible by their home institution.

Theoretical courses and hands-on sessions will be organised on I/O, HPC computation (DSLs, C++ and coupling software), data analytics, containers, efficient programming and usage of Data Centre resources, using the teaching material created in this Work Package (WP). We will also invite external experts to give tutorials about selected topics. Selected presentations will be recorded and made available online after the summer school for a wider audience.

A survey will be conducted to evaluate the feedback of participants about the content of the tutorial itself (D6.1[1]).

[1] This document.

3 Summer School on Effective HPC for Climate and Weather – 2020

3.1 Introduction

The Summer School on Effective HPC for Climate and Weather has brought together young researchers and software engineers interested in the current technological developments in the field of climate modelling. Together, they explored hot topics in high-performance computing and climate and weather applications.

Making effective use of HPC environments becomes increasingly challenging for PhD students and young researchers. As their primary intent is to generate insight, they often struggle with the technical nature of the tools and environments that enable their computer-aided research: computation, integration, and analysis of relevant data. The scope of this event was the training of young researchers and software engineers in methods, tools, and theoretical knowledge to make effective use of HPC environments and generate insights for the field and their research.

While the school aimed to prepare the attendees for large scale simulation runs and data processing, it also covered a representative selection of modern concepts such as machine learning, domain-specific languages, containerisation, and analysis of climate/weather data using Python. The event also provided an outlook of challenges and strategies for HPC for climate and weather. Additionally, it was an opportunity to foster networking among scientists bringing together users of specific models and tools and enabling them to exchange their knowledge.

Due to COVID-19 restrictions, the Summer School 2020 was held as an online-only event from 24 to 28 of August, 2020. For more accessible communication, all participants were registered to our mailing list. For further information about the event, check the official webpage.

We provided a Tutorial to set up a Virtual Machine (VM) with Ubuntu and all the software for the training and lab sessions pre-installed. Participants were asked to set up the VM on their PC before the summer school to fully participate in the tutorial sessions and perform most of the training.

A certificate of attendance was provided to all participants that attend the event regularly. The Blackboard Collaborate stored the log details of every person connected to the event, and this information was used to select the participants that were worthy of a certificate.

The summer school supports the mission of the European Network for Earth System modelling (ENES). ENES is developing a common climate and Earth system modelling distributed research infrastructure in Europe integrating the European community on Earth's climate System Models (ESMs) and their hardware, software, and data environments.

The overarching goals of ENES are to:

- Further integrate the European climate modelling community;
- Ease the development of full ESMs;
- Foster the execution and exploitation of high-end simulations;
- Support the dissemination of model results and the interaction with the climate change impact community.

3.2 Structure

The ESiWACE Summer School was structured along with topical sessions in the morning/afternoon, from Monday to Thursday, a Q&A session on Friday morning and a Keynote Talk on Friday afternoon.

A topical session typically consists of an academic lecture, and it may contain hands-on/lab practicals. Experts in the respective field organised each of these sessions.

The hands-on tutorials/lab practicals worked as follows:

- A video tutorial was pre-recorded and provided an introduction/walk-through to the topic.
- The tutorials were scheduled for each session, but the usage was up to the participant's discretion.
- We provide a tutorial to set up a Virtual Machine (VM). The VM comes with Ubuntu and all the software for the training pre-installed. Attendees may install the VM on their PC to perform most of the training. Some training may be on a dedicated cluster. You should set up the VM before the summer school if you like to participate in the tutorial sessions.
- A lab practical may list additional exercises and suggestions for further learning.
- At the end of the day, a time slot for a Virtual Lab Session is given to allow independent and self-paced learning, and participants can decide what and when they want to engage.

On Friday, the dedicated Q&A slot per lab practical was scheduled. The session offered the participants the opportunity to contact the organiser of each hands-on session and ask questions regarding the topic and particularly regarding the tutorial and exercises.

Throughout the event, additional support was provided by the mailing list, in which participants may have posted questions and cooperated with other students and the organisers.

All sessions were recorded and the recordings, together with the presentations and material for each session, were made available in the main summer school webpage.

3.3 Topics and Sessions

The topics covered in the summer school were:

- Extreme-Scale Computation – Section 3.3.1
- Parallel Programming in Practice – Section 3.3.2
- Modern Storage – Section 3.3.3
- Input/Output and Middleware – Section 3.3.4
- Machine Learning – Section 3.3.5
- ECMWF – Virtual Visit – Section 3.3.6
- High-Performance Data Analytics and Visualisation – Section 3.3.7
- Performance Analysis – Section 3.3.8
- Containers – Section 3.3.9

On Friday, we organised a Q&A session a Keynote Talk:

- Q/A Session – Section 3.3.10
- Keynote Talk – Section 3.3.11

3.3.1 Extreme-Scale Computation

Abstract

This session will introduce the concept of extreme-scale computing with an explanation of the trends in the computer architectures that provide the underlying computing power. In particular, the increasing use of parallelism and heterogeneity in these architectures will be discussed.

A high-level overview will then be given of the performance, portability and productivity (3P's) requirements that Weather and Climate models have to run successfully on these computer architectures. It will be shown how current approaches can struggle to meet all three of these requirements.

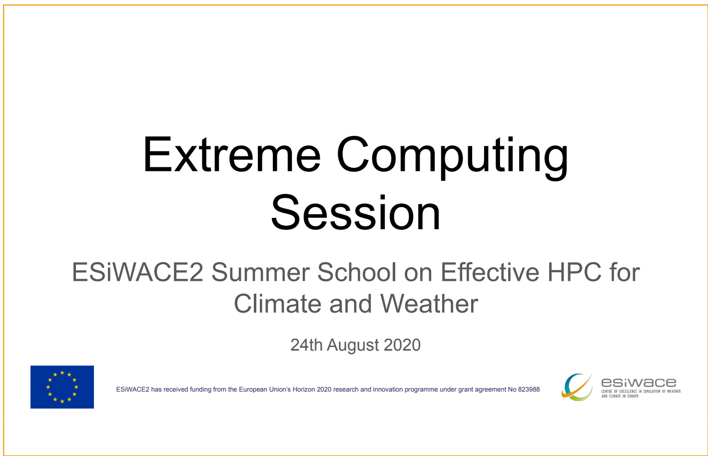
Lastly, a relatively new, Domain-Specific Language (DSL), approach to programming Weather and Climate models will be introduced with examples from two existing DSLs – DAWN and PSyclone. It will be shown that the DSL approach offers the possibility of supporting all three of the above requirements, by separating the implementation of the science code from its parallelisation and optimisation on the underlying computer architecture.

Learning Objectives

- Illustrate the complexity and diversity of extreme-scale computing on examples in climate and weather
- State the Performance, Portability and Productivity requirements of Weather and Climate models (3P's)
- Describe how Domain-Specific Languages (DSLs) can provide a solution to the problem of providing the 3P's
- Use PSyclone and Gridtools DSLs for small applications

Sessions

Extreme-Scale Computation		
Chair: Rupert Ford (STFC, UK) Chair: Carlos Osuna (MeteoSwiss, Switzerland)		
Time	Title	Speakers
09:00	Extreme Computing Session – Overview	Rupert Ford
09:05	Supercomputer Trends	Simon McIntosh-Smith
09:30	Performance, Portability and Productivity	Rupert Ford
10:00	Introduction to DSLs	Ben Weber and Rupert Ford
10:45	An Introduction to PSyclone	Andrew Porter
11:15	Dusk & Dawn – Introduction and Overview	Giacomo Serafini
11:45	Tutorial Introduction	Carlos Osuna
12:15	Lab Tutorial: Extreme-Scale Computation	
	PSyclone Tutorial	
	Dusk & Dawn Tutorial	



