Report on summer school and its material

Deliverable D6.1

The project Centre of Excellence in Simulation of Weather and Climate in Europe Phase 2 (ESiWACE2) has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 823988

About this document:
Work package in charge: WP6
Actual delivery date for this deliverable: 2021-12-31
Dissemination level: PU (for public use)

Lead authors:
Luciana Pedro (University of Reading)
Julian Kunkel (GWDG)
Jack Ogaja (GWDG)

Internal review: Julian Kunkel (GWDG), Sophie Valcke (CERFACS)

Contact details:
Project Office: esiwace@dkrz.de
Visit us on: https://www.esiwace.eu

Access our documents in Zenodo:
https://zenodo.org/communities/esiwace

Contents

1 Abstract /publishable summary

This report describes the organisation and outcome of the two ESiWACE summer schools in 2020 and 2021, in particular, the offered training sessions. Additionally, the results of the conducted surveys regarding the quality and effectiveness of the summer schools are illustrated.

The summer schools were organised as a one-week virtual event in August 2020 and 2021 due to COVID-19 restrictions. The scope of the schools is in the training of young researchers and software engineers in methods, tools, and theoretical knowledge to make effective use of HPC environments and generate insights. The sessions included about 34 hours of lectures and 18 hours of hands-on and lab practicals.

Overall 289 people registered for the schools – 87 obtained a certificate of attendance because they attended most of the sessions in the respective school. The created videos of the sessions were published on YouTube after the event to allow even more attendees to enjoy the training. The post-event survey suggests that more than 95% of the attendees had at least a good experience and generally, a favorable appraisal was given. In the CoE proposal, we expected to host 30-40 students per school which were surpassed by the total number of attendees of the summer schools. We conclude the lectures and lab practicals were generally well-received and despite being a virtual event we successfully delivered the planned training events.

2 Conclusion & Results

The main outcome of the effort is the training of the attendees and the long-term provision of the generated training material on the webpage of the summer school(s) including the videos of the recorded sessions in YouTube. The topics covered were:

- Extreme-Scale Computation
- Parallel Programming in Practice

- Modern Storage
- Input/Output and Middleware
- Machine Learning
- High-Performance Data Analytics and Visualisation
- Performance Analysis
- Containers

3 Project objectives

This deliverable contributes directly and indirectly to the achievement of all the macro-objectives and specific goals indicated in section 1.1 of the Description of the Action:

Macro-objectives	**Contribution of this deliverable**
(1) Enable leading European weather and climate models to leverage the available performance of pre-exascale systems with regard to both compute and data capacity in 2021.	X
(2) Prepare the weather and climate community to be able to make use of exascale systems when they become available.	X

Specific goals in the workplan	**Contribution of this deliverable**
Boost European climate and weather models to operate in world-leading quality on existing supercomputing and future pre-exascale platforms	X
Establish new technologies for weather and climate modelling	-
Enhance HPC capacity of the weather and climate community	X
Improve the toolchain to manage data from climate and weather simulations at scale	-
Strengthen the interaction with the European HPC ecosystem	X
Foster co-design between model developers, HPC manufacturers and HPC centres	-

4 Detailed report on the deliverable

The detailed report is enclosed after this header document.

5 Changes made and/or difficulties encountered, if any

The COVID pandemic required us to flexibly react on the rapidly changing lockdown rules in 2020. While initially planned as in-person meeting, close to the event in August 2020, it was decided to move the event to a fully virtual event overhauling the planing for the onsite visits and projects. Thankfully, all contributers adapted to the changing format. We aimed to exploit the virtual event by dropping any attendance restriction and, hence, allowing everyone to participate in the summer school. This lead to a high number of registrations – many students only attended few topics that they were most interested in, though.

6 How this deliverable contributes to the European strategies for HPC

The summer school contributed to the European ecosystem by allowing young researchers to better harness HPC resources and by promoting HPC software technology (that also is) developed in the EU.

7 Sustainability

The summer school benefited from other training activities of ESiWACE. Some of the material was used across topic-specific trainings offered in WP6, such as for the DSLs. A large portion of training material generated for the school is also used (full, in parts, or remixed) in projects related to the software that we trained for.

The experience with the summer school indicate:

- Generally, researchers in the domain acknowledge the need for training and there is a high demand, i.e., young researchers still need more training.
- The virtual event allows theoretically for multiplying the attendance. The video recording and provisioning on YouTube contributes to long reach of the event.
- Drawbacks of a virtual event is that in-depth engagement with and between the attendees is limited. Therefore, the individual networking benefits for individuals is suboptimal.

From the experience, we conclude that a mix of in-person and virtual events would be optimal to reach out to participants. However, sharing slides and videos on the event page(s) and promoting them offers prospect attendees worldwide the opportunity to gain training.

8 Dissemination, Engagement and Uptake of Results

8.1 Target audience

As indicated in the Description of the Action, the audience for this deliverable is:

X	The general public (PU)
	The project partners, including the Commission services (PP)
	A group specified by the consortium, including the Commission services (RE)
	This reports is confidential, only for members of the consortium, including the Commission services (CO)

8.2 Record of dissemination/engagement activities linked to this deliverable

See table 1.

The number of person reached is the unique number of people that attended at least one of the sessions.

8.3 Publications in preparation OR submitted

None

8.4 Intellectual property rights resulting from this deliverable

None

Table 1: Record of dissemination / engagement activities linked to this deliverable

Type of dissemination and communication activities	Details	Date and location of the event	Type of audience	Zenodo Link	Estimated number of persons reached
Organisation of a workshop	https://hps.vi4io.org/events/2020/esiwace-school	24-28 August 2020	Scientific Community		90+
Organisation of a workshop	https://hps.vi4io.org/events/2021/esiwace-school	23-27 August 2021	Scientific Community		65+

D6.1 – Report on Summer School and Its Material

Luciana Pedro Julian Kunkel Jack Ogaja

Work Package:	Work Package 6
Responsible Institution:	University of Reading (UREAD)
Contributing Institutions:	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
Date of Submission:	December 13, 2021

Contents

1 Introduction

This report summarises the outcomes of hosting two summer schools on Effective HPC for Climate and Weather in 2020 and 2021.

1.1 Executive Summary

Due to COVID-19 restrictions, the Summer School on Effective HPC for Climate and Weather has been held online twice for the past two years (2020 and 2021). Both events provided five days of learning and debating HPC systems and climate and weather applications.

The 2020 school which was held from 24th to 28th August attracted 162 applications out of which 54 participants who achieved more than 70% attendance were offered certificates 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.

Similarly, the 2021 school which took place from 23rd to 27th August, also attracted participants from all over the world with significantly varying scientific backgrounds and levels of qualifications. In the 2021 event we received 125 registrations for the school with most affiliations of participants from from Indonesia (17%), Germany and Italy (13% each), India (8%), France and UK (6% each), and Austria and Cyprus (2%). The 2021 attendance was between 30 - 50 people per session affiliated to institutions in 31 countries. 33 of the participants who attended more than 60% of all sessions received certificates of attendance.

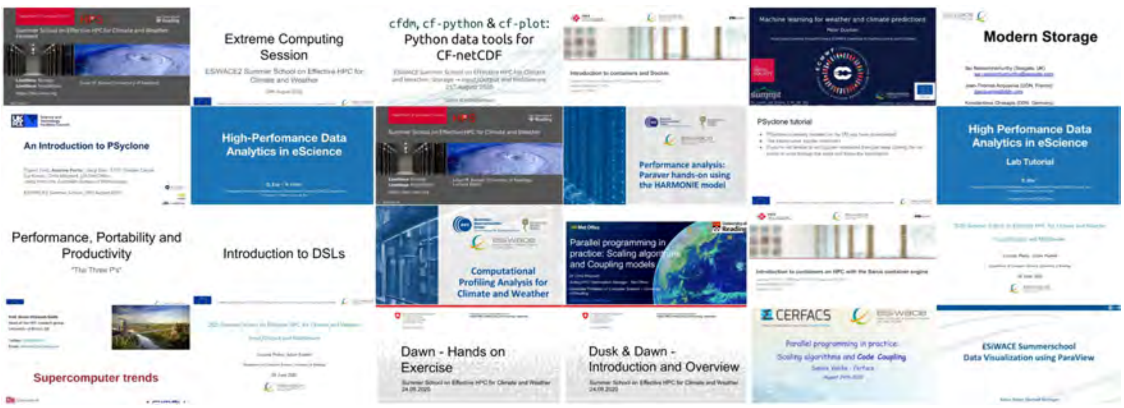


Table 1.1: Snapshot of the presentations.

In both schools, the format was slightly changed compared to previous schools to accommodate online participation. 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).

A different format was adopted for the Friday's program where extra sessions in a Q&A format for each lab that happened during the week, and finished with a keynote talk. For the 2020 event, Daniel Klocke, from the German Meteorological Service (DWD), presented his work on the global storm- and ocean eddy-resolving coupled climate simulations and the contributions of projects DYAMOND and DYAMOND2. For the 2021 event, two invited talks took place. Miguel Castrillo (BSC), Dorotea Iovino (CMCC), and Clement Bricaud (Mercator Ocean) presented "Preparing NEMO and EC-Earth models for very high-resolution production experiments", and Sebastian Paczkowski from the University of Goettingen presented "Project FORESTCARE: An HPC application case study for a high spatial and temporal resolution forest vitality assessment".

For both schools all sessions were recorded and the recordings, together with the presentations and material for each session, are available on both summer school webpages, ss-2020 and ss-2021. All applicants were also added to a mailing list to facilitate communication between them and the organizers, and to inform the participants about related events afterward.

After both events, we asked the participants to fill out a survey. The following are some of the main numbers and comments about the 2020 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, 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".

For the 2021 school, we also received similar feedback with 92% of good or very good experiences.

1.2 This Document

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 into the targets of the project, **Excellence in Simulation of Weather and Climate in Europe, Phase 2 (ESiWACE2)**. The topics are described in paragraphs 2.0.0.0.1 to 2.0.0.0.6 and the summer school proposal can be found in paragraph 2.0.0.0.7. Chapter 3 reports the Summer School on Effective HPC for Climate and Weather 2020 and Chapter 4 is about the Summer School 2021. The material for each session is available on both summer school webpages, ss-2020 and ss-2021. The links are also available throughout this document, in their respective sessions, and in Chapter 5.

As part of the delivery for the school, we collaborated with the HPC-Certification Forum (HPC-CF) contributing various skills of the individual sessions to the certification standard. As the certification requires multiple-choice questions and interactive scenarios, we realised that the required amount of work would exceed what we can contribute as part of the ESiWACE project. Nevertheless, we found the certification a general support-worthy approach.

2 The Proposal

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).

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 paragraphs, we describe the different ESiWACE2 trainings as they were originally planned in the Description of Work.

