September 3, 2021

WEATHER CLIMATE WATER
TEMPS CLIMAT EAU

Exascale Computing & Data

Opportunities and challenges for weather, climate and environmental prediction

Kris Rowe

Argonne Leadership Computing Facility

Mark Govett

NOAA, Global Systems Laboratory

WMO OMM

World Meteorological Organization
Organisation météorologique mondiale

WORLD METEOROLOGICAL ORGANIZATION

NOAA

Met Office

ECMWF

Environment and Climate Change Canada
Environnement et Changement climatique Canada

JÜLICH Forschungszentrum
JÜLICH SUPERCOMPUTING CENTRE

Argonne NATIONAL LABORATORY

University of Reading

気象庁 Japan Meteorological Agency

ISAC

中国气象局 CHINA METEOROLOGICAL ADMINISTRATION

NCAR

AIMS

GEOMAR Helmholtz Centre for Ocean Research Kiel

The Team

WMO Research Board Task Team on Exascale, Data, and AI

Mark Govett	NOAA
Bubacar Bah	African Institute for Mathematical Sciences
Peter Bauer	ECMWF
Dominique Berod	WMO
Veronique Bouchet	ECCC
Wenchao Cao	WMO
Susanna Corti	Institute of Atmospheric Science and Climate
Chris Davis	NCAR
Tim Graham	UK Met Office
Yuki Honda	WMO
Adrian Hines	UK Met Office
Junishi Ishida	Japan Meteorological Agency
Bryan Lawrence	University of Reading
Kris Rowe	Argonne Leadership Computing Facility
Martin Schultz	Jülich Supercomputing Centre
Martin Visbeck	Helmholtz Centre for Ocean Research
Keith Williams	UK Met Office
Yihong Duan	China Meteorological Administration

WMO OMM

The Big Picture

- The End (Goals)
- The New Computing
- Technical Challenges
- Critical Gaps
- Next Steps
- Summary

The End

(Goals)

Faster Prediction

- *Strong scaling*
- Weather*

<u>Benefits</u>

- Prevent loss of life and property
- Assimilate more recent data into predictions

Higher Fidelity

- *Weak scaling*
- Climate*

<u>Benefits</u>

- Model coupled systems
- Resolve critical small-scale processes
- Improve accuracy of long-time simulations

THE NEW COMPUTING

Form follows function

WMO OMM

What is "Exascale"?

$\geq$1 Exaflops/s (double precision) on the HPL benchmark

Compute + Memory + Network = Exascale?

Compute	Memory	Network
What about "AI Flops"?	CPU? GPU? Total?	Does cloud count?

The rigorous definition is not as important as the dramatic change in computer architecture!

Examples of Exascale Systems

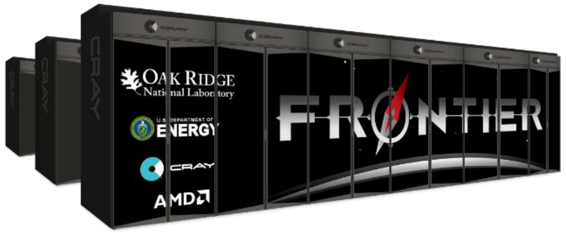
Key Details

- Multi-CPU + Multi-GPU
- Multiple HPC GPU vendors
- Network and storage must complement compute

ALCF

- >1 Exaflops/s
- Per node:
  - 2 Intel "Sapphire Rapids" CPUs
  - 6 Intel "Pointe Vecchio" GPUs
  - 8 fabric endpoints (NICs)
- Cray Slingshot
  - 3 hop Dragonfly topology
- DAOS Storage: ≥230 PB, ≥25 TB/s

OLCF



- >1.5 Exaflops/s
- Per "blade":
  - 2 AMD EPYC CPUs
  - 8 AMD Radeon Instinct GPUs
  - Multiple fabric endpoints (NICs)
- Cray Slingshot
  - 3 hop Dragonfly topology
- Lustre Storage: ≥250 PB, 5-10 TB/s

WMO OMM

Why this architecture?