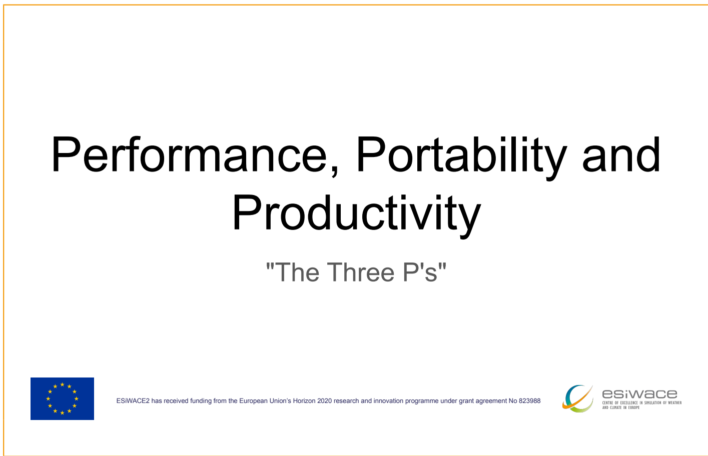
(slide above)	Slides	Video

Table 3.1: Extreme Computing Session – Overview

<table>
<tr><td>

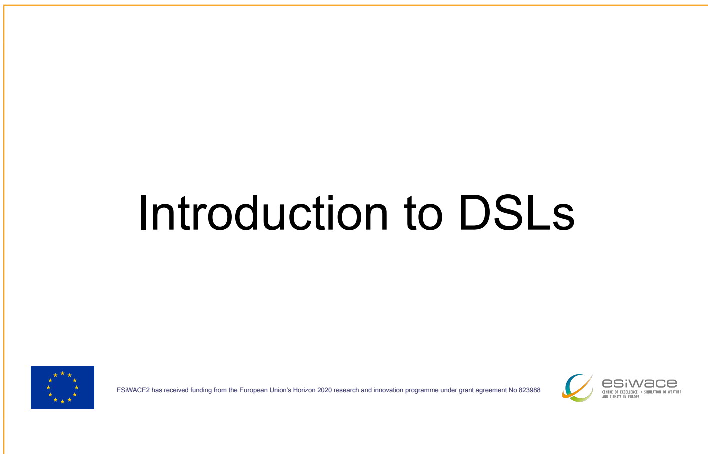

</td><td>Slides</td><td>Video</td></tr>
</table>

Table 3.2: Supercomputer Trends

<table>
<tr><td>


</td><td>Slides</td><td>Video</td></tr>
</table>

Table 3.3: Performance, Portability and Productivity

<table>
<tr><td>


</td><td>Slides</td><td>Video</td></tr>
</table>

Table 3.4: Introduction to DSLs

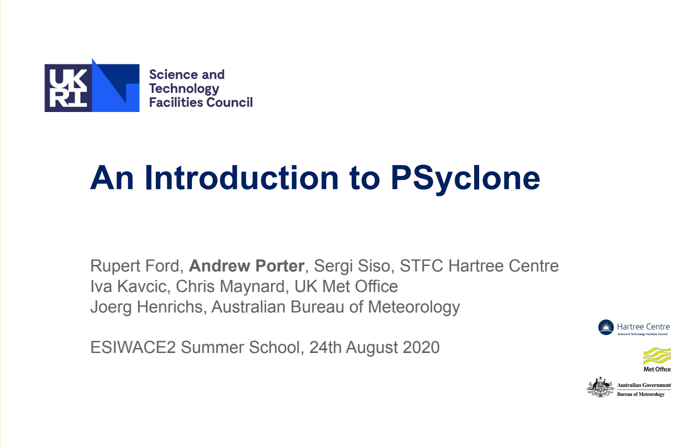
	Slides	Video

Table 3.5: An Introduction to PSyclone

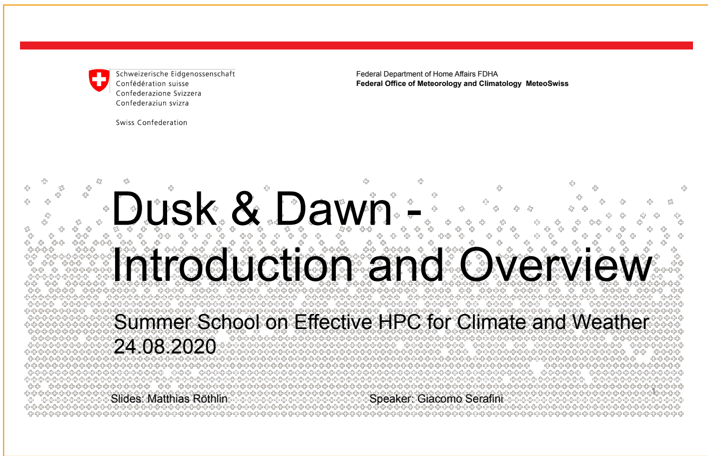
	Slides	Video

Table 3.6: Dusk & Dawn – Introduction and Overview

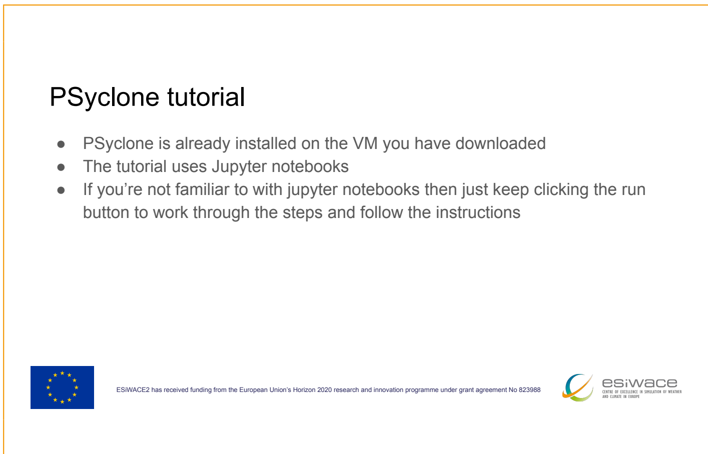
	Slides	Video

Table 3.7: PSyclone Tutorial

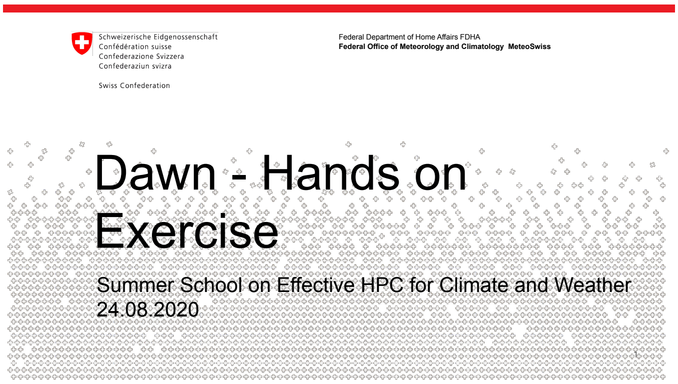
<table>
<tr>
<td>
</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.8: Dusk & Dawn Tutorial

Comments from the Survey

"I found it quite hard to follow, but I like the idea of different half an hour topics."

"The use of Anaconda for the PSyclone tutorial was a really helpful and interactive tool!"

3.3.2 Parallel Programming in Practice

Abstract

In this session, we will provide a global overview of how the main concepts of parallel programming are implemented in weather and climate codes. We will detail the different parallel programming models for distributed and shared memory systems and describe the resulting scalability of commonly-used algorithms implementing those models.

Particular attention will be devoted to specific features that may inhibit scaling and performance of weather and climate codes. This analysis will be done at the level of the code routine itself but also in the more general context of code coupling, the latter being a specific implementation of coarse grain parallelism.

Learning Objectives

- Describe the scaling characteristics of commonly used algorithms in weather and climate models
- Discuss issues which may inhibit scaling and performance
- Classify programming models for distributed and shared memory systems
- Identify performance features and potential issues for computer processor architectures
- Describe the concepts of coupling software

- Classify coupling software implementations given their main characteristics
- Evaluate qualitatively the impact of different coupling configurations (sequential vs concurrent, multi vs mono-executable, etc.) on coupled model performance
- Describe the most used coupling software in climate and weather applications

Sessions

Parallel Programming in Practice		
Chair: Sophie Valcke (Cerfacs, France) Chair: Christopher Maynard (University of Reading, UK)		
Time	Title	Speakers
13:30	Scaling Algorithms	Christopher Maynard
15:15	Code Coupling	Sophie Valcke

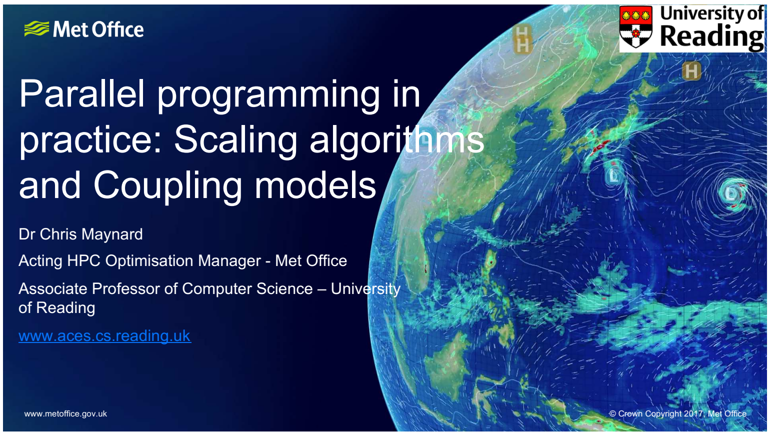
	Slides	Video

Table 3.9: Scaling Algorithms

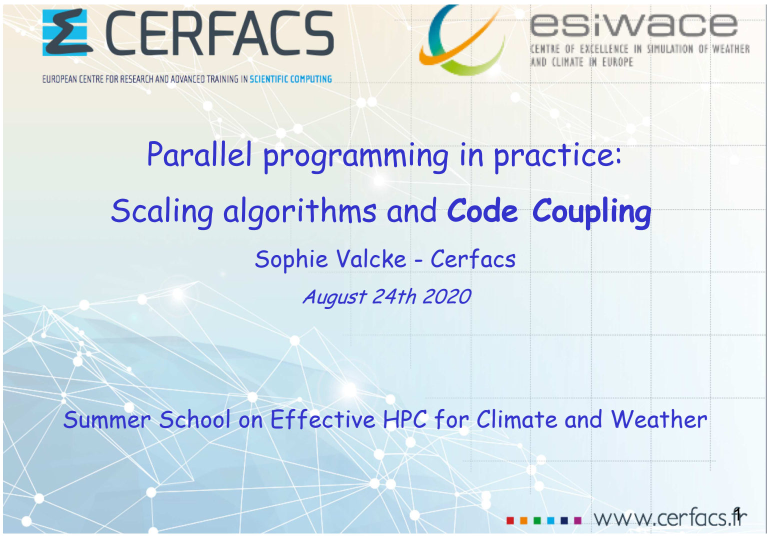
	Slides	Video

Table 3.10: Code Coupling

3.3.3 Modern Storage

Abstract

Learning Objectives

- Describe the architecture and architectural implications of modern storage architectures and object stores suitable for extreme-scale computing
- Discuss the storage stack with its semantics and potential performance implications on different levels: in particular POSIX vs MPI-IO vs NetCDF and high-level I/O middleware
- Execute the Darshan tool to identify I/O patterns and assess the performance
- Apply benchmarking tools to assess the performance