2.0.0.0.1 Training on I/O and HPC Awareness This task covers training on the I/O hardware and software stack, and targets scientists accessing data on storage and developers that aim to utilise storage APIs efficiently. We created 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 also explored how the analysis stack fits into it (in-depth analysis workflows are part of training on high-performance data analytics, see below). Furthermore, we covered 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.

2.0.0.0.2 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 provided insights on DSLs considered in ESiWACE2 which includes PSyclone, CLAW and GridTools ecosystem, and demonstrated how to apply them to weather and climate models. Participants theoretically and practically learnt how to use the DSL languages to implement PDE operators. During a hands-on session, were 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.

2.0.0.0.3 Training on C++ for HPC ETH Zurich has organised an advanced course for C++ in HPC as part of its regular training. This training presented 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.

2.0.0.0.4 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, focusing 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 covers 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.

2.0.0.0.5 Training on High-Performance Data Analytics The objective of this training 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 addresses 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, includes 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.

2.0.0.0.6 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 in this session were made available online to reach a wider audience.

2.0.0.0.7 Summer School on HPC for Weather and Climate In this task, two summer schools were organised for Earth system scientists, including PhD students and postdocs, covering the HPC aspects to efficiently run scientific workflows on a large scale HPC environments. The five-day summer schools took place in years two and three of the project and supported about 30 students each year by funding their attendance. The selection procedure for subsidising students were established six months before the summer school taking into account the experiences gained with summer schools held within the IS-ENES projects. It prioritised diversity and internationality in which students were funded depending on the financial support possible from their home institution. Theoretical courses and hands-on sessions were 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). External experts were also invited to give tutorials about selected topics. Selected presentations were recorded and made available online after the summer schools for a wider audience. A survey was then 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 attended 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 who had earned 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 provided an introduction/walk-through to the topic. It was either pre-recorded, or recorded during a live session and quickly provided on the webpage after the session.
- The tutorials were scheduled for each session, but the participants had the option to use them after the sessions.
- We provided a tutorial to set up a Virtual Machine (VM). The VM came with Ubuntu and all the software for the training pre-installed. Attendees could install the VM on their PC to perform most of the training. Some training were on a dedicated cluster. Participants had to set up the VM before the summer school if they liked 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 was given to allow independent and self-paced learning, and participants decided what to learn and when to engage.

For Friday, a dedicated Q&A slot per lab practical was scheduled. The session offered 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, through which participants could post questions and cooperate 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 on 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

3.3.1.1 Abstract

This session introduced 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 was discussed.

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

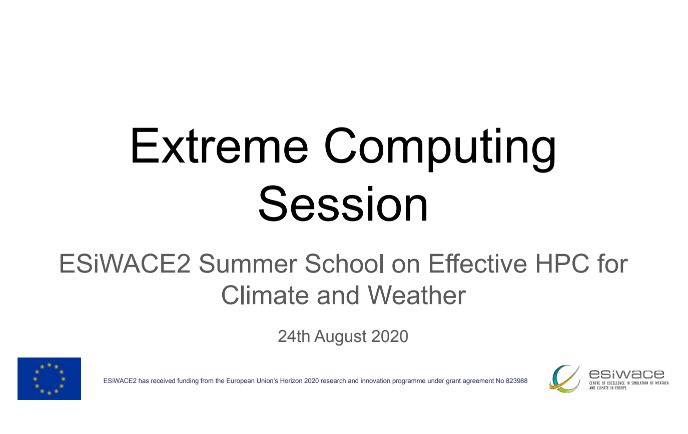
Lastly, a relatively new, Domain-Specific Language (DSL), approach to programming Weather and Climate models was introduced with examples from two existing DSLs – DAWN and PSyclone. It was shown that the DSL approach offers the possibility of supporting all three of the below requirements, by separating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

3.3.1.2 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

3.3.1.3 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

	Slides	Video

Table 3.2: Supercomputer Trends

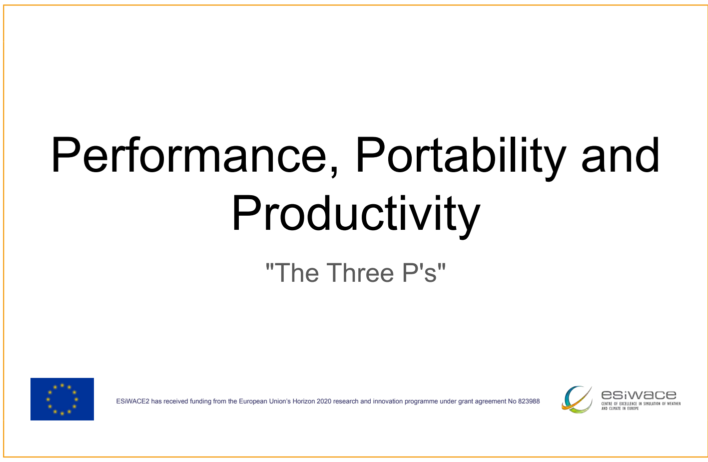
	Slides	Video

Table 3.3: Performance, Portability and Productivity

	Slides	Video

Table 3.4: Introduction to DSLs

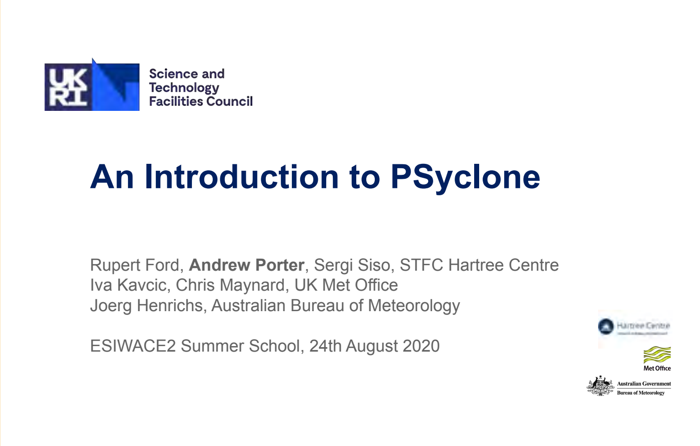
<table>
<tr>
<td>
</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.5: An Introduction to PSyclone

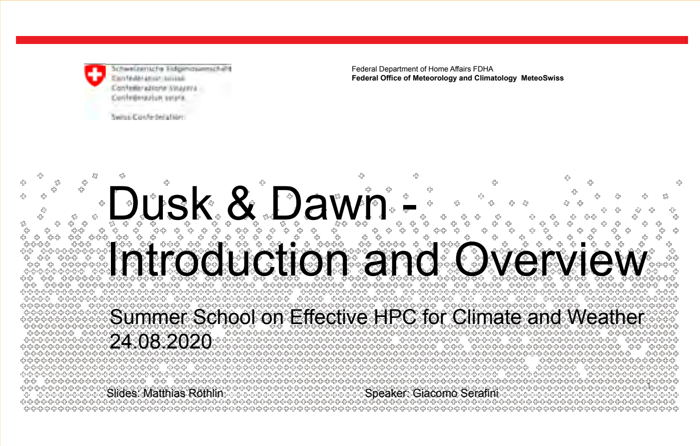
<table>
<tr>
<td>
</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.6: Dusk & Dawn – Introduction and Overview

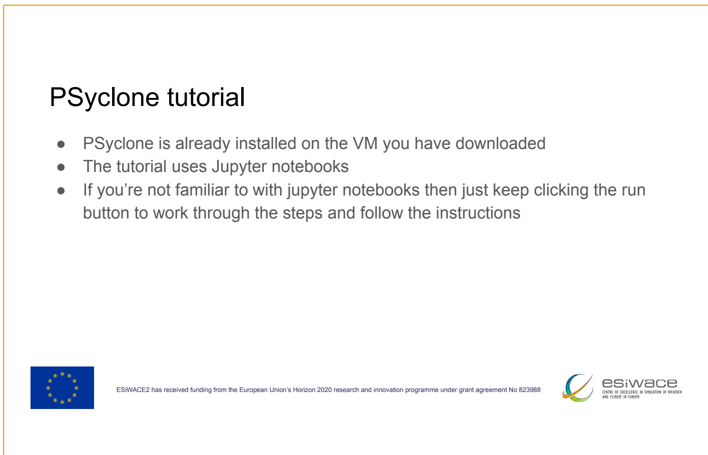
<table>
<tr>
<td>
</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

Table 3.7: PSyclone Tutorial

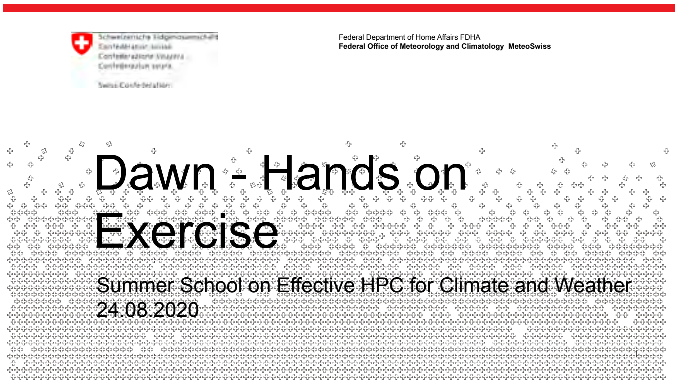
<table>
<tr><td>
</td><td>Slides</td><td>Video</td></tr>
</table>

Table 3.8: Dusk & Dawn Tutorial

3.3.1.4 Comments from the Survey

Some of the comments we received from the survey were: "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

3.3.2.1 Abstract

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

Particular attention was devoted to specific features that may inhibit scaling and performance of weather and climate codes. This analysis was 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.