Multiple GPUs	HBMe	Multiple NICS

Design Requirements

Flops, Bandwidth	Power Envelope	Cost Constraints

TECHNICAL CHALLENGES

New computers, old problems

WMO OMM

New Computers, Old Problems

- Traditionally end-users of prediction models focus on scientific challenges
- Concurrent with these are numerous technical challenges related to hardware, software, and human-factors
- Many of these challenges are not new, however, their difficulty and complexity are amplified in the exascale context
- Challenges are not independent: addressing or failing to address one may reduce or increase the difficulty of another

Challenge #1

Cost

Hardware

- U.S exascale systems in the range $300m-$600m
- Electricity costs for 30 MW are more than $12m USD
- Facilities and maintenance costs are also significant

Software

- Development and maintenance costs are often overlooked
- Funding a team of research software engineers can cost millions/yr.
- Example: ECP has made significant investment in this area

Environmental Impact

- Carbon footprint for 30 MW is over 100 Mt per year!
- EU Green New Deal requires data centres to be carbon neutral by 2030

Challenge #2

Data

- Estimated 0.5 PB storage for 10-day forecast with
  - 3km resolution
  - 192 vertical
  - 3-hour output interval
- Storage for climate simulation will be significantly larger
- Data-in-place strategies now are fundamental
- Data loss or corruption must be addressed at this scale
- *In situ* analysis and visualization are essential tools
- With 5G/IOT volume of data to be assimilated will continue to grow