Sessions

Modern Storage		
Chair: Sai Narasimhamurthy (Seagate, UK) Chair: Jean-Thomas Acquaviva (DDN, France)		
Time	Title	Speakers
09:00	Modern Storage	Sai Narasimhamurthy, Konstantinos Chasapis
10:45	Lab Tutorial: Modern Storage	
	Darshan Demonstration – Hands-on Session	Konstantinos Chasapis
	Installing Darshan for I/O Performance Analysis	Konstantinos Chasapis
	Introduction to Using Darshan for I/O Performance Analysis	Konstantinos Chasapis

	Slides	Video

Table 3.11: Modern Storage

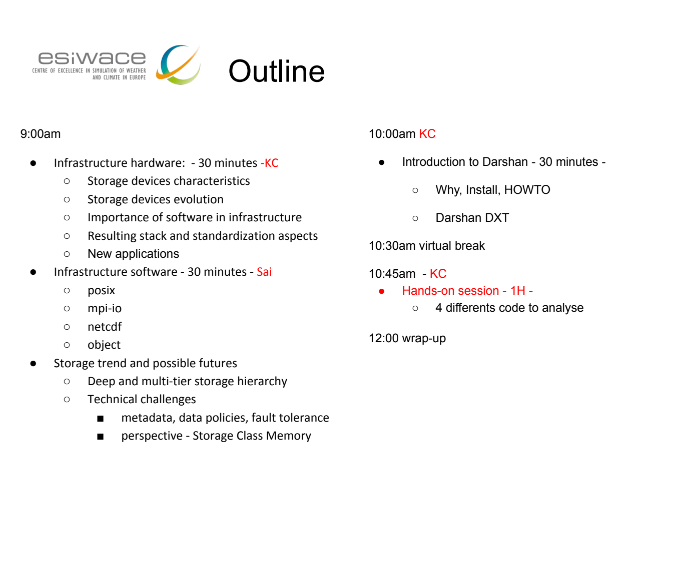
	Slides	Video

Table 3.12: Darshan Demonstration – Hands-on Session

	Slides	Video

Table 3.13: Installing Darshan for I/O Performance Analysis

	Slides	Video

Table 3.14: Introduction to Using Darshan for I/O Performance Analysis

Comments from the Survey

"All the session was overall interesting. Konstantinos' presentation was great; the constant interruptions by his colleague, however, was a little disturbing."

3.3.4 Input/Output and Middleware

Abstract

Climate and weather research is typically data-intensive, and applications must utilise input/output efficiently. Often, a user struggles to assess observed performance leading to superflux attempts to tune the application and optimise performance in a wrong layer of the stack.

The content of this session is twofold. Firstly, we discuss storage layers focusing on the NetCDF middleware and provide a performance model that aids users to identify inefficient I/O. Secondly, we introduce the NetCDF Climate and Forecast (CF) conventions that are often used as a standard to exchange data.

Learning Objectives

- Discuss challenges for data-driven research
- Describe the role of middleware and file formats
- Identify typical I/O performance issues and their causes
- Apply performance models to assess and optimise I/O performance

- Design a data model for NetCDF/CF
- Analyse, manipulate and visualise NetCDF data
- Execute programs in C and Python that read and write NetCDF files in a metadata-aware manner
- Implement an application that utilises parallel I/O to store and analyse data
- Describe ongoing research activities in high-performance storage

Sessions

Input/Output and Middleware		
Chair: Julian Kunkel (University of Reading, UK) Chair: Luciana Pedro (University of Reading, UK)		
Chair: Sadie Bartholomew (University of Reading, UK)		
Time	Title	Speakers
13:30	Input/Output and Middleware	Luciana Pedro
15:15	Python Data Tools for CF-netCDF	Sadie Bartholomew
16:45	Lab Tutorial: Input/Output and Middleware	
	An Introduction to NetCDF Using C Language	Luciana Pedro
	CF-NetCDF with cfdm, cf-python and cf-plot	Sadie Bartholomew

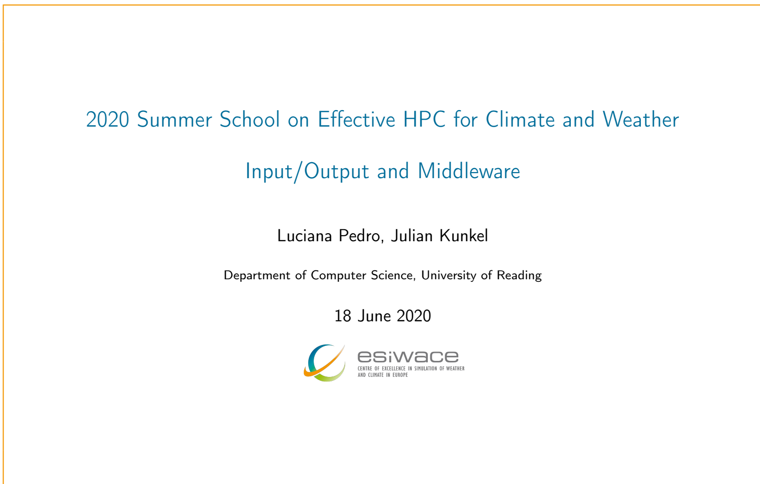
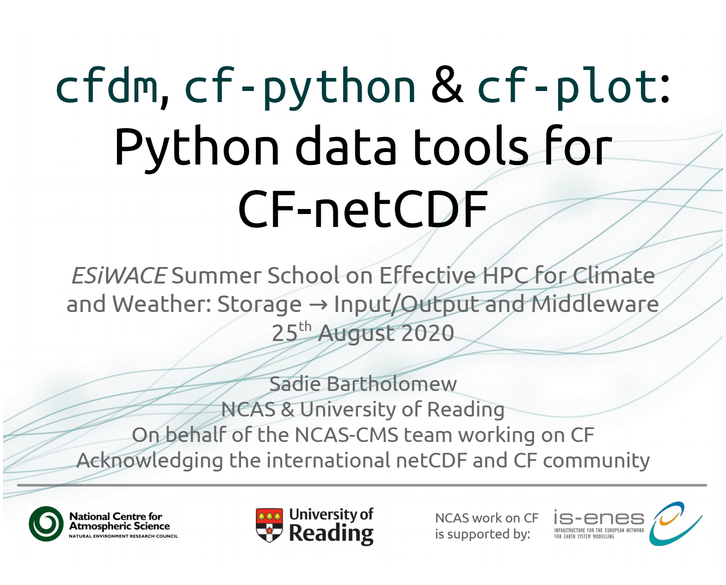
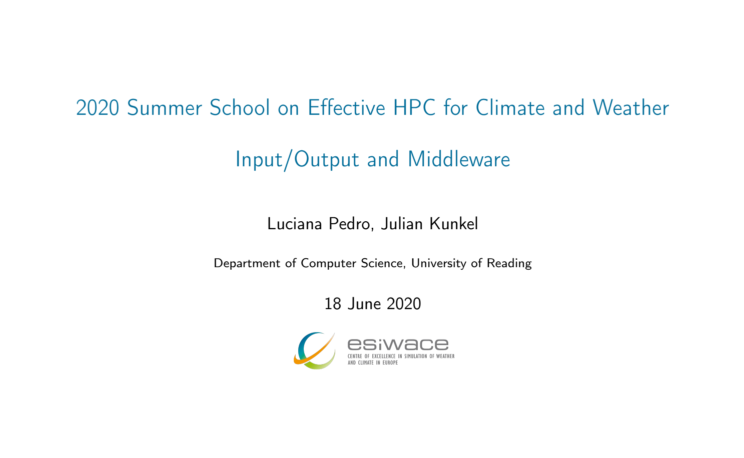
	Slides	Video

Table 3.15: Input/Output and Middleware

<table>
<tr>
<td>


</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.16: Python Data Tools for CF-netCDF

<table>
<tr>
<td>


</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.17: An Introduction to NetCDF Using C Language

Course order

The recommended order in which to work through the full course Notebooks is as follows, where all paths given are relative to `full_course/notebooks/` :

	Topic	Format	Notebook name & location
1.	Introduction	Overview & walkthrough	`introduction.ipynb`
2.	Reading and manipulating data	Walkthrough	`cf_python.ipynb`
3.	Reading and manipulating data	Exercises	`exercises/cf_python_exercise_*.ipynb` [`*` =1-5]
4.	Contour plots	Walkthrough	`contour_plots.ipynb`
5.	Contour plots	Exercises	`exercises/contour_exercise_*.ipynb` [`*` =1-2]
6.	Regridding data	Walkthrough	`cf-python_regridding.ipynb`
7.	Regridding data	Exercises	`exercises/regridding_exercise_*.ipynb` [`*` =1-2]
8.	Vector and line plots	Walkthrough	`vector_line_significance.ipynb`
9.	Wrap up	Summary	`wrap_up.ipynb`

Slides | Video

Table 3.18: CF-NetCDF with cfdm, cf-python and cf-plot

Comments from the Survey

“I really liked both presentations.”

“I enjoyed a lot this session, was really descriptive! I would have liked to have this session before the Modern Storage session to be able to have a better understanding.”

“The lab session was awesome. I learned so much.”

“Sadie gave an exceptional lab session of the Python data tools for CF-netCDF. A lot of great tools and libraries to use.”

3.3.5 Machine Learning

Abstract

(1) Predicting weather and climate require modelling the Earth System – a huge system that consists of many individual components that show chaotic behaviour and for which conventional tools often struggle to provide satisfying results.

(2) A huge amount of data of the Earth System is available from both observations and modelling.

(3) Machine learning methods allow learning complex non-linear behaviour from data if enough data is available and to apply the learned tools efficiently on modern supercomputers.