3.3.2.2 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

3.3.2.3 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

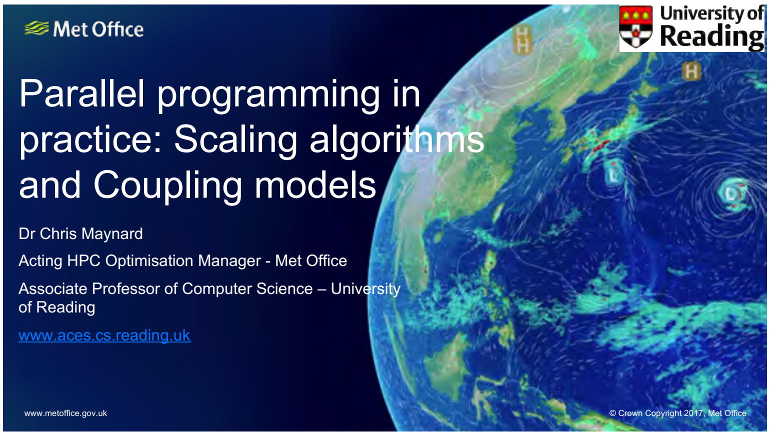
	Slides	Video

Table 3.9: Scaling Algorithms

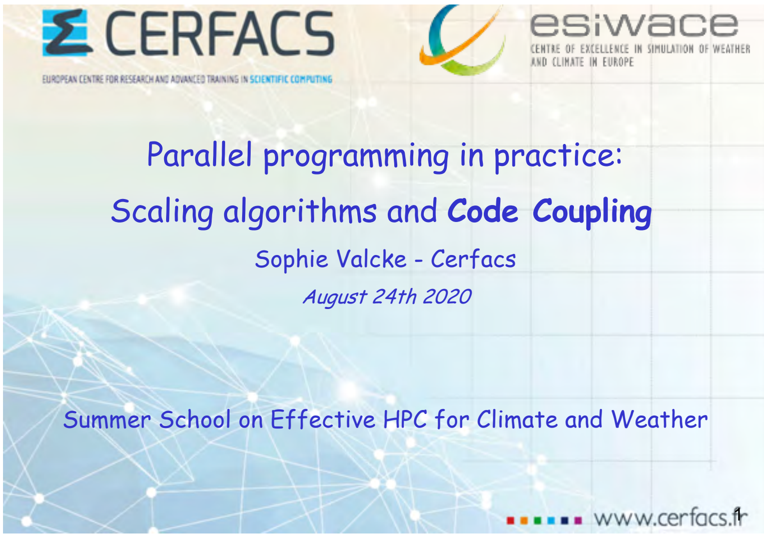
	Slides	Video

Table 3.10: Code Coupling

3.3.3 Modern Storage

3.3.3.1 Abstract

3.3.3.2 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

3.3.3.3 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

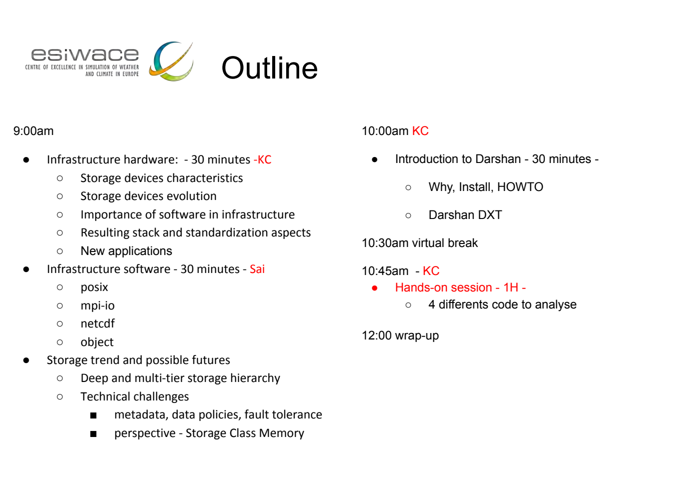
	Slides	Video

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

3.3.4 Input/Output and Middleware

3.3.4.1 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 was twofold. Firstly, we discussed storage layers focusing on the NetCDF middleware and provided a performance model that aids users to identify inefficient I/O. Secondly, we introduced the NetCDF Climate and Forecast (CF) conventions that are often used as a standard to exchange data.

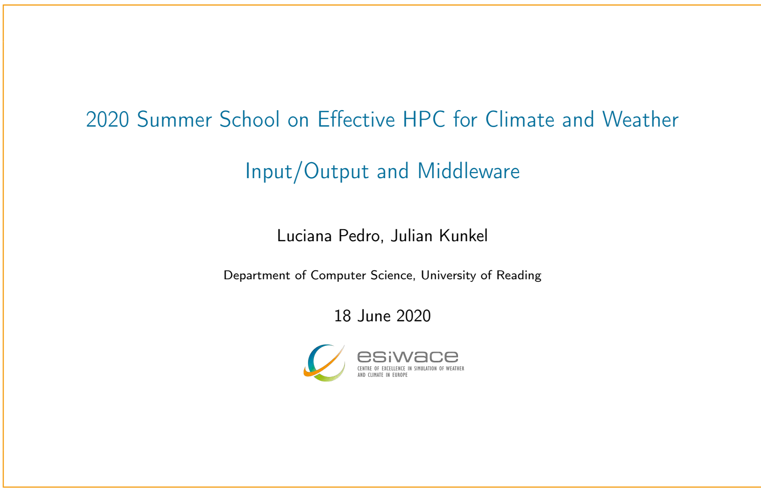
3.3.4.2 Learning Objectives

- 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

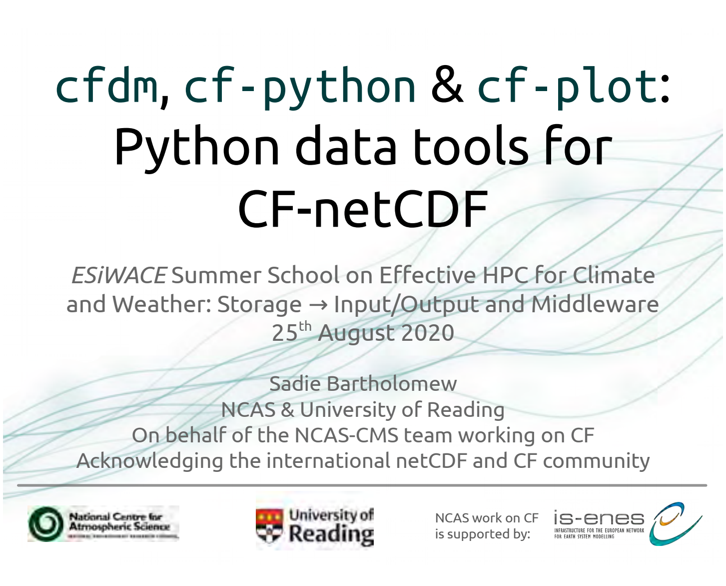
3.3.4.3 Sessions

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



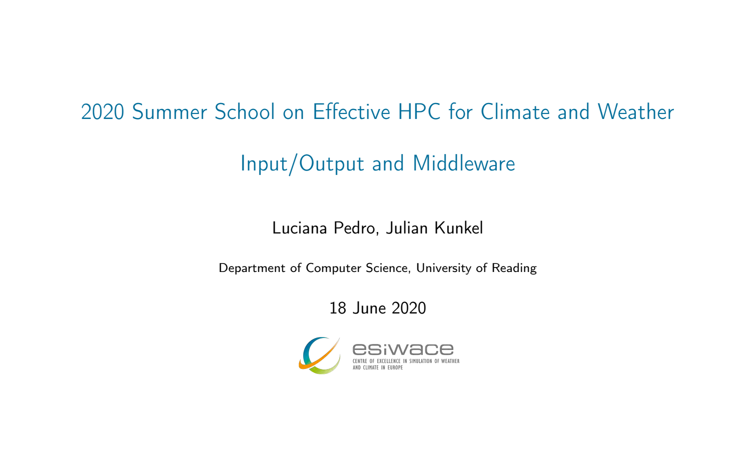
(slide above)	Slides	Video

Table 3.15: Input/Output and Middleware



(slide above)	Slides	Video

Table 3.16: Python Data Tools for CF-netCDF



(slide above)	Slides	Video

Table 3.17: An Introduction to NetCDF Using C Language

<table>
<tr>
<td>
Course order
The recommended order in which to work through the full course Notebooks is as follows, where all paths given are relative to <code>full_course/notebooks/</code> :
<table>
<tr><th></th><th>Topic</th><th>Format</th><th>Notebook name & location</th></tr>
<tr><td>1.</td><td>Introduction</td><td>Overview & walkthrough</td><td><code>introduction.ipynb</code></td></tr>
<tr><td>2.</td><td>Reading and manipulating data</td><td>Walkthrough</td><td><code>cf_python.ipynb</code></td></tr>
<tr><td>3.</td><td>Reading and manipulating data</td><td>Exercises</td><td><code>exercises/cf_python_exercise_*.ipynb</code> [* =1-5]</td></tr>
<tr><td>4.</td><td>Contour plots</td><td>Walkthrough</td><td><code>contour_plots.ipynb</code></td></tr>
<tr><td>5.</td><td>Contour plots</td><td>Exercises</td><td><code>exercises/contour_exercise_*.ipynb</code> [* =1-2]</td></tr>
<tr><td>6.</td><td>Regridding data</td><td>Walkthrough</td><td><code>cf-python_regridding.ipynb</code></td></tr>
<tr><td>7.</td><td>Regridding data</td><td>Exercises</td><td><code>exercises/regridding_exercise_*.ipynb</code> [* =1-2]</td></tr>
<tr><td>8.</td><td>Vector and line plots</td><td>Walkthrough</td><td><code>vector_line_significance.ipynb</code></td></tr>
<tr><td>9.</td><td>Wrap up</td><td>Summary</td><td><code>wrap_up.ipynb</code></td></tr>
</table>
</td>
<td>Slides</td>
<td>Video</td>
</tr>
</table>

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

3.3.4.4 Comments from the Survey

Some of the comments we received from the survey were: “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

3.3.5.1 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.

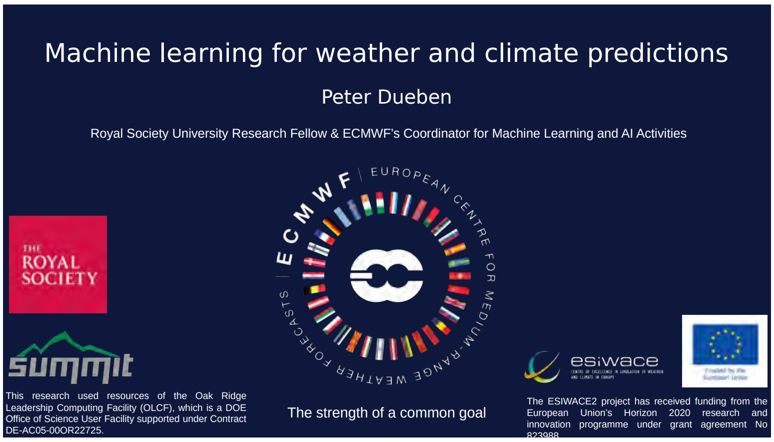
This talk provided an introduction to machine learning, outlined how to apply machine learning in Earth System modelling, showed examples for the application of machine learning throughout the weather and climate modelling workflow, and discussed the challenges that will need to be tackled.

3.3.5.2 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

3.3.5.3 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

3.3.6.1 Computer Hall Tour

The aim of this session was to learn about the performance and specifications of the ECMWF High-Performance Computing Facilities, and the way the supercomputer is used for operations, storage and research by ECMWF and its 34 Member and Co-operating States.

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

3.3.6.2 Weather Room Tour

This tour introduced participants to ECMWF Forecasting products and activities. A member of the ECMWF Forecasting team introduced to the participants maps, charts and plots that are produced daily in the “Weather Room” for weather prediction and analysis.