WMO OMM

Challenge #3

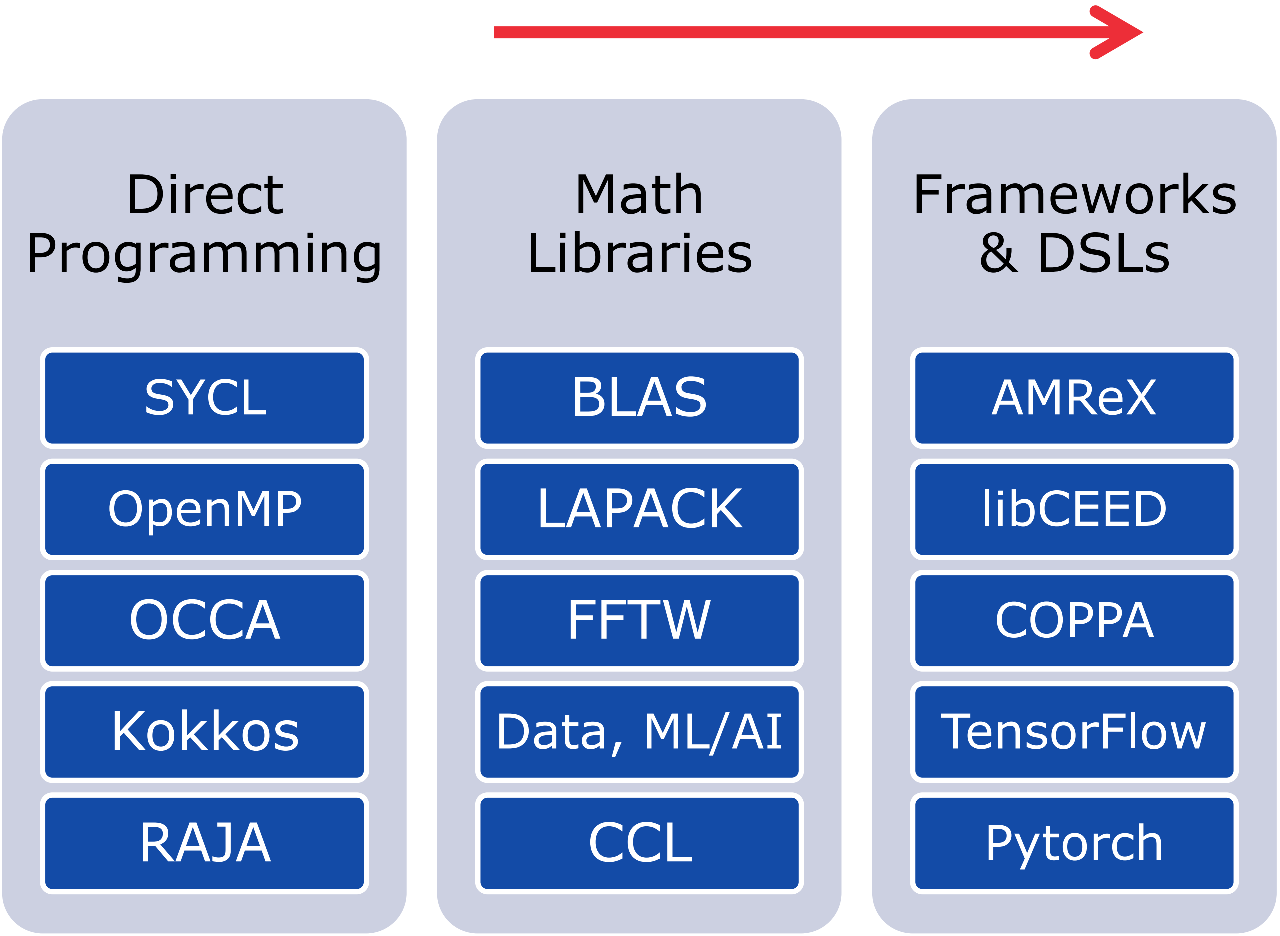
Performance

- Runtime is still the main performance metric
  - *Energy-to-solution* also key consideration
- Internode communication is still an issue for parallel scaling
  - e.g., halo exchange, global reductions
- GPUs require different data-layouts and algorithms which expose more parallelism
  - Performance tuning can be notoriously difficult
  - Subtle differences between vendor microarchitecture can be important
- Denser compute nodes require greater focus on intranode communication and optimization
- Algorithmic changes can sometimes provide the greatest benefit
- Mixed-precision techniques have significant momentum

Challenge #4

Portability

- Portable software can run on
  - Different types of hardware
  - Different vendors' hardware
- Goal is to minimize
  - Lines of source code needed to achieve portability
  - Effort to run existing code on new and future types of hardware
- Want turnkey performance
  - Otherwise with minimal (automatic) parameter tuning



Challenge #5

Productivity

The *Better Scientific Software* website is a great resource from the US ECP

https://bssw.io/

- Ease with which software is developed, tested, shared, maintained, documented
- Following best practices is critical for creating high-quality scientific software
- Software which is modular, composable, and extensible retains greater value, can be more easily ported/adapted
- A co-design approach is optimal
  - Scientists and research software engineers working collaboratively, communicating effectively
  - Examples:
    - ECMWF's Scalability Programme
    - German Climate Computing Center + Max Planck Institute for Meteorology
    - US ECP Co-design Centers: CEED, AMReX, COPA