If you combine (1), (2) and (3), it is easy to see that there are a large number of potential application areas for machine learning in weather and climate science that are currently explored. However, whether these approaches will succeed is still unclear as there are also a number of challenges for the application of machine learning tools in weather predictions.

This talk will provide an introduction to machine learning, outline how to apply machine learning in Earth System modelling, show examples for the application of machine learning throughout the weather and climate modelling workflow, and discuss the challenges that will need to be tackled.

Learning Objectives

- Describe the relevance of Machine Learning and its application to judge why there is such a hype around the topic at the moment
- Explore how machine learning can be used in weather and climate modelling
- List several specific examples for the use of machine learning at ECMWF
- Discuss challenges for machine learning in weather and climate science

Sessions

Machine Learning		
Chair: Peter Dueben (ECMWF, UK)		
Time	Title	Speakers
09:30	Machine Learning for Weather and Climate Predictions	Peter Dueben

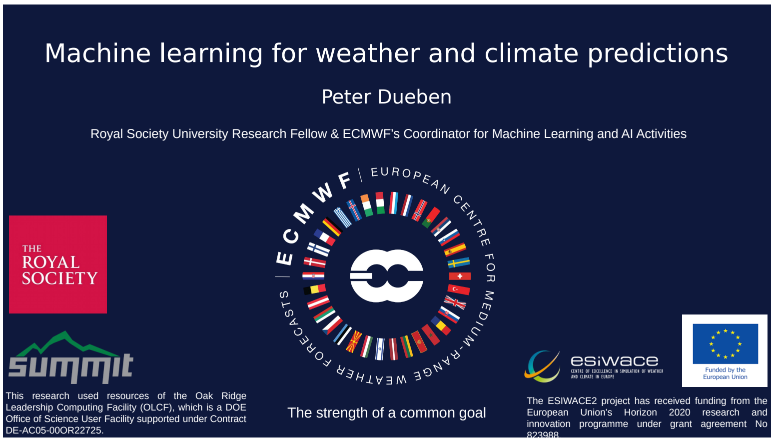
	Slides	Video

Table 3.19: Machine Learning for Weather and Climate Predictions

3.3.6 ECMWF – Virtual Visit

Computer Hall Tour

Learn about the performance and specifications of the ECMWF High-Performance Computing Facilities, and the way this supercomputer is used for operations, storage and research by ECMWF and its 34 Member & Co-operating States.

The presentation will include a video tour of the computing facilities currently located in our HQ in Reading and a preview of what the new data centre will look like when it opens in Bologna (Italy) next year.

Weather Room Tour

Learn about ECMWF Forecasting products and activities. A member of the ECMWF Forecasting team will introduce you to the maps, charts and plots that are produced daily in the "Weather Room" for weather prediction and analysis.

Sessions

ECMWF – Virtual Visit		
Chair: Peter Dueben (ECMWF, UK)		
Time	Title	Speakers
11:00	Introduction to ECMWF	Peter Dueben
11:30	Computer Hall Tour	Umberto Modigliani
12:00	Weather Room Tour	David Lavers

	Slides	Video

Table 3.20: Introduction to ECMWF

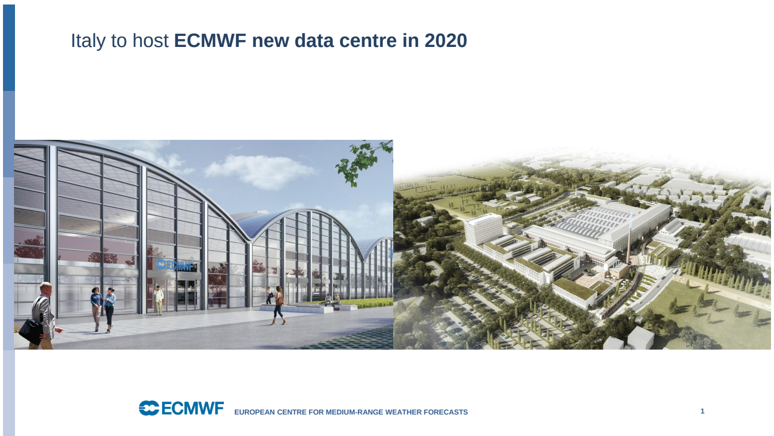
	Slides	Video

Table 3.21: Computer Hall Tour

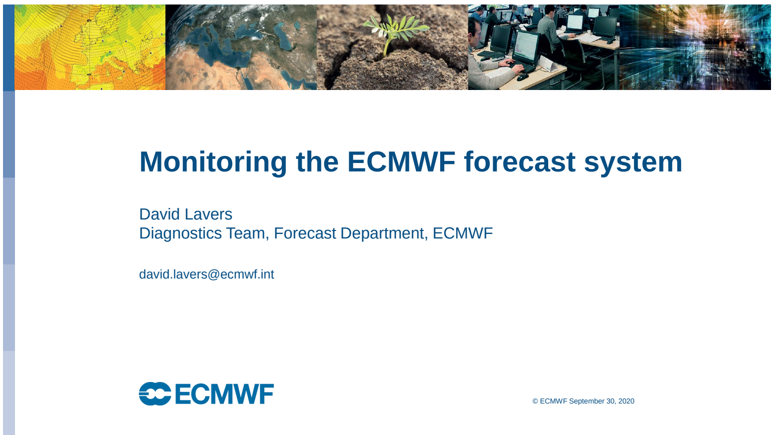
	Slides	Video

Table 3.22: Weather Room Tour

Comments from the Survey

"That was very interesting."

3.3.7 High-Performance Data Analytics and Visualisation

Abstract

Analysis and visualisation of scientific data, such as those in the field of climate and weather, requires solution capable of effectively and efficiently handling massive data.

In this session, we will discuss some of the main challenges concerning scientific data management and in particular, those related to data analytics and visualisation. Software

solutions for high-performance data analytics and visualisation, as well as examples of applications of these systems for real use cases in the climate and weather domain, will be presented.

The lab tutorial will provide a more practical introduction about some tools and modules for data analysis and how to apply these on climate data, as well as a walk-through of the VMI for the virtual lab.

Learning Objectives

- Discuss the main challenges of joining big data and HPC for scientific data management, in particular for data analytics and visualisation
- Put into action practical hints about some HPDA tools and their application to scientific data at scale
- Apply techniques and knowledge acquired during the course to real case studies in the weather and climate domain

Sessions

High-Performance Data Analytics and Visualisation		
Chair: Donatello Elia (CMCC, Italy) Chair: Niklas Röber (DKRZ, Germany)		
Time	Title	Speakers
13:30	Data Visualization Using ParaView	Niklas Röber
14:00	Hands-on: Data Visualization Using ParaView	
15:15	High-Performance Data Analytics and Visualisation	Donatello Elia, Sandro Fiore (CMCC, Italy)
16:45	Lab Tutorial: High-Performance Data Analytics and Visualisation	
	High-Performance Data Analytics and Visualisation	Donatello Elia