3.3.6.3 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

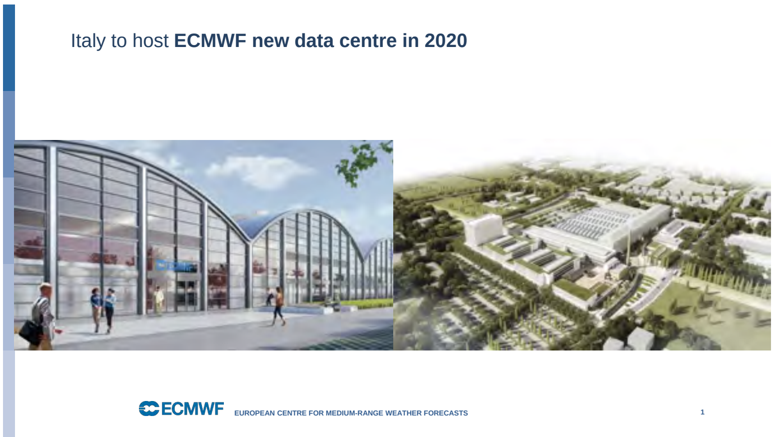
 	Slides	Video

Table 3.21: Computer Hall Tour

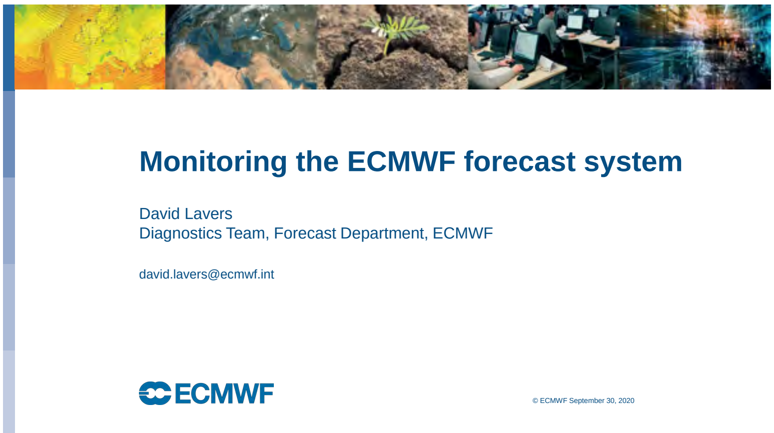
 	Slides	Video

Table 3.22: Weather Room Tour

3.3.6.4 Comments from the Survey

Some of the comments we received from the survey were: "That was very interesting."

3.3.7 High-Performance Data Analytics and Visualisation

3.3.7.1 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 discussed 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 were presented.

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

3.3.7.2 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

3.3.7.3 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 Visualisation 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

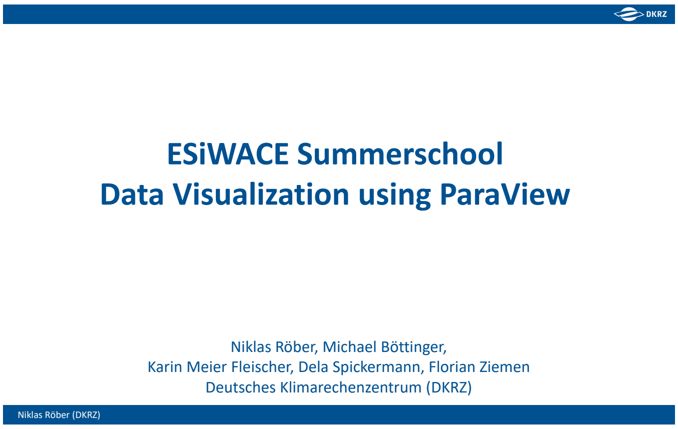
	Slides	Video

Table 3.23: Data Visualization Using ParaView

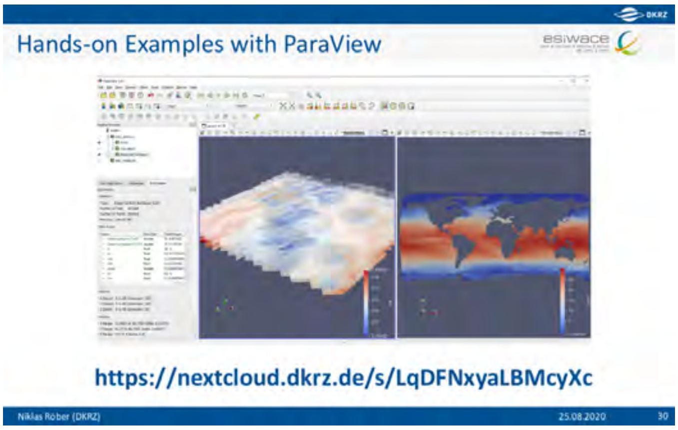
	Slides	Video

Table 3.24: Hands-on: Data Visualisation using ParaView

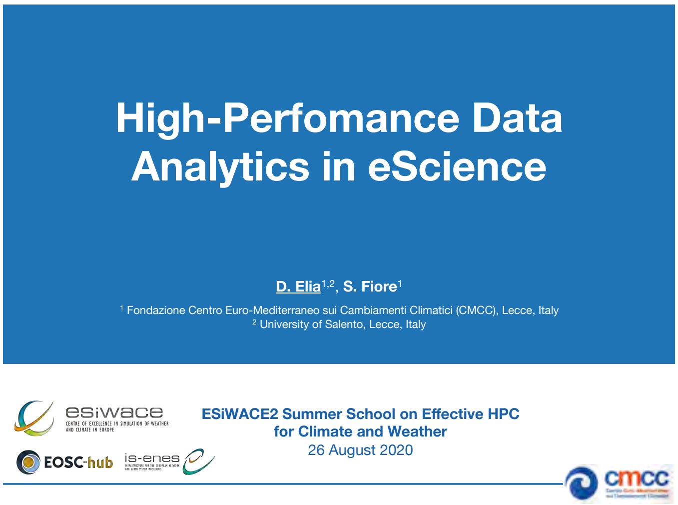
	Slides	Video

Table 3.25: High-Performance Data Analytics and Visualisation

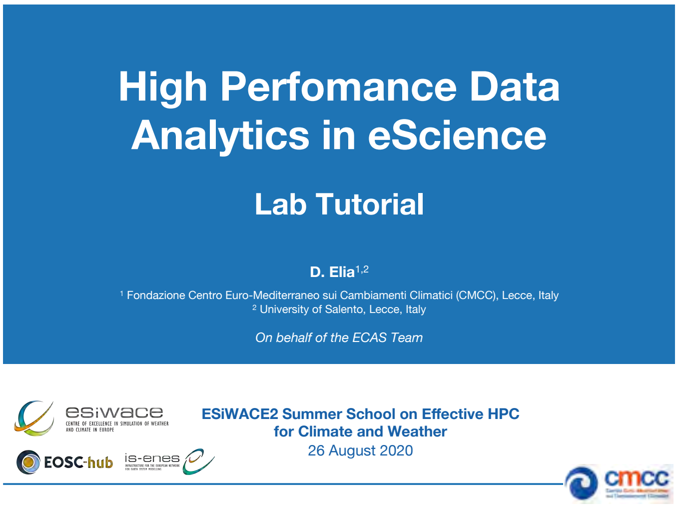
	Slides	Video

Table 3.26: High-Performance Data Analytics and Visualisation

3.3.7.4 Comments from the Survey

Some of the comments we received from the survey were: “This session was very good explained and very easy to follow.”

3.3.8 Performance Analysis

3.3.8.1 Abstract

3.3.8.2 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

3.3.8.3 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

	Slides	Video

Table 3.27: Computational Profiling Analysis for Climate and Weather

	Slides	Video

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

3.3.9 Containers

3.3.9.1 Abstract

This session presented an introduction to an end-to-end scientific computing workflow utilising Docker containers. Attendees learnt about the fundamentals of containerisation and the advantages it brings to scientific software. Participants then familiarised 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 was given to software resources that enable highly-efficient scientific applications, like MPI libraries and the CUDA Toolkit.

The second part of the lecture focussed 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.

3.3.9.2 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 that drove the creation of HPC-focused container solutions
- Highlight differences and similarities between Docker and Sarus

3.3.9.3 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	

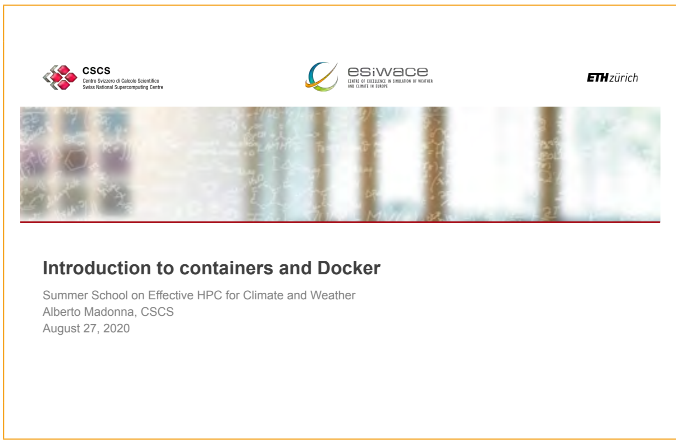
	Slides	Video

Table 3.29: Introduction to Containers and Docker

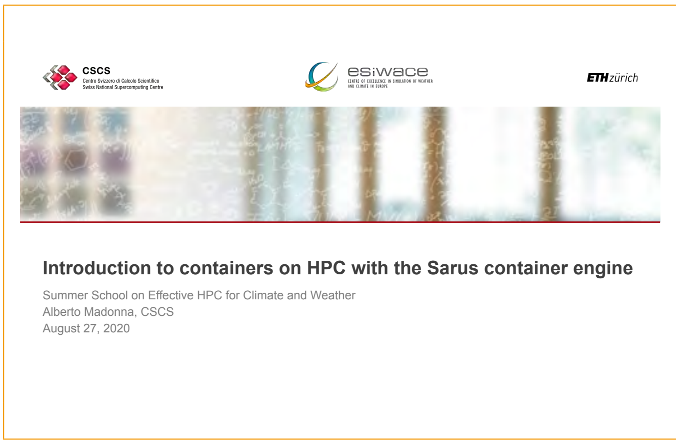
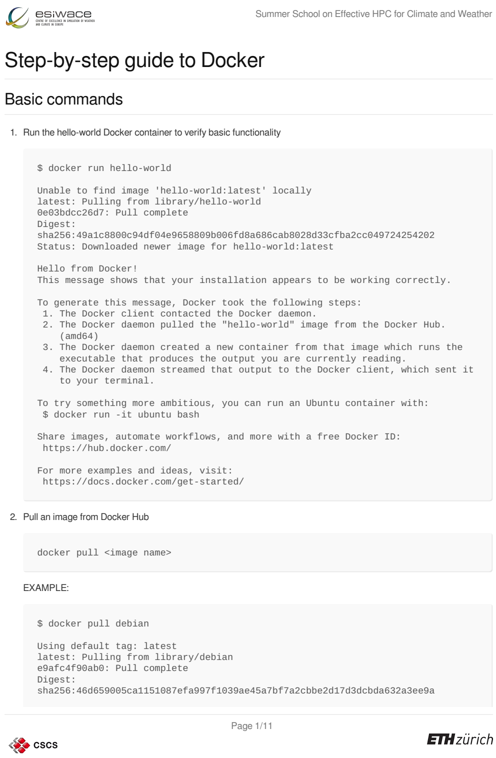
	Slides	Video

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

3.3.10 Q&A Session

3.3.10.1 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.

