Breakout group 1

Ensemble sub-selection: is a common criteria/minimum set of evaluation metrics/framework to serve all communities possible or desired?

Christian Steger, Deutscher Wetterdienst

Anika Ruess, European Space Agency (ESA)

Questions:

- What are current practices?
- How do non-scientific criteria influence the decision/selection?
- How can user needs and application relevant processes be taken into account?
- How to deal with the fact that different regions may have different requirements?
- Is it possible to align the different requirements (across regions and/or activities/applications)?
- Can something already be done for CMIP7?

Key questions/issues/challenges

- Data availability + costs to store the data
- Use of emission scenarios
- General set of criteria for specific regions – common set for different regions and different stakeholders is going to be challenging
- Making sure to not exclude the best models for certain regions

Role for collaboration and who should be involved

- Setting up several guided meetings with all partners to get updated and a streamlined process (which group will provide the data for which scenario, which group are planning to do the extension and when,...)
- ESMValTool and Rapid Evaluation Framework as starting point to generate information for a later point in time to get the criteria – linked to the common set of diagnostics, basis for an initial selection

Suggested actions and next steps

- Follow-up meetings (CMIP, CORDEX, ISIMIP, and others)
- Contact with the responsible from ESMValTool (DLR) and Rapid Evaluation Framework
- CMIP: Modelling centres should provide information, so that the downstream activities can expect any upcoming delays

Breakout group 2

Ensemble sub-selection: is a common criteria/minimum set of evaluation metrics/framework to serve all communities possible or desired?

Key questions/issues/challenges (all details on padlet)

- Taking a step back: Do we want a common set of criteria for model selection?
  - Yes, CORDEX cannot downscale all 50-60 models
  - But challenges remain with 'costliness' of having too many prescriptions
- Choosing 'families' of criteria or creating a framework of guidelines for selection
  - Rather than having specific criteria mandated for all uses
  - Basic QA is sometimes overlooked but should not be
  - [more] importantly – the selection process and rationale must be transparent (why and how the seleciton was made)
  - Plausability metrics that can be applicable / translatable from one region to another
- Global vs regional criteria?
  - Different priorities for different regions
  - Global metrics used in selection can be very different on the regional sclae (eg ECS as seen in cordex-core)
  - Experience from AR6 IPCC Global Sea Level assessment – confusing if different models are used for regional intercomparisons (hardder to compare acriss regions)

Key questions/issues/challenges

- Challenges when with selecting the best criteria
  - even the best GCM selected carefully for a specific can sometimes conflict with some RCMs (physics inconsistency)
- Comparison with observations
  - share how to compute errors / uncertainty on "observations " (e.g. reanalyses, jet position) and from multi-decadal variability

Role for collaboration and who should be involved

- Who should do the selection?
  - Guidelines / framework for the user communities, but not do the complete selection ourselves. Then they do the selection based on the guidelines on what is important for their regions.
  - Framework must be done in collaboration with experts from different regions not to overlook regionally critical criteria.
  - Evaluation will not be as extensive in some of these regions though. Many regions are under researched
- Collaborating with the global community models (FASTtrack)
  - refining the frameworks/tools that have already been developed would really benefit from collaboration with e.g., ISIMIP