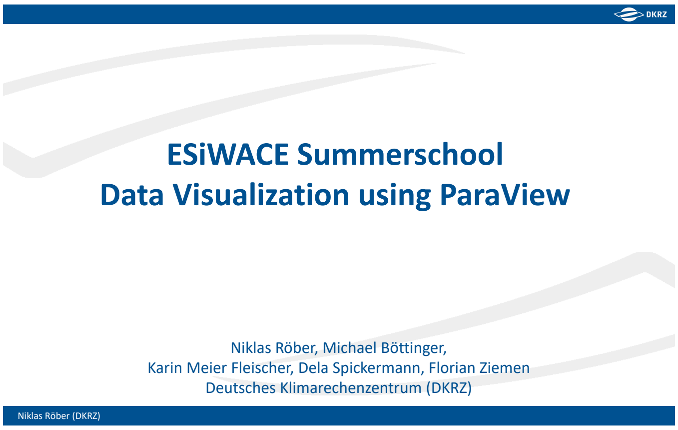
	Slides	Video

Table 3.23: Data Visualization Using ParaView

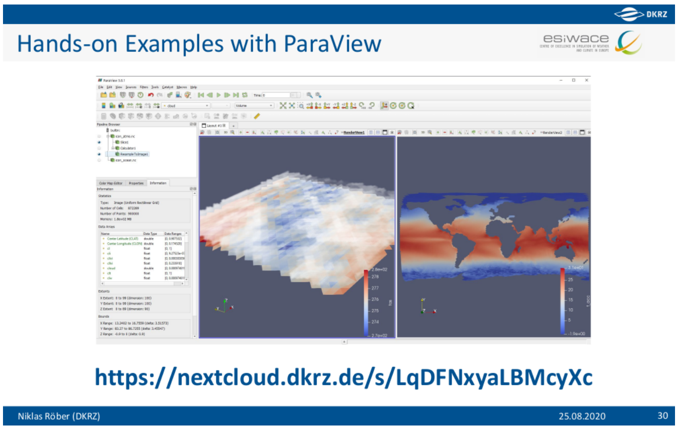
	Slides	Video

Table 3.24: Hands-on: Data Visualization Using ParaView

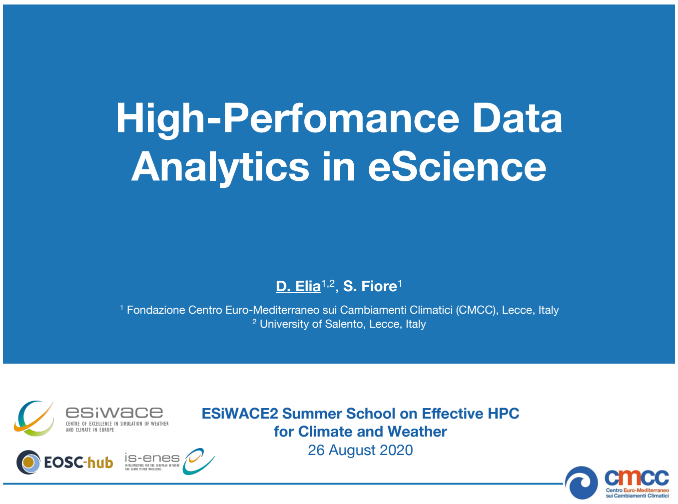
	Slides	Video

Table 3.25: High-Performance Data Analytics and Visualisation

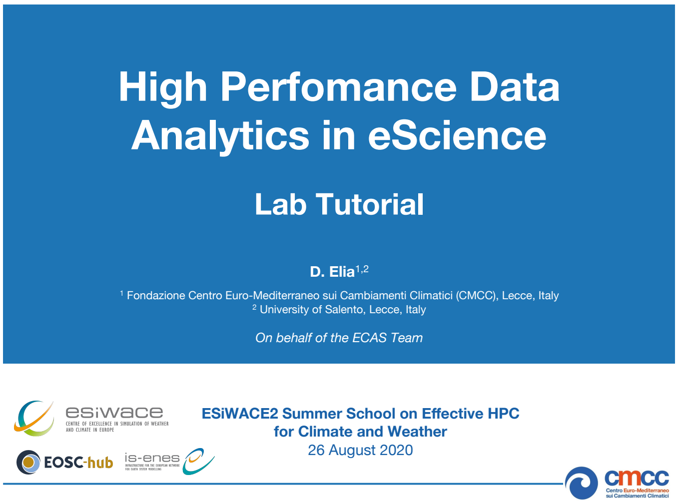
	Slides	Video

Table 3.26: High-Performance Data Analytics and Visualisation

Comments from the Survey

"This session was very good explained and very easy to follow."

3.3.8 Performance Analysis

Abstract

Learning Objectives

- Define performance analysis fundamentals (objectives, methods, metrics, hardware counters, etc.)
- Describe the BSC performance analysis tools suite (Extrae, Paraver, Dimemas)
- Interpret uses cases from Earth System Models (IFS, NEMO, etc.) that illustrate how to identify and solve performance issues
- Apply profiling techniques to identify performance bottlenecks in your code
- Summarise typical performance problems
- Discuss specific knowledge about performance analysis applied to earth system modelling

Sessions

Performance Analysis		
Chair: Mario C. Acosta (BSC, Spain) Chair: Xavier Yepes (BSC, Spain)		
Time	Title	Speakers
09:00	Computational Profiling Analysis for Climate and Weather	Xavier Yepes, Mario C. Acosta
10:45	Lab Tutorial: Performance Analysis	
	Paraver Hands-on Using the HARMONIE Model	Mario C. Acosta, Xavier Yepes

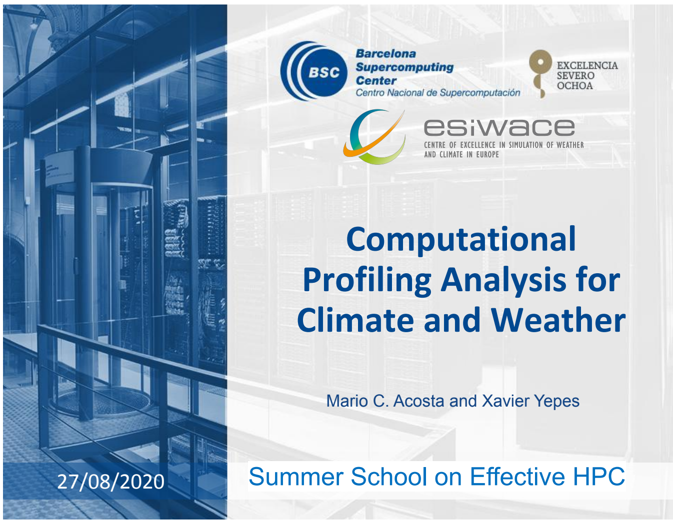
<table>
<tr>
<td>
</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.27: Computational Profiling Analysis for Climate and Weather

<table>
<tr>
<td>
</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.28: Performance Analysis: Paraver Hands-on Using the HARMONIE Model

Comments from the Survey

3.3.9 Containers

Abstract

This session will present an introduction to an end-to-end scientific computing workflow utilising Docker containers. Attendees will learn about the fundamentals of containerisation and the advantages it brings to scientific software. Participants will then familiarise with

Docker technologies and tools, discovering how to manage and run containers on personal computers, and how to build applications of increasing complexity into portable container images. Particular emphasis will be given to software resources which enable highly-efficient scientific applications, like MPI libraries and the CUDA Toolkit.

The second part of the lecture will focus on deploying Docker images on high-end computing systems, using a container engine capable of leveraging the performance and scalability of such machines, while maintaining a consistent user experience with Docker.

Learning Objectives

- Describe the difference between a container and a virtual machine
- Explain the relationship between a container and a container image
- Outline the basic workflow for the distribution of an image
- List advantages of using containers for scientific applications
- Write a Dockerfile
- Build a container image using Docker
- Run containers on personal computers using Docker
- Perform basic management of Docker containers and images
- Explain the motivations which drove the creation of HPC-focused container solutions
- Highlight differences and similarities between Docker and Sarus

Sessions

Containers		
Chair: Alberto Madonna (ETH Zürich, Switzerland) Chair: Simon Wilson (NCAS, UK)		
Time	Title	Speakers
13:30	Introduction to Containers and Docker	Alberto Madonna
15:15	Introduction to Containers on HPC with the Sarus Container Engine	Alberto Madonna
16:45	Lab Tutorial: Containers	
	Containers Hands-on	

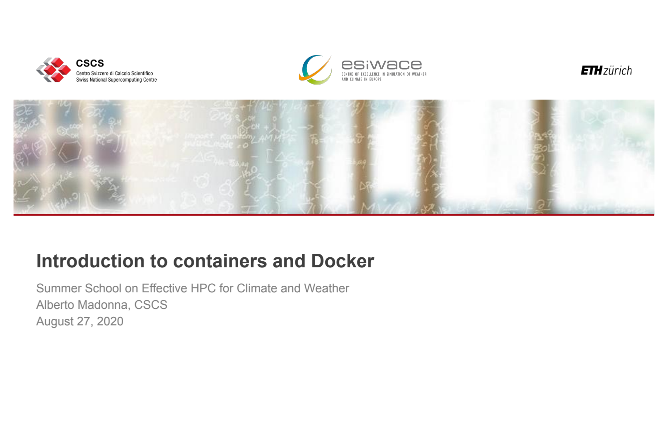
	Slides	Video

Table 3.29: Introduction to Containers and Docker

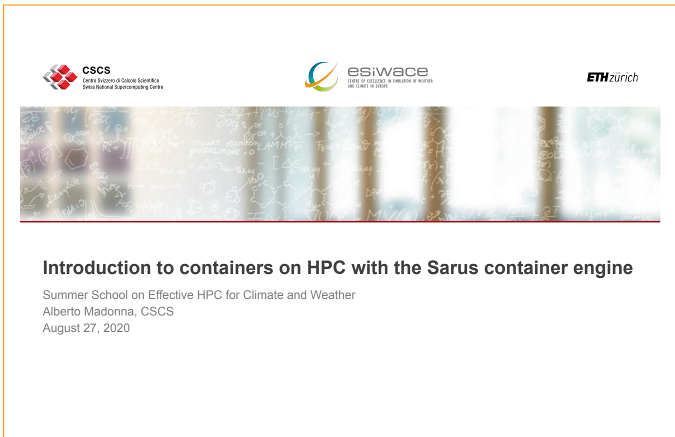
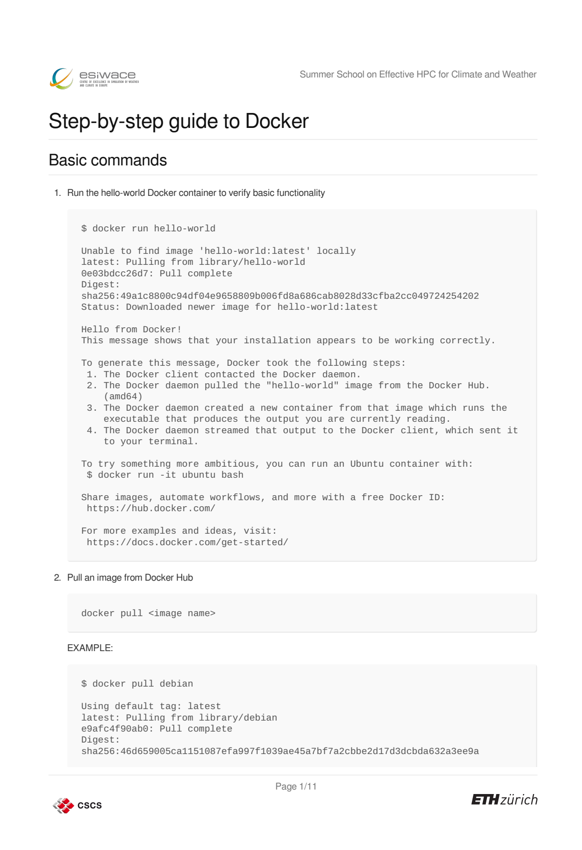
	Slides	Video