3.3.10.2 Sessions

Q&A Session		
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

3.3.10.3 Comments from the Survey

Some of the comments we received from the survey were: "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 DYAMOND and DYAMOND2.

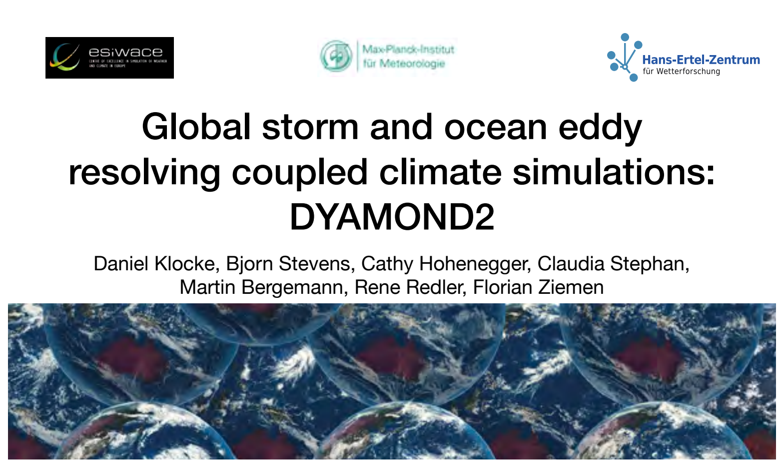
	Slides	Video

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 managed 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 fill out a simplified Google Form to register for the event. All applicants were added to a mailing list to facilitate the communication between them and the organisation, and this list will now be used to advertise future related 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 the most participants were: UK (19.8%), India (12.3%), and Germany (11.1%) (Figure 3.1).

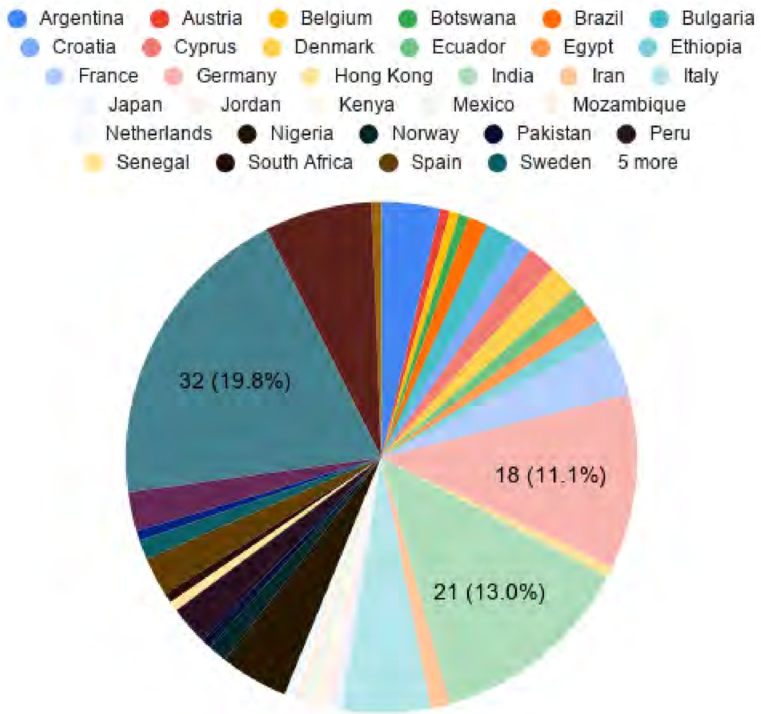


Figure 3.1: Country of the participants

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).

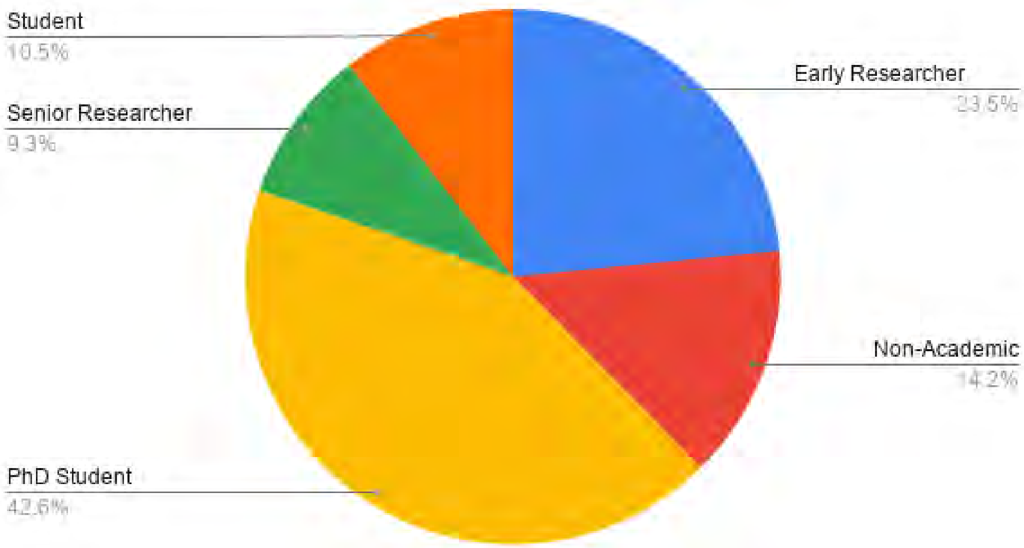
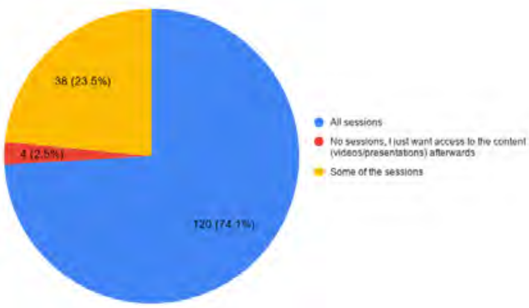
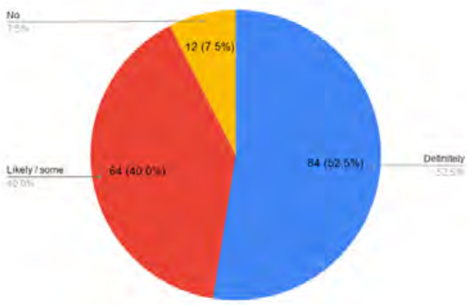


Figure 3.2: Occupation of the participants

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).

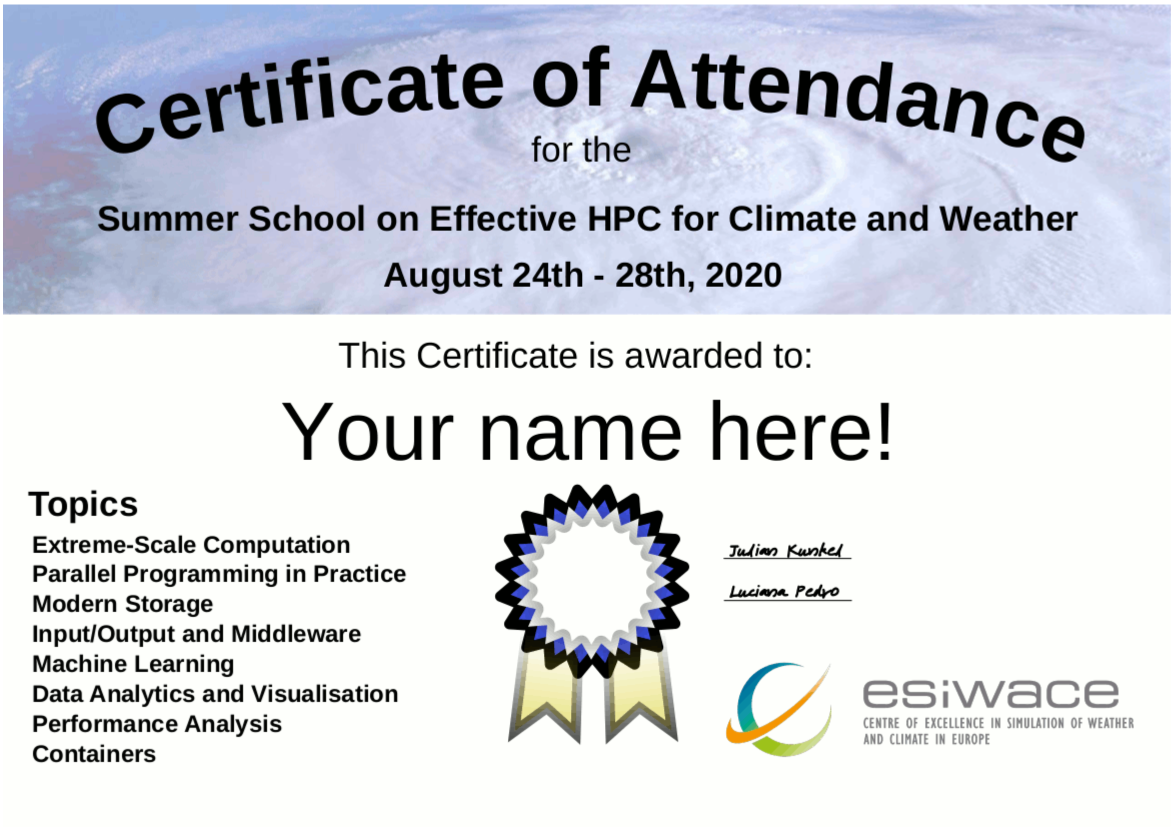


Figure 3.4: Certificate of Attendance

3.6 Survey

After the event, we asked the participants to fill out 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 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:
  - "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