CRITICAL GAPS

And ways to address them

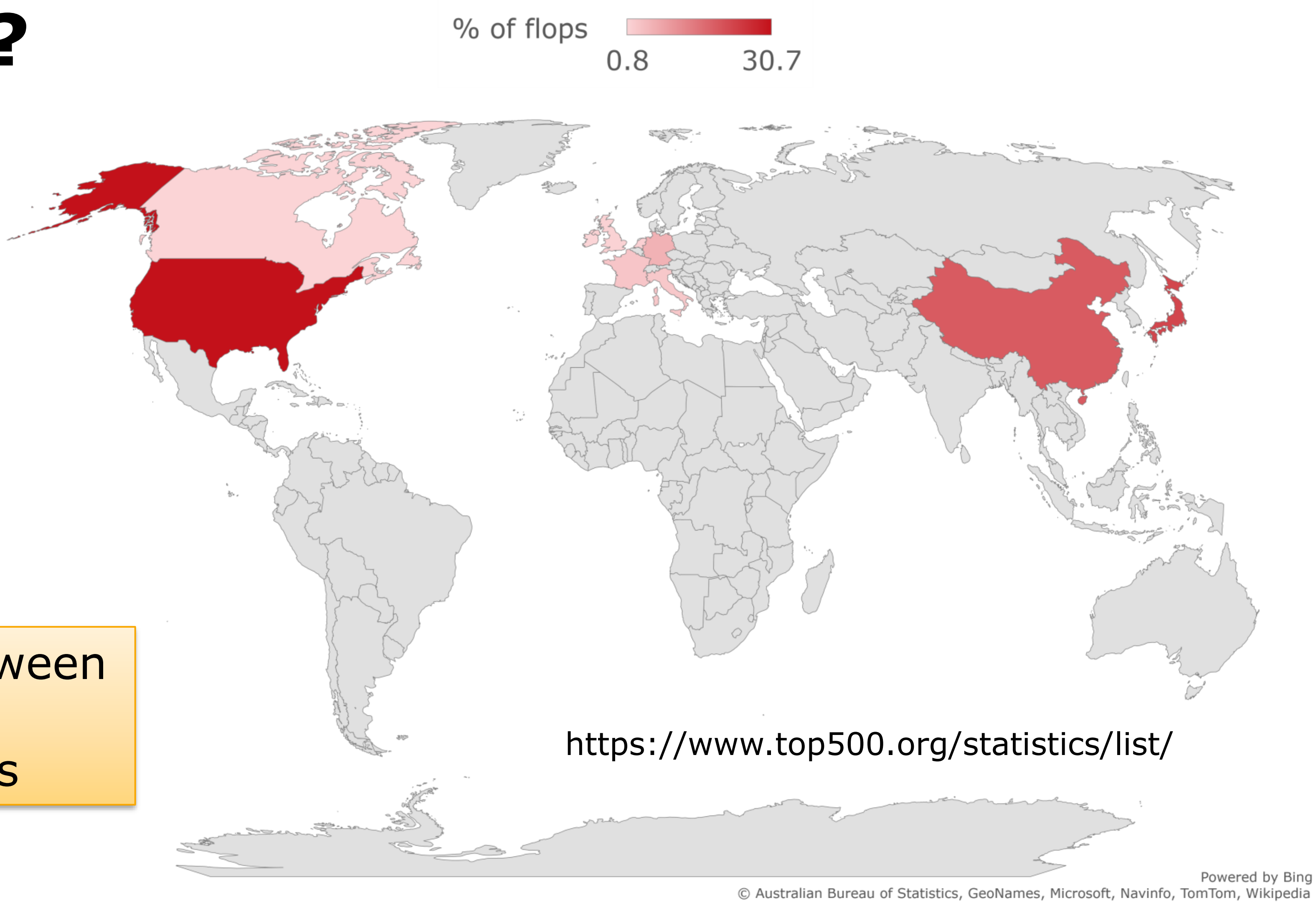
WMO OMM

The Benefits of Exascale—for Who?

- Over 98% of computing power worldwide is in Europe, Asia, and North America
- Over 72% belongs to China, Japan, and the U.S.
- Regions most at risk from climate change have few or no computing resources

Goal: Identify critical gaps between members with and without significant computing resources

TOP 500 Share of Performance



https://www.top500.org/statistics/list/

WMO OMM

Gap #1: Access to Sufficient HPC Resources

Observations

- Many HPC centers can be accessed worldwide
- Remote visualization and data analysis require only a laptop
- Significant software development can be completed remotely
- Compute requirements cannot be avoided!