Table 3.30: Introduction to Containers on HPC with the Sarus Container Engine

	Slides	Video

Summer School on Effective HPC for Climate and Weather

Step-by-step guide to Docker

Basic commands

1. Run the hello-world Docker container to verify basic functionality

```
$ docker run hello-world

Unable to find image 'hello-world:latest' locally
latest: Pulling from library/hello-world
0e03bdcc26d7: Pull complete
Digest:
sha256:49a1c8800c94df04e9658809b006fd8a686cab8028d33cfba2cc049724254202
Status: Downloaded newer image for hello-world:latest

Hello from Docker!
This message shows that your installation appears to be working correctly.

To generate this message, Docker took the following steps:
 1. The Docker client contacted the Docker daemon.
 2. The Docker daemon pulled the "hello-world" image from the Docker Hub.
    (amd64)
 3. The Docker daemon created a new container from that image which runs the
    executable that produces the output you are currently reading.
 4. The Docker daemon streamed that output to the Docker client, which sent it
    to your terminal.

To try something more ambitious, you can run an Ubuntu container with:
 $ docker run -it ubuntu bash

Share images, automate workflows, and more with a free Docker ID:
 https://hub.docker.com/

For more examples and ideas, visit:
 https://docs.docker.com/get-started/
```

2. Pull an image from Docker Hub

```
docker pull <image name>
```

EXAMPLE:

```
$ docker pull debian

Using default tag: latest
latest: Pulling from library/debian
e9afc4f90ab0: Pull complete
Digest:
sha256:46d659005ca1151087efa997f1039ae45a7bf7a2cbbe2d17d3dcbda632a3ee9a
```

Page 1/11

Table 3.31: Containers Hands-on

Comments from the Survey

"Unfortunately, I was lost with the topic."

3.3.10 Q/A Session

Abstract

On Friday, a dedicated Q&A slot per lab practical was scheduled. The session offered the participants the opportunity to contact the organiser of each hands-on session and ask questions regarding the topic and particularly regarding the tutorial and exercises.

Sessions

Q/A Session (CHAIRS TO BE DOUBLE-CHECKED.)		
Chair: Julian Kunkel (University of Reading)		
Time	Title	Chairs
09:00	Monday Lab: Extreme-Scale Computation	Carlos Osuna
09:30	Tuesday Lab: Modern Storage	Konstantinos Chasapis
10:00	Tuesday Lab: Input/Output and Middleware	Julian Kunkel, Luciana Pedro, Sadie Bartholomew
10:30	Wednesday Lab: High-Performance Data Analytics and Visualisation	Donatello Elia
11:00	Thursday Lab: Performance Analysis	Mario C. Acosta, Xavier Yepes
11:30	Thursday Lab: Containers	Alberto Madonna

Comments from the Survey

"Its like exam after the course."

3.3.11 Keynote Talk

Daniel Klocke, from DWD (Germany), presented his work with Global Storm and Ocean Eddy Resolving Coupled Climate Simulations and the contributions of projects DIAMOND and DIAMOND2.

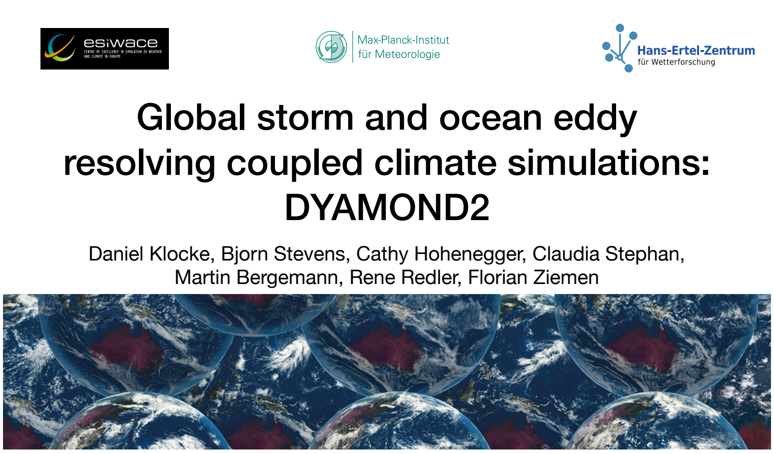
	Slides	Video

Table 3.32: Global Storm and Ocean Eddy Resolving Coupled Climate Simulations

3.4 Applicants

3.4.1 Letters

Everyone interested in joining the Summer School 2020 was first required to make a formal registration. The selection procedure for applicants was conducted by a committee that selected the registrations to ensure a balance across countries and gender and support to those in need. All applicants had to submit:

- Up to one-page motivational letter including:
  - A tentative idea of a project that can be conducted as part of the Academic Group Projects, including no more than five keywords, and
  - How you will act as a multiplier of the gathered information.
- Up to one-page CV showing that the applicant satisfies the description of the target audience including a reference to one paper/thesis/dissertation/project in a related area in which the applicant is the author (or one of the authors).

Considering that the main idea behind the summer school is spreading concepts on effective HPC for climate and weather, typically only one applicant will be selected per university or company.

Applicants that did not require funding from the summer school had priority to ensure that more people would benefit from the proposal.

3.4.2 General Application

We had three rounds of applications, with respective deadlines:

06 of March, 2020 Applications for attending the summer school open (with optional subsidy).

12 of May, 2020 General applications for attending the summer school.

23 of August, 2020 Applications for attending the summer school reopen. Event is now online and free!

We followed the first two deadlines and selected the participants for the event that would receive a subsidy. We went as far as booking the venue and the hospitality package for the attendees. However, due to the pandemic, the organisation decided to cancel the face-to-face event, and the agenda was adapted to make the most out of the virtual event.

The Summer School on Effective HPC for Climate and Weather was an online event. On the positive side, the summer school was open to participants free of charge, and we manage to record videos of the training. Lab sessions were still offered to participants, and the full material is available for download.

We ended up opting for requesting all participants to fulfil a simplified Google Form to register to the event. All applicants were inserted in a mailing list to facilitate the communication between them and the organisation, and this list will now be used to disseminate further events.

3.5 Participants

The Summer School on Effective HPC for Climate and Weather welcomed 162 participants. In total, we had 38 countries represented in the event. The countries with more participants were: UK (19.8%), India (12.3%), and Germany (11.1%) (Figure 3.1).

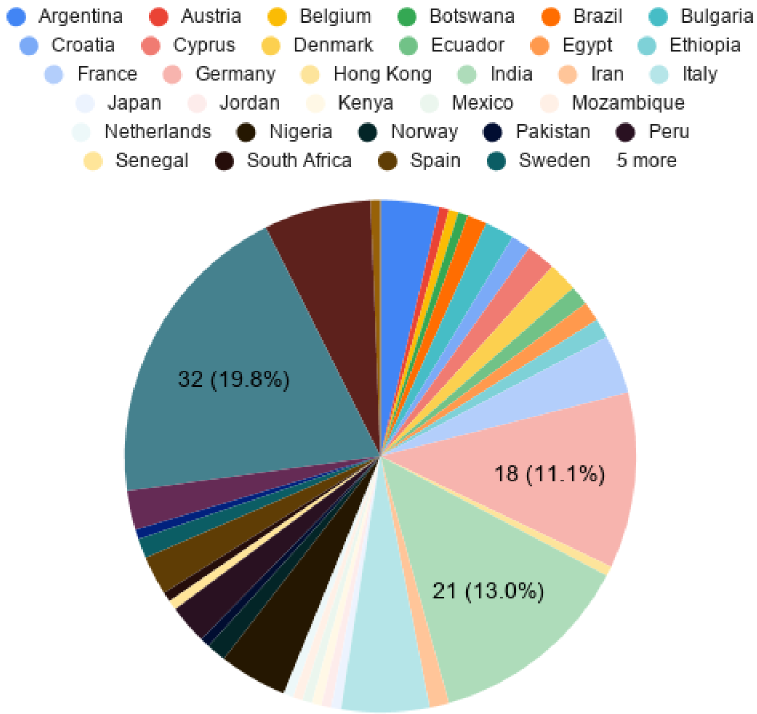


Figure 3.1: Country of the participants

The significant majority of the participants were PhD students (42.6%), but we also had a substantial attendance of researchers (32.8%), non-academic professionals, and students (not yet in PhD level) (Figure 3.2).

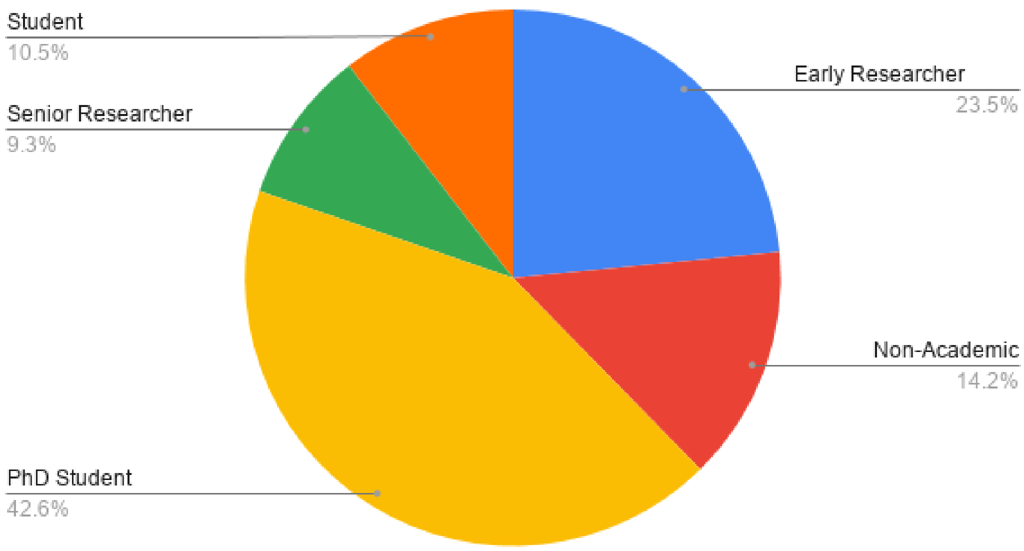
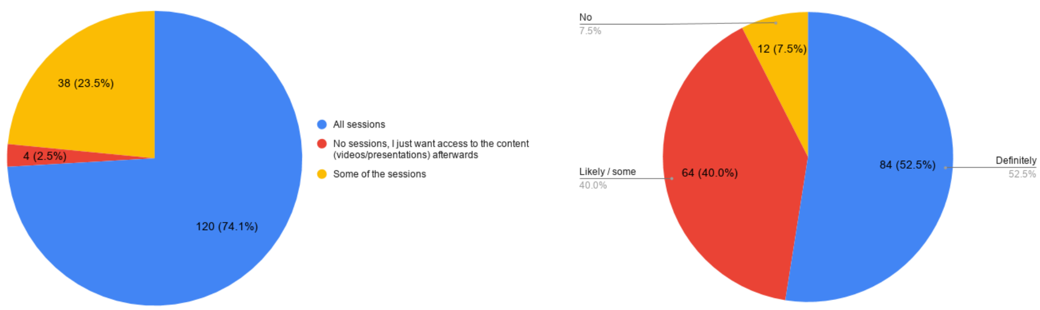