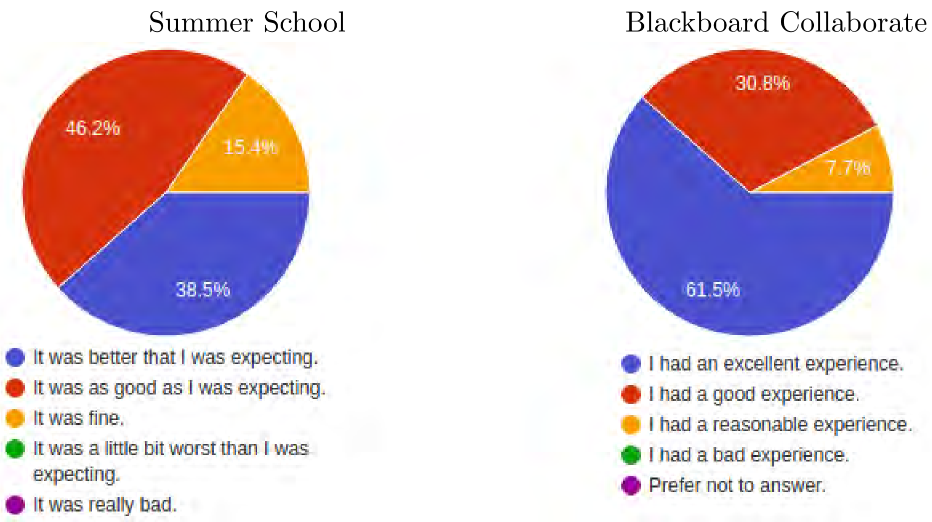


Figure 3.5: General assessments

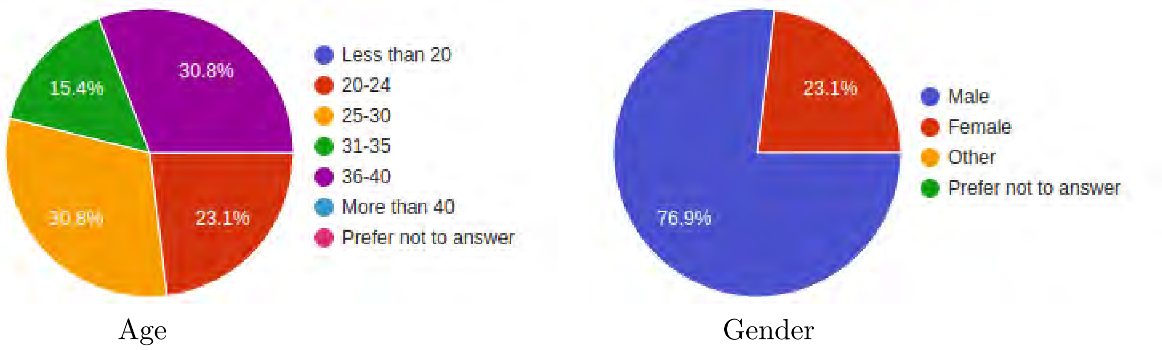


Figure 3.6: Personal information

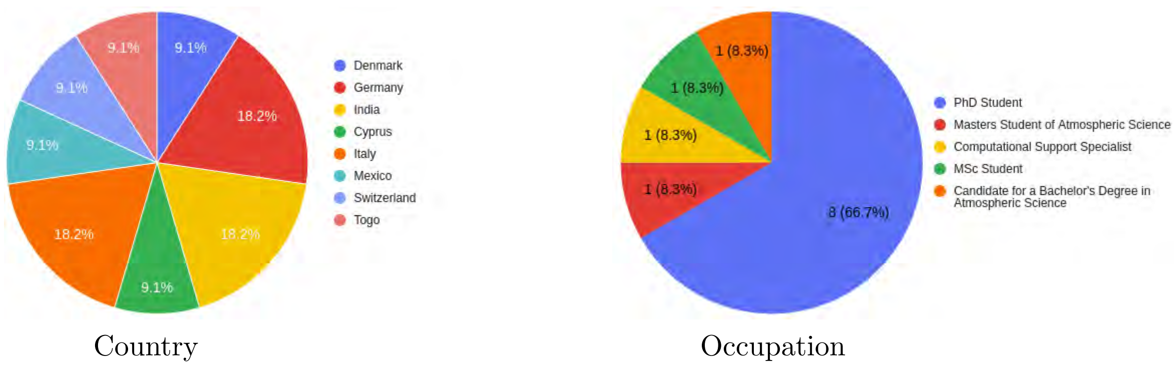


Figure 3.7: Personal information

3.6.2 Comments

3.6.2.1 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.”

3.6.2.2 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."

3.6.2.3 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 used to improve the event in 2021. Even though we had hoped to have the event in Reading, UK, due to the COVID-19 the 2021 event was held virtually, but we applied some lessons from the 2020 event regardless.

- 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 organisers and the attendees, the expected secondary benefit (support the attendees) was not achieved.

We had hoped that the list would 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.

Encouraging 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-togethers at the end of the day. When transforming 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 face-to-face interactions.

  Here we can mention the mailing list again, which gave attendees the opportunity to contact one another and improve their personal network. For the 2021 event, similar virtual opportunities were used to facilitate the interaction among attendees and speakers.

- Virtual Machine

  The Virtual Machine (VM) was a fantastic idea that allowed all the participants to work with tools developed and used in the ESiWACE project in the same environment. By doing that, we saved time that would have been used for configuring each machine and for dealing with operating systems specifications. However, not everyone was able to install the VM. Without the VM installed, it was nearly impossible to follow the labs hands-on.

  Again, for the event in Reading, we had been planning to have physical labs in which the system would have already been appropriately configured for all the attendees to focus their attention only on 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 on 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 planned to use 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 intended to guarantee that everyone would have the VM adequately installed and running, and they would 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 being 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 re-connection.

The organisers decided to evaluate other platforms in case the summer school 2021 would also be designed as an online event. The 2021 event was also held virtual and similar tools were used.

- 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 short 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. A similar structure was used for the 2021 event.

  - "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. Similar strategy was used for the virtual 2021 event.

  - "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 contact data and offered to provide further help during and after the event.

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

    The 2021 summer school was also held virtually.

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

4.1 Introduction

The Summer School on Effective HPC for Climate and Weather – 2021 was a virtual event held on August 23 - 27, 2021. It is the second time the summer school was affected by COVID-19 restrictions after the 2020 event. The scope of the summer school is the training of young researchers and software engineers in methods, tools, and theoretical knowledge to make effective use of HPC environments and generate insights.

Not only does the school aim to prepare the attendees for large scale simulation runs and data processing, it also covers a representative selection of modern concepts such as machine learning, domain-specific languages, containerisation, and analysis of climate/weather data using Python.

In the summer school, an outlook of challenges and strategies for HPC for climate and weather was provided. At the end of the school certificates of attendance were awarded to participants with over sixty percent attendance.

The summer school was funded by the ESiWACE2 project and supports the mission of the European Network for Earth System modelling - ENES.

Sections 4.2 and 4.3 of this chapter describe the structure and topics covered in the 2021 event. As the general structure of the 2020 event proved effective and, unfortunately, had to use the virtual format again, we kept the general structure. Hence, the following descriptions are identical or slightly improved versions over the 2020 event. They are here for completion.

4.2 Structure

The Summer School comprised topical sessions in the morning and afternoon. A topical session organised by experts in respective fields consisted of an academic lecture and hands-on and lab practicals, group work, and discussion.

The hands-on tutorials and lab practicals organised by the experts provided an introduction and/or walk-through to each topic. Most of those were live sessions. However, some experts had pre-recorded them to allow independent or self-paced learning, and participants had the opportunity to decide what and when they wanted to engage with the experts. The Q&A session on Friday (the last day of the Summer School) offered the participants the opportunity to contact the organisers of each of the hands-on sessions and ask questions regarding the lab practicals.

Virtual Machines (VM) were provided together with a tutorial to set them up. The VM included Ubuntu with all the software needed for the training pre-installed. Participants could also install the VM on their PC to perform most of the training tasks. On Friday, a Q&A slot was scheduled for all lab practicals.

The hands-on sessions also offered participants the opportunity to contact the session organisers, to ask questions on the topics, tutorials and exercises. Additional support was provided by the mailing list which participants could post questions and cooperate with other students and organisers. Some students organised themselves into study groups.

4.3 Topics Covered

The following topics were covered in the summer school:

- Extreme-Scale Computation – Section 4.3.1
- Parallel Programming in Practice – Section 4.3.2
- Modern Storage – Section 4.3.3
- Input/Output and Middleware – Section 4.3.4
- Machine Learning – Section 4.3.5
- ECMWF – Virtual Visit – Section 4.3.6
- High-Performance Data Analytics and Visualisation – Section 4.3.7
- Performance Analysis – Section 4.3.8
- Containers – Section 4.3.9
- Q&A Session – Section 4.3.10
- Keynote Talk – Section 4.3.11

4.3.1 Extreme-Scale Computation

4.3.1.1 Abstract

This session introduced 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 were discussed.

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

Lastly, a relatively new, Domain-Specific Language (DSL), approach to programming Weather and Climate models was introduced with examples from two existing DSLs – DAWN and PSyclone. It was 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.

4.3.1.2 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

4.3.2 Parallel Programming in Practice

4.3.2.1 Abstract

In this session, participants received a global overview of the implementation of the main concepts of parallel programming in weather and climate codes. The different parallel programming models for distributed and shared memory systems were detailed, and the resulting scalability of commonly-used algorithms implementing those models was described.

Particular attention was devoted to specific features that may inhibit scaling and performance of weather and climate codes. This analysis was 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.

4.3.2.2 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 frequently used coupling software in climate and weather applications

4.3.3 Modern Storage

In this session implications of modern storage architectures and object stores among others were presented.

4.3.3.1 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

4.3.4 Input/Output and Middleware

4.3.4.1 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 superfluous attempts to tune the application and optimise performance in a wrong layer of the stack.

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

4.3.4.2 Learning Objectives

- 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

4.3.5 Machine Learning

In this session, application of Machine Learning in weather and climate sciences and modelling was presented and discussed.

4.3.5.1 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 being 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.

The talk provided an introduction to machine learning, outlined how to apply machine learning in Earth System modelling, showed examples for the application of machine learning throughout the weather and climate modelling workflow, and discussed the challenges that will need to be tackled.

4.3.5.2 Learning Objectives

- Describe the relevance of Machine Learning and its application to assess 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

4.3.6 ECMWF – Virtual Visit

4.3.6.1 Computer Hall Tour

The session gave participants an opportunity to 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 and Co-operating States.

The presentation included a video tour of the computing facilities currently located in the ECMWF headquarters in Reading and a preview of what the new data centre will look like when it opens in Bologna (Italy) the following year.

4.3.6.2 Weather Room Tour

This session gave participants an opportunity to learn about ECMWF Forecasting products and activities. A member of the ECMWF Forecasting team introduced participants to the maps, charts and plots that are produced daily in the "Weather Room" for weather prediction and analysis.

4.3.7 High-Performance Data Analytics and Visualisation

4.3.7.1 Abstract

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

In this session, some of the main challenges concerning scientific data management and in particular, those related to data analytics and visualisation were discussed. 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 were also presented.

The lab tutorial provided a more practical introduction to some tools and modules for data analysis and for applying these to climate data, as well as a walk-through of the Vendor Managed Inventory (VMI) for the virtual lab.