Potential Solutions

- International allocation programs (e.g., INCITE) exist for research
  - Create awareness of these programs; connect researchers
  - Not available for operation forecasting needs
- Collaboration between Members with & without HPC resources
- Possible service opportunities for industry

WMO OMM

Gap #2: Access to Data Resources, Storage, Analysis Tools

Observations

- Data now often too big to transfer, remains on-site
- Learning to use specialized or niche tools can be a barrier
- Consumers have different use cases, need different aspects
  - Potentially at-odds with *in-situ* analysis, data reduction

Potential Solutions

- Large centers can provide shared external access to
  - Data sets
  - Storage
  - Visualization & analysis nodes
- Advocate for free and open-source community-based tools
- Develop models to accept external analysis "plugins"
- Include at-risk regions in the creation of data standards

WMO OMM

Gap #3: Access to Specialized Knowledge & Skills

Observations

- Developing models for large-scale HPC requires advanced knowledge in
  - applied mathematics
  - computer science
  - software engineering
- Allocation programs require that researchers and their codes can make effective use of resources
- Members with significant computing resources have pipelines to train researchers
  - E.g., workshops, summer schools, internships/fellowships

Potential Solutions

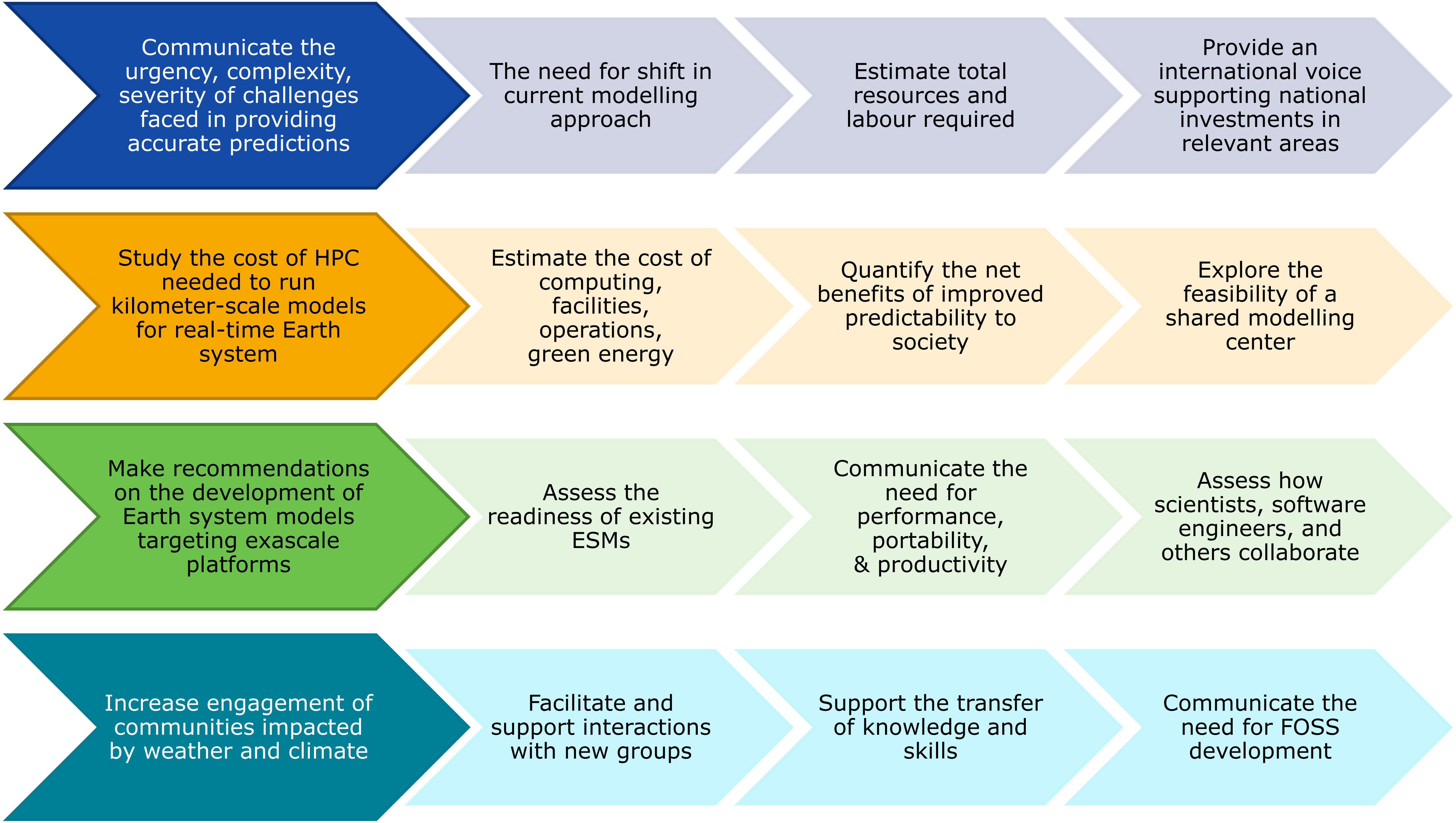
- Create broader awareness of existing resources:
  - Training material from hardware vendors, major computing centers, government funded projects
  - Workshops/summer schools
    - E.g., ATPESC
- WMO develop shared, specialized training resources focused on applying HPC, AI, data techniques in the in the weather/climate
- International technical meeting using ECP Annual Meeting as a prototype