Figure 3.2: Occupation of the participants

In the registration form, 120 people assured they would participate in all topical sessions, and 84 would definitely participate in the hands-on sessions (Figure 3.3).



All sessions Hands-on sessions

Figure 3.3: Expected participation in the summer school

The 54 people that attended more than 70% of the event were granted a personalised certificate of participation (Figure 3.4).

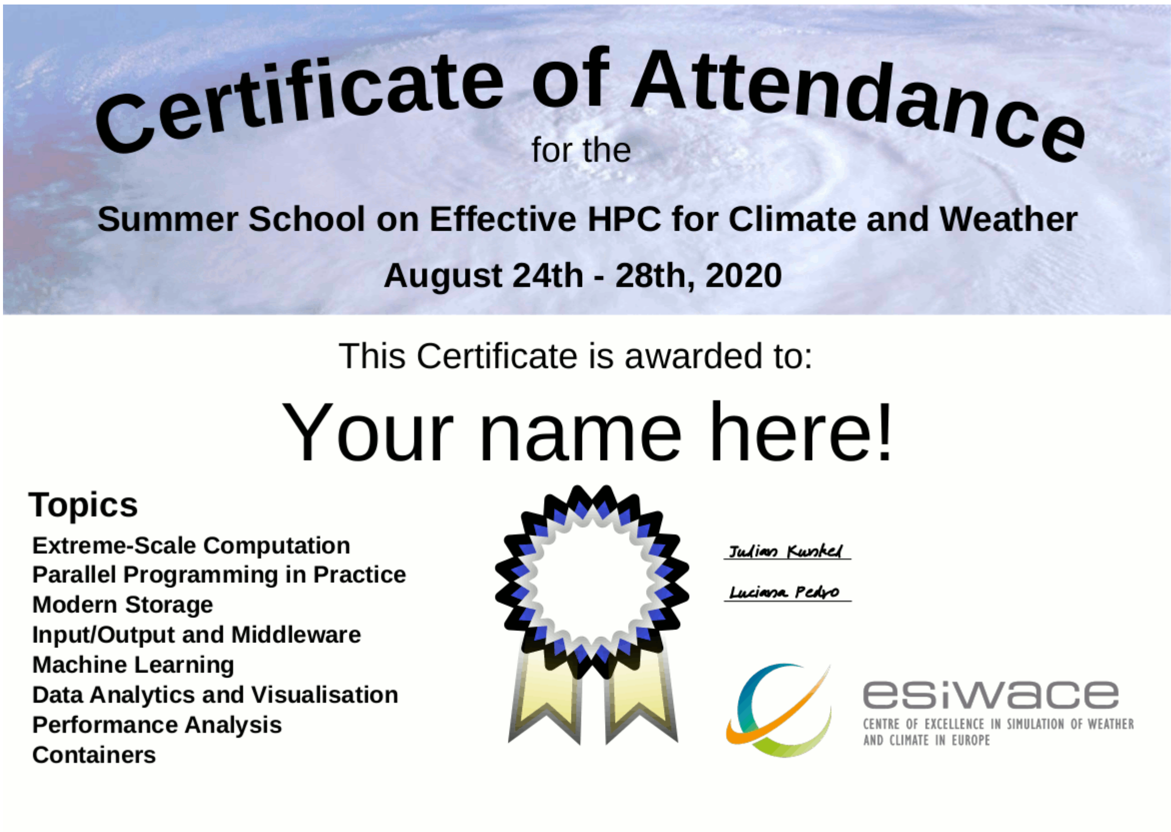


Figure 3.4: Certificate of Attendance

3.6 Survey

After the event, we asked the participants to fulfil a Survey. The survey was sent to all participants registered in the mailing list of the event. We had 13 complete answers. The next subsections cover the statistics and the answers provided by the participants.

Here are some of the main numbers and comments about what they most liked about the summer school:

- 38.5% said the summer school was better than they were expecting and 46.2% said it was as good as they were expecting;
- 92.3% said they had an excellent or good experience;
- Selected quotes from attendees:
  - “The combination of learning the theory, getting some hands-on it, and working with the learned things”;
  - “The interdisciplinary spirit of the presentations and profound knowledge of people invited to talk”;
  - “High-quality presentations and I can revisit it when I need on youtube”.

3.6.1 Statistics

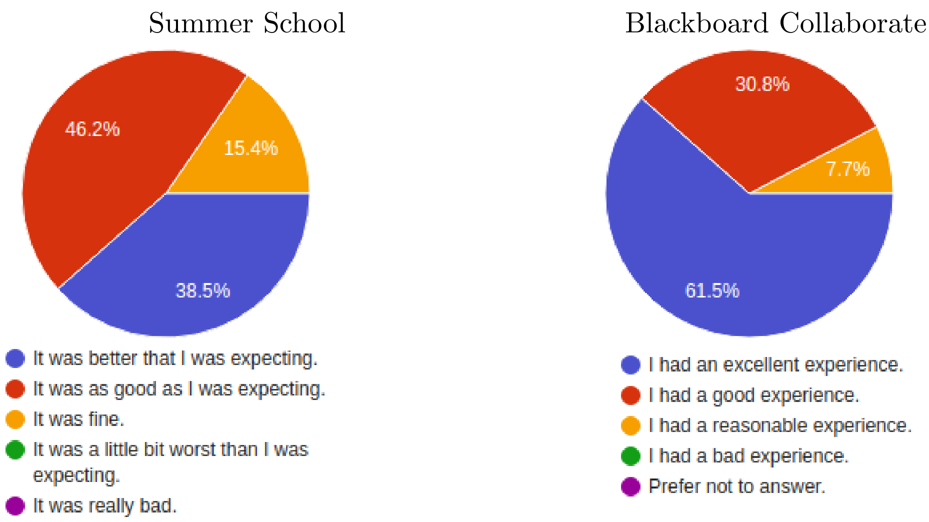


Figure 3.5: General assessments

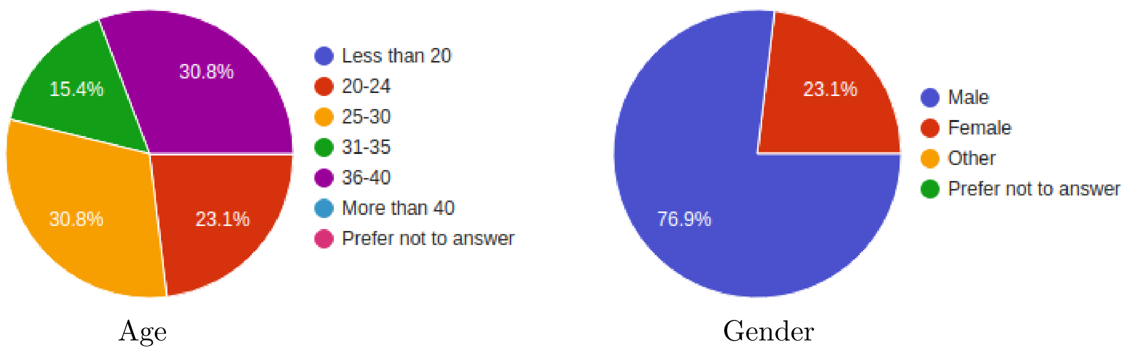


Figure 3.6: Personal information

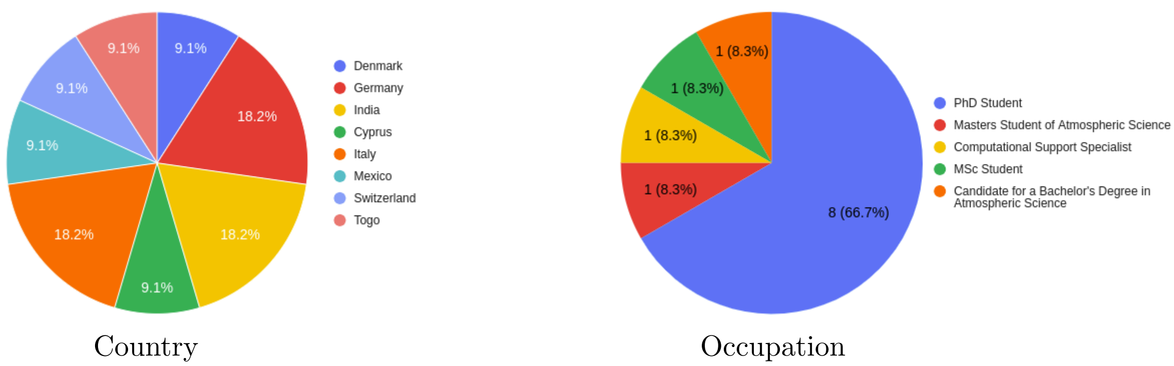


Figure 3.7: Personal information

3.6.2 Comments

What did you most like about the summer school?