4.3.7.2 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

4.3.8 Performance Analysis

In this session the fundamentals of performance analysis among other performance related topics were presented with examples of applying Barcelona Supercomputing Center (BSC) performance tools.

4.3.8.1 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

4.3.9 Containers

4.3.9.1 Abstract

This session presented an introduction to an end-to-end scientific computing workflow utilising Docker containers. Attendees learnt about the fundamentals of containerisation and the advantages containerisation brings to scientific software. Next, participants familiarised themselves 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 was given to software resources that enable highly-efficient scientific applications, such as MPI libraries and the CUDA Toolkit.

The second part of the lecture focused 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.

4.3.9.2 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 that drove the creation of HPC-focused container solutions
- Highlight differences and similarities between Docker and Sarus

4.3.10 Q&A Session

4.3.10.1 Abstract

One of the Friday's sessions was dedicated to a joint Q&A for all lab practicals. The session offered the participants the opportunity to contact the organisers of all hands-on session and ask questions regarding the topics presented together with tutorials and exercises.

4.3.11 Keynote Talk

For the 2021 event, the keynote included two invited talks. Miguel Castrillo (BSC), Dorotea Iovino (CMCC), and Clement Bricaud (Mercator Ocean) presented *"Preparing NEMO and EC-Earth models for very high-resolution production experiments"*, and Sebastian Paczkowski from University of Göttingen presented *"Project FORESTCARE: An HPC application case study for a high spatial and temporal resolution forest vitality assessment"*.

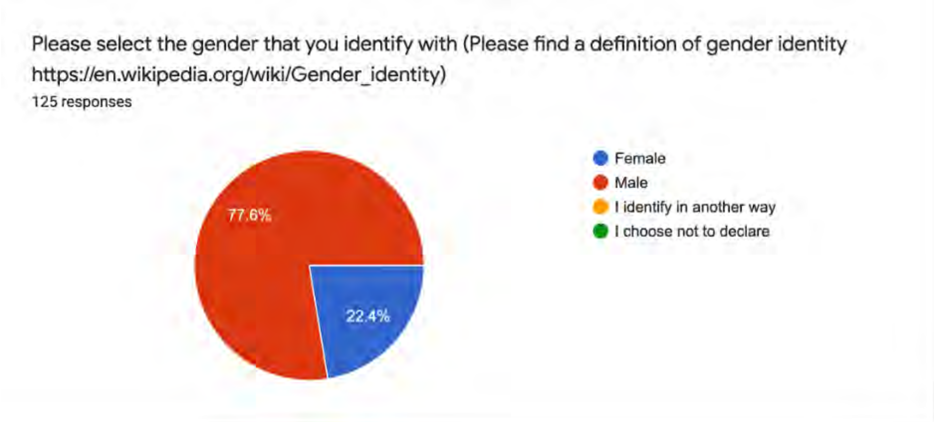
4.4 Survey

4.4.1 Pre-event Survey and Registration

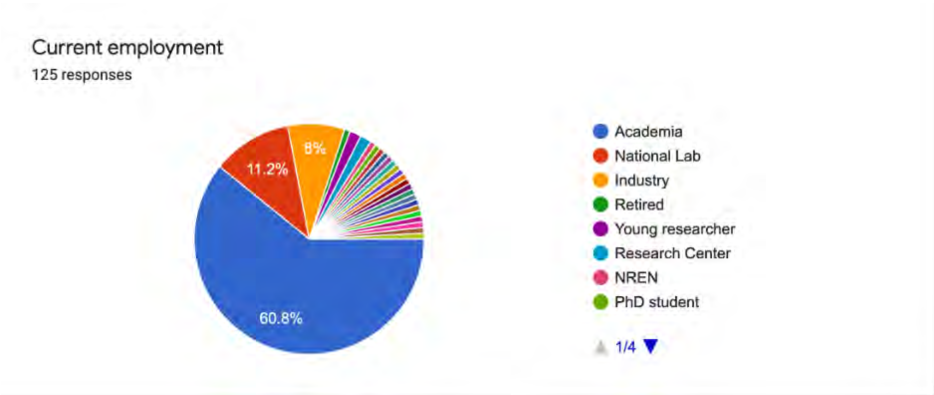
During registration, participants were asked to participate in a survey which was sent to the event's mailing list. The survey included 10 questions, and we have analysed those with completed answers. In the following subsections we give a summary of the survey results with some statistical numbers and comments on the summer school.

4.4.1.1 Statistics

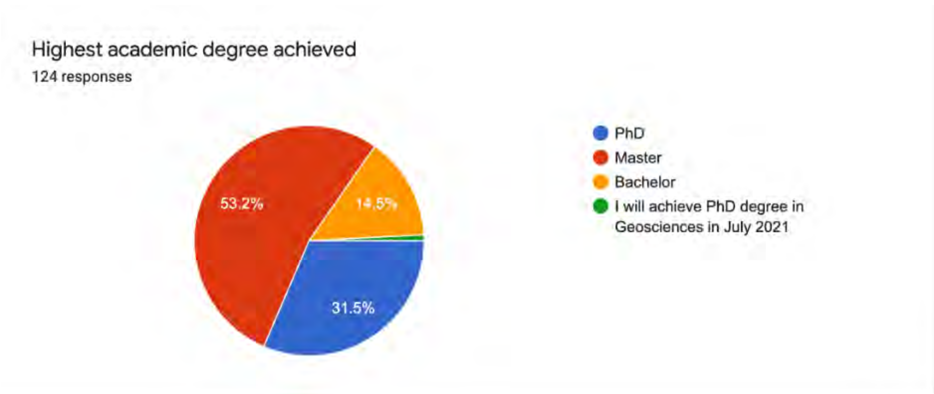
We had more registrations from males than from females for the summer school. Out of a total of 125 registered participants 77% identified as male and 22% as female, as shown in the following figure.



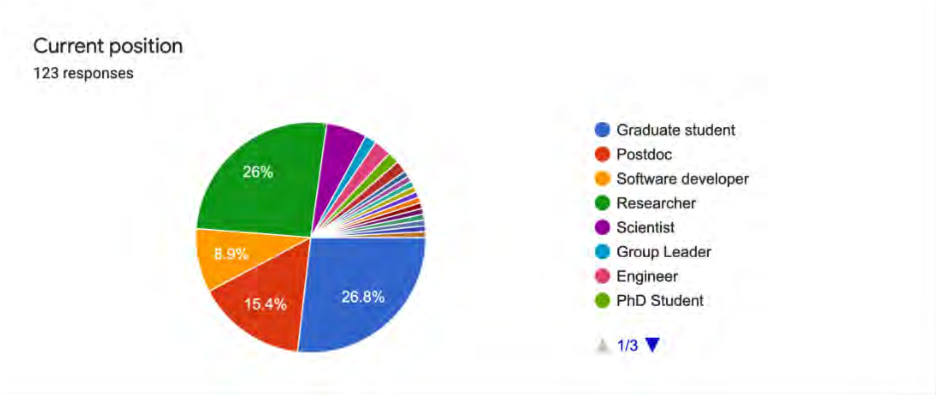
Of these participants more than 60% were employed in or affiliated to academia with another 11% and 8% affiliated to national laboratories and industry respectively as shown in the figure below.



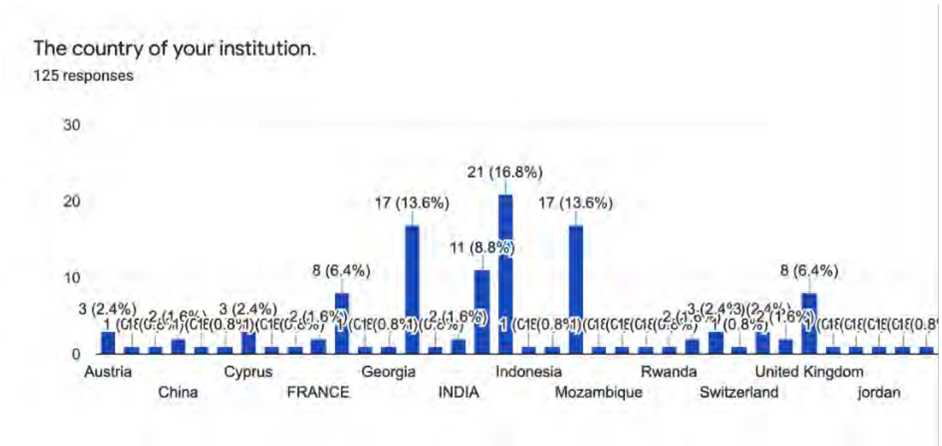
Many of the attendees had high qualifications with about 53% having obtained a master's degree and 31% having obtained PhD. There rest either had a bachelor's degree or were preparing to obtain their PhDs in 2021 as shown in the following figure.



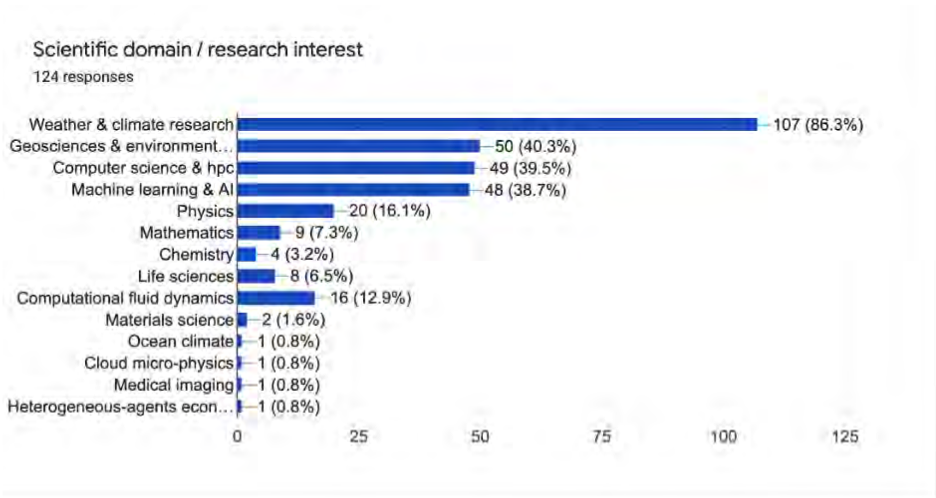
Many of the attendees were either graduate students or researchers each representing 26% of the total number of participants. About 15% were postdoctoral researchers and about 9% were software developers.