Suggested actions and next steps

Breakout group 3

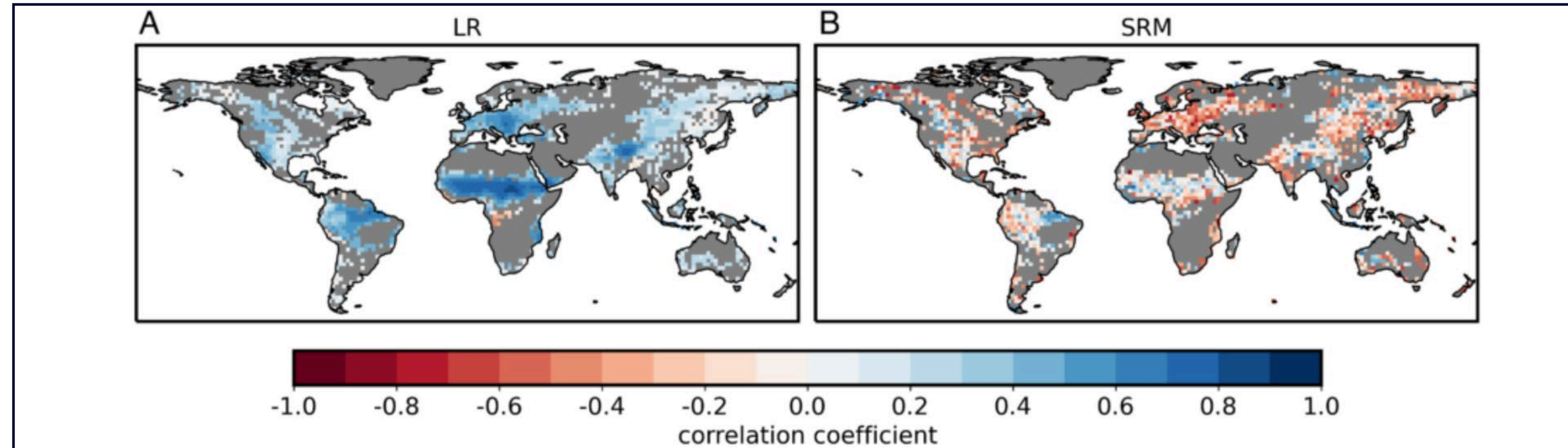What do model development innovations mean for model selection e.g., the CMIP7 focus on CO2 emission driven simulations, increasing resolution and AI/ML?

Couple of inputs to spark the discussion:

(a) Cumulative Emission
- Fossil, E-HIST (Forcing)
- Fossil, C-HIST (14ESMs)
- LUC, C-HIST-NOLU minus C-HIST (10ESMs)

(e) GSAT anomaly
- E-HIST (12ESMs)
- C-HIST (14ESMs)

Hajima et al. 2024
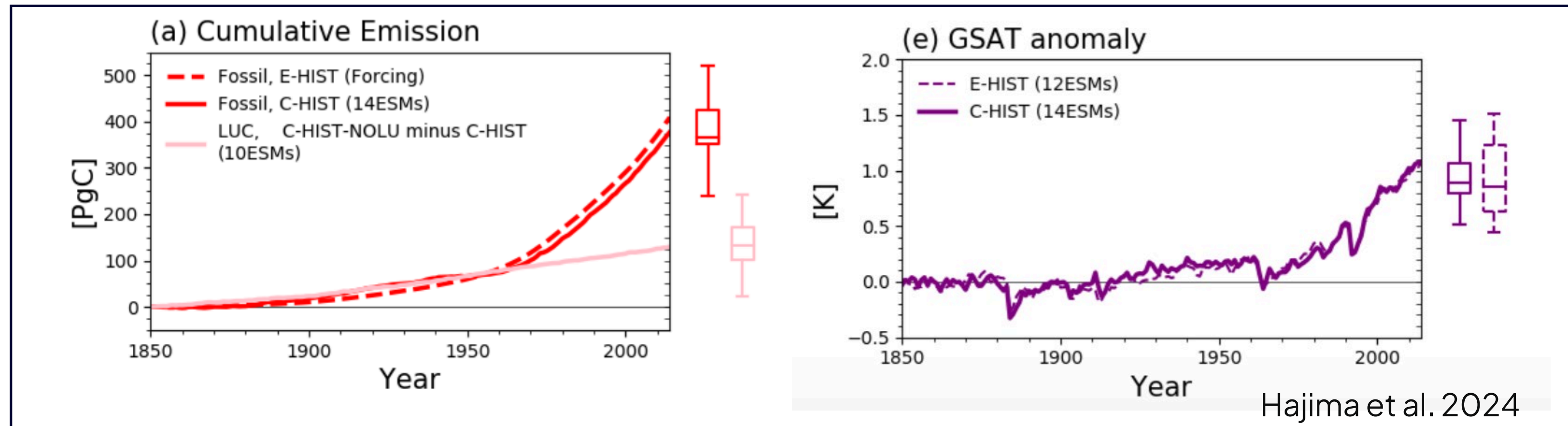
Emission-driven configurations
Could propagate uncertainties from G-ESMs to R-ESMs:
- Initial state/LBCs
- Response to a given scenarios (but more realistics)