“High quality, great variety.”

“Switching to a free online summer school instead of cancelling.”

“Coupling, container, MPI.”

“The combination of learning theory and get some hands-on it and work with the learned things.”

“Interactive talks.”

“High-quality presentations and I can revisit it when I need on youtube.”

“The presentations wich application examples.”

“It was new for me but very knowledgeable sessions program.”

“The interdisciplinary spirit of the presentations and profound knowledge of people invited to talk.”

“Practical Sessions.”

“Theory as well as practical sessions.”

“Feedback questions on presentation and hand-on practicals.”

“I liked that it was accessible to anyone around the world. That there was a wide variety of topics and that the experts of each session were people who were passionate about their field of study. Also, everything was really good planned and that we had access to all the software in the VM.”

What did you most dislike about the summer school?

“No room for personal presentation, 70 participants and no general overview or insight of who was attending from where. Also: maybe too little breaks.”

“Compatibility problems with the VMs (host: mac).”

“I am not sure.”

“I wasn’t really into the topic before and some sessions were really hard to understand because some basic knowledge was missing.”

“Hard to work online on tutorials.”

“Bad English of some of the presenters.”

"Time shortage in the hands-on session."

"Nothing in particular."

"None."

"Difficulty to get personalised coaching during practicals (probably due to the virtual nature of the summer school)."

"That I had to wake up at 3 AM in Mexico."

Please, let a general comment about your experience in the summer school.

"Best professionals, over a wide range, brought together into one great, interesting event."

"I've learned several things from the computational point of view. Many of which I would never even have the ability to formulate the questions if were it not for a course geared towards climate scientists."

"The introduction was quite steep and dry for a person with few knowledge about processors etc."

"It was a good experience to see some topics I am familiar with (CS) applied in an interesting field that I don't know in deep details (climate)."

"Everything was just wonderful."

"My personal experience at the summer school was great! I had a great overview of the recent advancements and orientations in HPC for the weather and climate domain."

"My experience in the course was really good. I am Mexican, and because of the time difference I had to wake up at 3 AM for the summer school, but it was totally worth it. I enjoyed every session and learned a lot. Many topics were new for me, but the speakers were excellent. It is definitely an excellent HPC introductory course and also an excellent course for those who already have some experience in the field."

3.7 Improvements

We learned many things with this first summer school that we will use to improve the event in 2021. The environment for the summer school 2021 is expected to be different (we hope we have the event in Reading, UK), but there are some lessons that we can apply to the next event, regardless of how the event will happen next year.

- Mailing List

  It was a very good idea to create a mailing list with all participants of the event, including the speakers. However, although the list really facilitated the communication between the organisation and the attendees, the secondary benefit (support the attendees) was not achieved.

We would like that list to be used by the participants to ask questions and interact with each other and the speakers. The participants had the chat of the Blackboard Collaborate to contact the speakers during their presentation and to exchange information among themselves during the Virtual Lab Session. The former was used successfully, but the later was almost not used at all.

Encourage the attendees to use the mailing list to interact would have been an extra incentive for them to go through the recorded labs, pose different questions, help each other and extend their personal network. It was a missed opportunity that we intend to emphasise in the next summer school.

- Networking

  In the physical event, we had planned many opportunities for the participants to do networking. We had group projects and daily get together at the end of the day. When translating the event to virtual, we lost those options, and we could not think about something extra to replace the interaction they would have in the projects and confraternisations.

  Here we can mention the mailing list again, that would help the attendees to contact one another and improve their personal network. Besides that, in case we need to plan another virtual event, this time we have extra time to find creative solutions for helping the participants connect to each other and also to the speakers. Hopefully, we will be able to put the original plan in practice in 2021.

- Virtual Machine

  The Virtual Machine (VM) was a fantastic idea that allowed all the participants to work with tools developed and used in ESiWACE project in the same environment. By doing that, we saved time that would be used configuring each machine and dealing with operating systems specifications. However, not everyone was able to install the VM. In this case, it was nearly impossible to follow the labs hands-on.

  Again, for the event in Reading, we were planning to have physical labs in which the system would be already appropriately configured for all the attendees to focus their attention only in the tools they are learning, and not waste time configuring systems. But even in a physical event, the VM is a good idea because it allows the participants to practice on their own and in their personal machine.

  The VM was delivered to the participants a month in advance, to give them time to fix any possible incompatibility with their machine. For the event in 2021, we plan on having the VM again, together with a dedicated assistant (possibly one of Reading's students) to provide support for the participants that are facing problems that they cannot solve. With that, we intend to guarantee that everyone will have the VM adequately installed and running, and they will then have the opportunity to attend all labs prepared for the event.

- Platform

  The platform chosen for the summer school was the Blackboard Collaborate Ultra (BBCU). The University of Reading uses Blackboard for organising the online learning experience for students, and BBCU is a part of it. The platform has advantages such as it is easy to use and works by web technology – you only need a modern browser to connect to it. BBCU also allows automatic recording of the sessions and provides different levels of access: moderator or guest. Another major disadvantage is that there is an easy to use chat, in which one can send direct messages (DM) or global messages (GM) to all logged-in the session.

One of the disadvantages of this platform is that all the participants have to connect as guest users, which removes a more personalised experience, such as inserting photos in their profile. The chat is local to the session (both DM and GM), and it is not recorded when the session ends. Also, if someone disconnects, the discussion restarts after each reconnection.

The organisers are evaluating other platforms in case the summer school 2021 is also designed as an online event. Even if we manage to have a face-to-face the event in 2021, we intend to record the event and then we need to decide on the best tool to achieve good results.

- Survey

  From the survey, we also noticed further room for improvements:

  - "No room for personal presentation, 70 participants and no general overview or insight of who was attending from where."

    We hope to have a better networking structure in the next event.

  - "Maybe too little breaks."

    We had a structure with 15 minutes break during the morning and afternoon sessions and 45 minutes for a lunch break. The small duration of the intervals is standard in this kind of event, and we took advantage of the virtual structure to have a more compact lunch break. If we need to run the event online again, we may rethink these options. If a physical event is possible in 2021, the whole structure will change, and we will address the breaks more carefully.

  - "Time shortage in the hands-on session."

    The hands-on sessions were pre-recorded to give the participants the freedom to choose which session to check out. Again, modifications in the format will depend on the kind of event we can organise in 2021.

  - "Difficulty to get a personalised coaching during practicals (probably due to the virtual nature of the summer school)."

    When we were planning the event, we had designed group projects. In this case, each group of four to five people would have a designated supervisor. With the virtual event, the number of participants was too high to have personalised coaching. But many of the speakers left their contacts and offered to provide further help during and after the event.

  - "That I had to wake up at 3 AM of Mexico."

    We hope to welcome all the participants in Reading (UK) for the summer school in 2021. Hence everyone can attend the sessions in reasonable hours.

4 Summer School on Effective HPC for Climate and Weather – 2021

5 OER Materials

5.1 Summer School 2020

This chapter summarises the OER material developed as a result of the Summer School on Effective HPC for Climate and Weather 2020.

5.1.1 Extreme-Scale Computation

Extreme Computing Session – Overview	Slides	Video

Supercomputer Trends	Slides	Video

Performance, Portability and Productivity	Slides	Video

Introduction to DSLs	Slides	Video

An Introduction to PSyclone	Slides	Video

Dusk & Dawn – Introduction and Overview	Slides	Video

PSyclone Tutorial	Slides	Video

Dusk & Dawn Tutorial	Slides	Video

5.1.2 Parallel Programming in Practice

Scaling Algorithms	Slides	Video

Code Coupling	Slides	Video

5.1.3 Modern Storage

Modern Storage	Slides	Video

Darshan Demonstration – Hands-on Session	Slides	Video

Installing Darshan for I/O Performance Analysis	Slides	Video

Introduction to Using Darshan for I/O Performance Analysis	Slides	Video

5.1.4 Input/Output and Middleware

Input/Output and Middleware	Slides	Video

Python Data Tools for CF-netCDF	Slides	Video

An Introduction to NetCDF Using C Language	Slides	Video

CF-NetCDF with cfdm, cf-python and cf-plot	Slides	Video

5.1.5 Machine Learning

Machine Learning for Weather and Climate Predictions	Slides	Video

5.1.6 ECMWF – Virtual Visit

Introduction to ECMWF	Slides	Video

Computer Hall Tour	Slides	Video

Weather Room Tour	Slides	Video

5.1.7 High-Performance Data Analytics and Visualisation

Data Visualization Using ParaView	Slides	Video

Hands-on: Data Visualization Using ParaView	Slides	Video

High-Performance Data Analytics and Visualisation	Slides	Video

High-Performance Data Analytics and Visualisation	Slides	Video

5.1.8 Performance Analysis

Computational Profiling Analysis for Climate and Weather	Slides	Video

Performance Analysis: Paraver Hands-on Using the HARMONIE Model	Slides	Video

5.1.9 Containers

Introduction to Containers and Docker	Slides	Video

Introduction to Containers on HPC with the Sarus Container Engine	Slides	Video

Containers Hands-on	Slides	Video

5.2 Summer School 2021