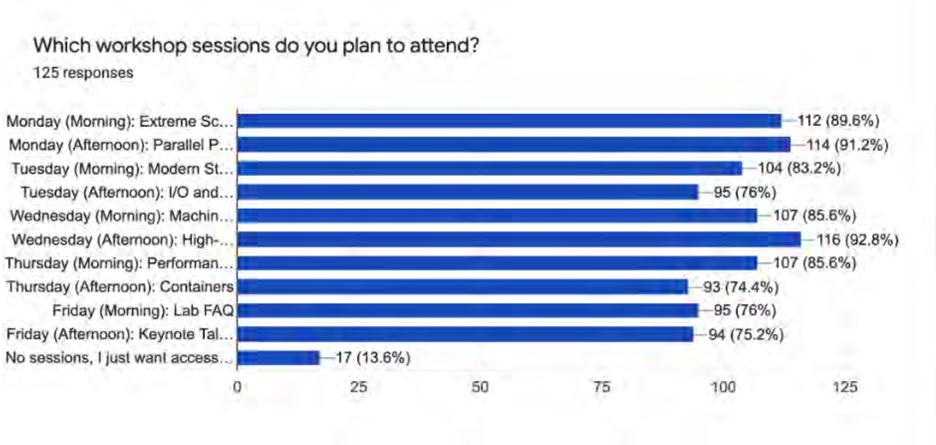
The participants were affiliated to different institutions from around the world with about 17% being in Indonesia, 13% each in Germany and Italy, 8% in India, 6% each in France and United Kingdom, 2% each in Austria and Cyprus. In total we had participants from institutions in 36 different countries as shown in the following figure.



About 86% of the attendees stated weather and climate research as their scientific domain or research interest as depicted in the following figure. Another 40% selected geosciences and environmental sciences, and yet another 39% mentioned either computer science and hpc or machine learning and AI as their scientific domain or research interest.



As shown in the following figure, many attendees participated in the Wednesday's afternoon session of High Performance Data Analytics and Visualisation and the Monday afternoon session on Parallel Programming in Practice with over 90% planning to attend both sessions. 13% of the attendees, however, had no session preference and just wanted to access the school's contents afterwards.



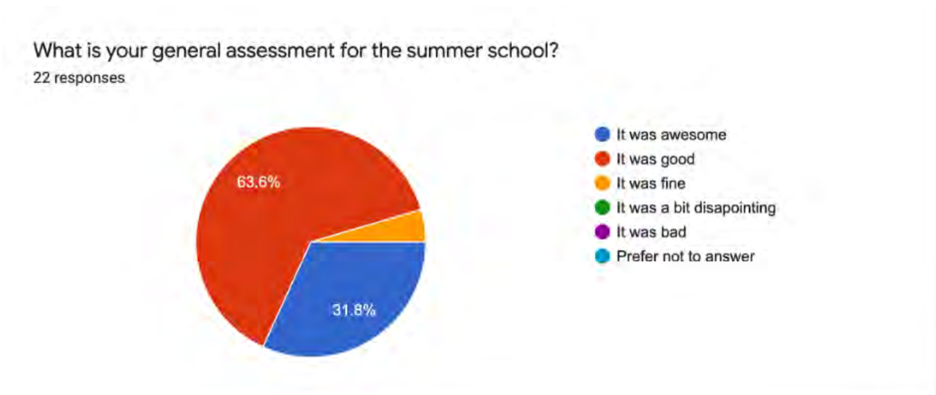
4.4.1.2 Comments

A number of attendees mentioned in the survey's comment section that they wanted to attend the school because they were setting up numerical weather prediction systems and platforms in their institutions. The summer school would give them worthwhile experience and prepare them to work with their own institutional computing facilities. Others also said that they were currently PhD students and that the summer school would be helpful in their PhD work.

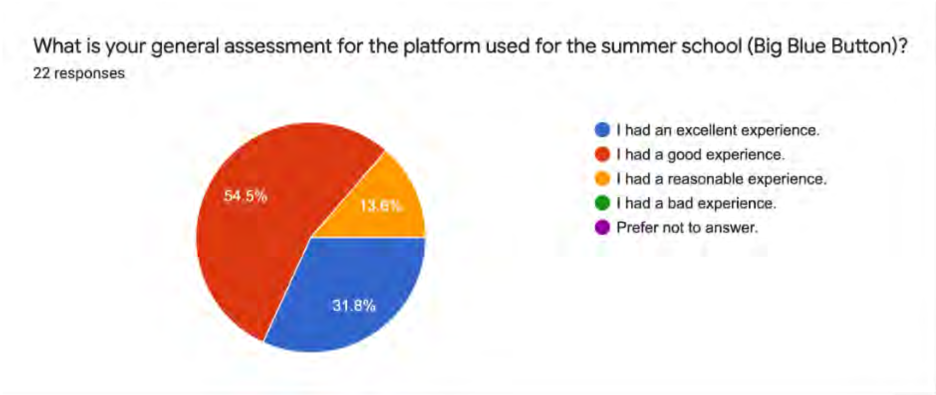
4.4.2 Post-event Survey

4.4.2.1 Statistics

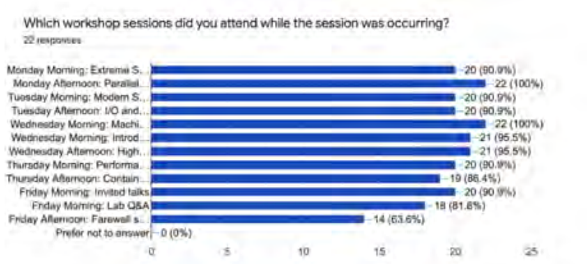
At the end of the summers school, participants were asked to give a feedback on the sessions via a survey. i.e. a response rate of 17.6% (22 out of 125 registered attendees) submitted their feedback on the school, with about 63% saying that they had a good experience and 31% saying they had an awesome experience as shown in the following chart.

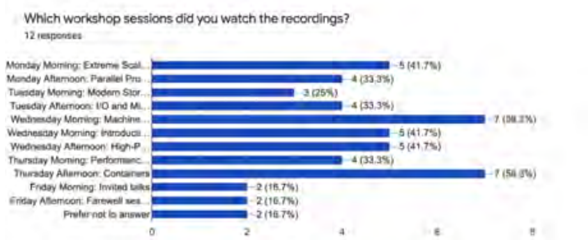


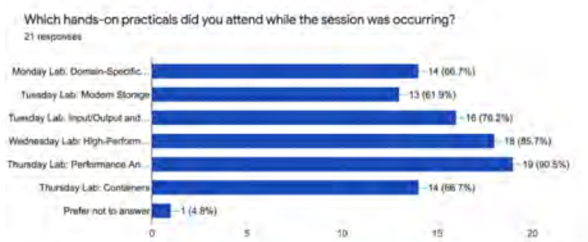
Regarding the platform used for the summer school, the Big Blue Button, 31% of the participants said that they had an excellent experience while 54% and 13% said they had good and reasonable experience respectively as shown in the following chart.

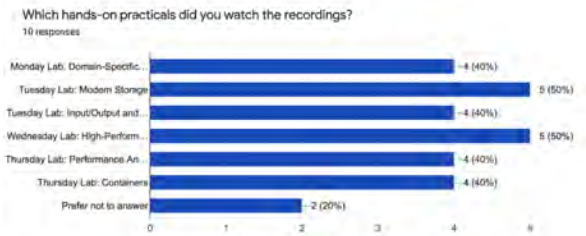


Majority of participants attended live sessions and hands-on sessions, and a minority watched recorded sessions and hands-on trainings. As shown in the following figures, among those answered the survey, overall attendance was over 90% for all sessions except for Lab Q&A, which was 80%, and the farewell session, which was 63%.









4.4.2.2 Comments

Of the 22 participants who participated in the post-event survey, 16 said what they liked about the summer school, for example, one stated "As someone who is more on the user side of weather models and not on the creation side, I liked to get an overview of what the problems and challenges are. I also liked to get more insight on how machine learning is used in NWP". Others praised the interesting content and good learning materials. 12 participants also provided suggestions for improvements many of which were technical issues with the internet connection, problems with the Virtual Machine, limited question and discussion time, non-interactive sessions where participants could not see each other, too much computer science material as opposed to meteorology and physics, as well some hard-to-follow lab sessions.

A reason for the good results might be that people who had bad experiences didn't fill out the survey.

4.5 Summary

The 2021 summer school was organised similarly to the 2020 summer school due to the COVID restrictions in both years. It was unfortunate that we could not offer the collaborative projects that we planned for the in-person event. Therefore, we aimed to exploit the virtual format as much as possible and created videos of the sessions to allow even more attendees to enjoy the training. The lecturers offered similar topics as in the 2020 event, but with revised and expanded materials. The sessions included 16 hours of lectures and 8 hours of hands-on and lab practicals.

The summer school attracted participants from all over the world with significantly varying scientific backgrounds and levels of qualifications. Overall, 125 people registered for the school, with some interested in receiving the information about the videos. There were 22 responses from the post-event survey with 95% of the respondents saying they had a good or awesome experience. The attendance was between 30 and 50 persons per session affiliated to institutions in 31 countries. 33 of the attendees received a certificate of attendance because they attended more than 60% of all sessions. Thus, we conclude the lectures and lab practicals were generally well-received and a successful replacement for the in-person summer school, which had been planned for 30 attendees as well.

5 Training Material with Open Access

This chapter summarises the material published in the Summer School on Effective HPC for Climate and Weather. It is provided with open access. Note that as part of the ESiWACE project, some OER materials will be published as well but this is outside the scope of the summer school.

5.1 Summer School 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

5.2.1 Extreme-Scale Computation

Extreme Computing Session – Overview	Slides	Video

Extreme Computing	Slides	Video

Introduction to DSLs	Slides	Video

A Brief Introduction to PSyclone	Slides	Video

Dawn DSL	Slides	Video

PSyclone Tutorial Introduction	Slides	Video

5.2.2 Parallel Programming in Practice

Scaling Algorithms	Slides	Video

Code Coupling | Slides | Video

5.2.3 Modern Storage

Modern Storage 1 | Slides | Video

Modern Storage 2 | Slides | Video

Darshan Demonstration – Hands-on Session | Slides | Video

Modern Storage: Lab Practicals Intro | Slides | Video

Modern Storage: Lab Practicals | Slides | Video

5.2.4 Middleware and File Formats

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

5.2.5 Machine Learning

Machine Learning for Weather and Climate Predictions | Slides | Video

5.2.6 ECMWF – Virtual Visit

Overview of ECMWF | Slides | Video

Computer Hall Tour | Slides | Video

Weather Room Tour | Slides | Video

5.2.7 High-Performance Data Analytics and Visualisation

High Perfromance Data Analytics in eScience | Slides | Video

Lab Practical	Slides	Video

Data Visualization using Paraview	Slides	Video

5.2.8 Performance Analysis

Computational Profiling Analysis for Climate and Weather	Slides	Video

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

5.2.9 Containers

Introduction to Containers and Docker	Slides	Video

Lab Session	Slides	Video

5.2.10 Keynote and Conclusions

Preparing NEMO and EC-Earth models for very high-resolution production experiments	Slides	Video

Project FORESTCARE	Slides	Video

Disccussion, Feedback, and Closing Remarks	Slides	Video