A LR   B SRM

correlation coefficient

Fig. 2. Correlation coefficient between SMI and subsequent 9-d mean precipitation for (A) the 60-y mean of the coarse-resolution model (LR) and (B) the SRM. Gray areas as in Fig. 1.

High-resolution configurations
Could also propagate uncertainties from G-ESMs to R-ESMs:
- Change in regional response and pattern
- A fewer ensemble of models and realizations are expected

Lee, J. & Hohenegger, C., 2024

2

Key questions/issues/challenges

- There are developments in both global and regional model developments and this required alignment (e.g., aerosols, land-use, ocean).
- What timescales are downstream communities requiring to ascertain fitness of purpose for GCM/ESMs.
- Problem of only using one ensemble member per GCM/ESMs – needs to be considered seriously
- Is there alignment in the foci of the different communities (CMIP, CORDEX, ISIMIP)?

Role for collaboration and who should be involved

- Need to foster closer collaboration and alignment of global and regional
- Everyone!

Suggested actions and next steps

- Coordination on key questions on selection criteria when considering e-driven simulations and high resolution.
- Common understanding around need for using more than one ensemble member.
- Engaging ML/AI community.
- Potential for Rapid Evaluation Framework to support consistent evaluation across model types.

5

Breakout group 4

Understanding, quantifying and communicating uncertainty

Key questions/issues/challenges

- Be aware of your definition of uncertainty – different stakeholders will have different definitions
- Being aware of model selection in a chain of processes/decisions
- Designing the ensemble to 'do what you want it to do'.
  - There are opportunities to use ideas from statistical design to inform the selection of ensemble members you choose.
- Selecting for different regions requires different considerations
  - Different regional drivers
  - Process representation (e.g. lakes)
  - Different amount of available output
- What will be most useful for applications, impacts, and decisions?
  - What uncertainty do you need to understand/quantity?
  - Storyline approach can be a good technique to cut through the uncertainty problems– no ensemble will be perfect!
- Other issues – be aware of keeping scientific curiosity alive!
  - E.g. hot models – appreciate the nuance

Role for collaboration and who should be involved

Suggested actions and next steps

Breakout group 5

Constraints on the availability of required GCM/ESM model data – timing, provision of required data including temporal frequencies and extensions

Key questions/issues/challenges

- How to get information from the centres to help with planning?
  - Communities need information for planning

- Is it worth it to push for CMIP7 based impact models for AR7?
  - Timing is key – cascade from ESM to impacts modellers is long!

- Would be very useful for users to understand modelling centre intentions – which simulations will they run?

- Is it helpful to prioritise variables which are needed on the fastest timescales?

- How can users get information on how will models be different from CMIP6 to CMIP7?
  - Could the development information be summarised – Essential Minimal Documentation will be a step forwards for this
  - Will help with understanding model families

2

Role for collaboration and who should be involved

- IPO has done a lot of user engagement – gives more people a voice. Could utilise this to get the end user 'voice' to the modelling centres

- Data Request Task Team have provided a model/framework for user engagement with the modelling centres
  - Could expand on this to discuss model selection.

- Data requirements should engage with Data Request Task Team process

Suggested actions and next steps

- CMIP IPO to engage with modelling centres to try and understand which centres will run which simulations
  - Regular updates
  - What they plan to output also useful for downstream planning

- Outline a mechanism for engagement with modelling centres
  - Impacts and Adaptation team (alongside other themes) have made good progress with this. MIPs were represented alongside modelling centre reps.