WMO OMM

NEXT STEPS

Concrete actions for the next 12-18 months

WMO OMM


Communicate the urgency, complexity, severity of challenges faced in providing accurate predictions
The need for shift in current modelling approach
Estimate total resources and labour required
Provide an international voice supporting national investments in relevant areas
Study the cost of HPC needed to run kilometer-scale models for real-time Earth system
Estimate the cost of computing, facilities, operations, green energy
Quantify the net benefits of improved predictability to society
Explore the feasibility of a shared modelling center
Make recommendations on the development of Earth system models targeting exascale platforms
Assess the readiness of existing ESMs
Communicate the need for performance, portability, & productivity
Assess how scientists, software engineers, and others collaborate
Increase engagement of communities impacted by weather and climate
Facilitate and support interactions with new groups
Support the transfer of knowledge and skills
Communicate the need for FOSS development

Summary of Recommendations

For the Research Board and WMO Members

We recommend urgency in dedicating efforts and attention to disruptions associated with evolving computing technologies that will be increasingly difficult to overcome, threatening continued advancements prediction capabilities.

The increasing scientific and computing complexity will require major efforts to adapt or rewrite earth system prediction models. In addition to scientific accuracy, models must be developed for performance, portability, and productivity.

The cost of computing resources, power consumption, and the related carbon footprint must be considered along with the benefit of improved predictability. Requirements to make data centers carbon neutral are already in force in a growing number of countries.

Scientists, model developers, computer scientists and software engineers need to work as equal partners on design, development, and maintenance of applications to overcome scientific, computing, and data challenges.

A data-in-place strategy is needed to support the increase in data volume from observations, model and ensemble output, and post processing. This will require co-location of HPC and data, with methods to access, extract, analyze, visualize, and store data by requesting processes & users.

WMO OMM

Summary of Recommendations

For the Scientific Advisory Panel

Few organizations will be able to fully address the software and data handling challenges, let alone provision the necessary supercomputing to continue to increase the scientific performance of their codes. A common, shared center could strengthen collaborative research on science, tools, software, and other development activities.

Scientists from regions lacking access to HPC resources face additional difficulties in adapting to this evolution. Large centers should be encouraged to provide open access to some shared resources as the most effective way for the community to collaborate, foster training and make improvements in all aspects of the prediction system.

WEATHER CLIMATE WATER
TEMPS CLIMAT EAU

WMO OMM

World Meteorological Organization
Organisation météorologique mondiale