Key questions/issues/challenges (detail)

- How to get information from the centres to help with planning?
  - E.g. model extensions, helpful for both glacier and ice sheet communities in particular. The more ensembles the better!
  - Providing modelling centres with information on how data will be used will help them to make decisions.

- Is it worth it to push for CMIP7 based impact models for AR7?
  - When is it possible to get the first batch of simulations? Need the data early to cascade from GCMs, to downscalers, to impact modellers. Cut off for AR7 is not too far away!
  - Earliest simulations likely available end of 2025
  - Would be really useful to understand modelling centre intentions – which simulations will they run?

- Is it helpful to prioritise variables which are needed on the fastest timescales?
  - Would also help to manage data on the nodes
  - Modelling centres will also make their own internal priorities. Some centres are still developing their models.

- How will models be different from CMIP6 to CMIP7?
  - What are the key model developments? Different groups document their developments to different levels
  - Could the development information be summarised – Essential Viable Documentation will be a step forwards for this
  - Will help with understanding model families
  - End users want to know the impacts of changes in models – different resource levels in centres will impact ability to do this.

Breakout group 6

Sustained and supported infrastructure to store, deliver and provide user friendly platforms for analysis and frameworks for coordinated cross-community exchange

Key questions/issues/challenges

- Data volume (user struggle to handle the data and access the data) – managing efficiently (which server? Who should maintain that?)
- Different levels of data – raw data, and data which is directly applicable
- Data normalization
- Users are confused, even if they are familiar with netCDF
- A central institutions for hosting it is costly
- Data storage of users
- Big barrier: online computing expenses (specifically developing nations)

Role for collaboration and who should be involved

- Coordination of CMIP for users to access the data more easily
- Enhancing cloud computing for users – CMIP providing proposals

Suggested actions and next steps

Potential solutions: specific user recommendations
- Infrastructure for clear guidance for a specific purpose on what to actually access
- Subselection of 3D-Variables
- Description in the metadata
- A centralised platform: Certain variable/time period – automatic download of a subsection? Just to the server and then download the final data output
- Data standardisation by CMIP
- CMIP providing proposal for funding resources for centralised cloud platform
- Standards on data compression (low-hanging-fruit) for the raw model data and maybe different standards for applicable data
- Not the whole globe is needed when downloading, just a specific region, this could also help the users to reduce data storage (e.g. land-masked with standard)
- Standardisation of interpolation

Breakout group 7

Speeding up the process from creation of global simulations until data/information reaches the end user including potential role of community developed tools to support model evaluation and selection

Motivation

- It takes a very long time for the information from the latest generation of CMIP simulations to reach the end users

CMIP6
downscaling
EURO–
CORDEX

2025

Impact
models

End user
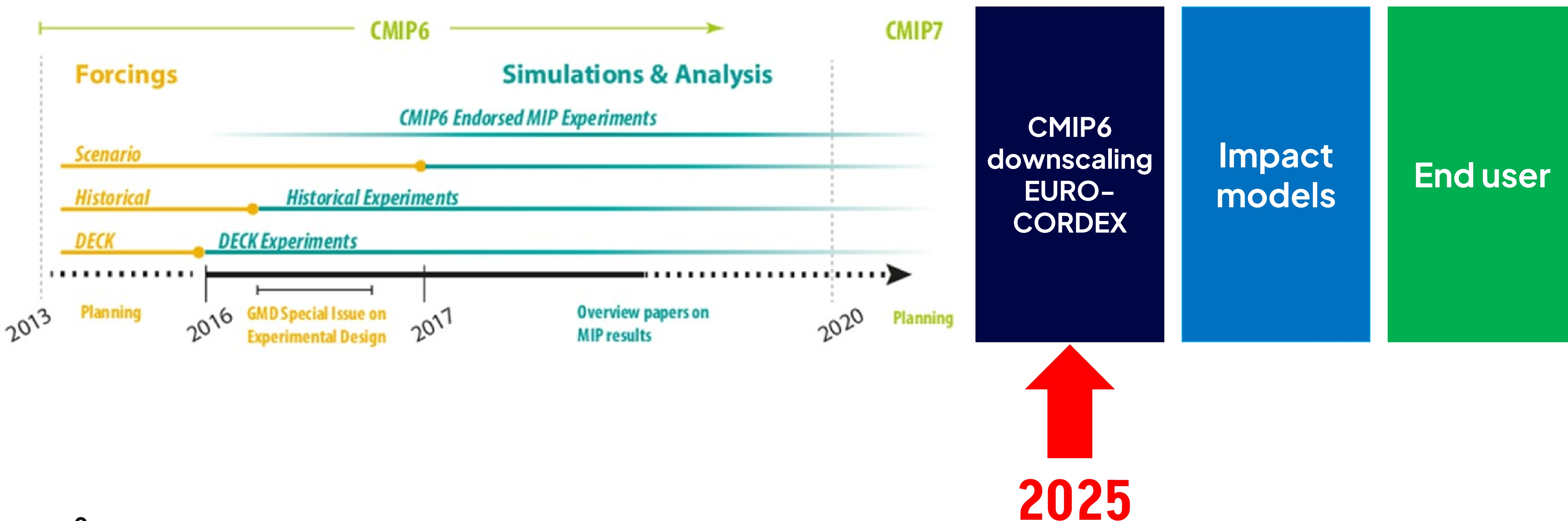
Questions

- What causes delays?
- Are there CORDEX domains or projects that have been able to realise a faster process?
- What would be necessary to ensure a faster process in the future?
- Can something already be done for CMIP7?

Key questions/issues/challenges

- Understanding why there is such a delay between CMIP6 and CORDEX CMIP6 – development of GCM framework selection, limited resources as not operationalised, and CMIP6 downscaling may delay start of CMIP7 downscaling, too close to AR6 deadline.
- Need to recognise the requirements of end users.
- How can tools like ML or other forcings provision be speeded up
- Communities need to work together – CMIP and CORDEX working closely but need to more tightly connect with ISIMIP and activities like DestinE.
- Does ML provide internal consistency for downscaling?

Role for collaboration and who should be involved

- What are the strengths of the different communities and how can this be brought together.
- All the different communities must come together in a structured framework and break down silos.

Suggested actions and next steps

- One quality checker for all communities.

- Collaboration on common indices/criteria (e.g. Rapid Evaluation Framework)

- Common tools and automation to deliver error free output.

- Standards of practice needed for ML and other tools.

Breakout group 8

Balancing competing needs – funding, politics and ensuring equity, and supporting training and capacity building in a global community effort

Key questions/issues/challenges (more details on the padlet)

- Funding context: CMIP is mainly voluntary / in-kind
  - More and more regions are struggling for funding. Regions can then prioritise on their own region
  - If the list of criteria is too long would be prohibitive for uptake from the modelling centres / model users?
  - Should / can WCRP pool funding and resources across the different groups of WCRP but also outside of WCRP to be more effective? also to stop duplication of efforts
  - Some smaller parts are funded (e.g. the IPO and the REF) – can we widen this?

- How to ensure a process that is inclusive for all CORDEX regions?
  - Regional imbalance: WCRP and the modelling centres have a GN imbalance
  - Digital Earth Hackathon could be a good model for building an inclusive process?
  - No ESGF in Africa . Would be a big statement to have this.
  - How to encourage more interaction? more alterative lines of communication, more reward for voluntary / in-kind contributions
  - Cloud hosting for data access. Issue of data continuity (if hosted by Amazon for e.g.)

2

Role for collaboration and who should be involved

Suggested actions and next steps

- WCRP investigate pooling resources
- CMIP look at the status of the data access task team – look at their options outside of the GN.
- Joint actions on ML and AI – investigate their role in accessing simulations (a more cost-effective solution?)

CMIP | WCRP

Thank You

@wcrp-cmip.org

wcrp-cmip

cmip-ipo@esa.int