Bringing it all together: Science and modelling priorities to support international climate policy.

I welcome open discussions and the opportunity to participate in scientific discussion connected to discussion papers such as these. Here I will provide comments and suggestions, some of which are deliberately critical to motivate refelcations and further discussions - such usually enhance our understanding and improve scientific papers.

My first reaction is that the title of this paper is a bit inappropriate, bringing it all together, as the paper exclusively focuses on modelling work. There is much more than models, although models represent an important aspect of science. Hence, there is not just modelling, but also observations, physics/chemistry/biology/geography and analysis - with a connection to the models.

L120. The paper could also ask specifically what kind of information has been used for decision-making and how it has been used. What information should have been used, but hasn't? (Atmospheric CO_2-concentrations keep rising despite IPCC reports and COPs; development within climate change adaptation is still slow) Is there really a need for even larger volumes of data, and does the information distilled thereof really improve? (https://DOI.org/10.1038/NCLIMATE3393)

What is meant by "internally consistent" in this case - and why not just consistent? (as used later on)

There is an unstated dilemma - HPC demands a lot of energy and contributes to our carbon footprint. This affects both ESMs, RCMs and AI/ML. Statistical methods, use of mathematics and clever analysis, on the other hand, can alleviate some of the computational needs. Don't forget that.

L212: "(RCMs) generate high-resolution regional simulations consistent with the ESM boundary condition…" - this is not true because the RCMs provide more detailed representation of topography and processes which result in different precipitation and hence fluxes than the driving GCM/ESM. GCMs/ESMs and RCMs are often inconsistent with each other (e.g. RCMs tend to generate a different cloud and precipitation climate than the GCM/ESM, and hence give different OLR aggregated over the same region, https://doi.org/10.1175/JAMC-D-20-0013.1). One question is whether this makes a big difference (can perhaps explain some biases). Furthermore, bias adjustment is often necessary, which often renders the results internally/physically inconsistent.

“CMIP simulations are extensively used to inform policy addressing global climate change risks …“ - this is perhaps true to some extent, but they have nevertheless not resulted in decisions that meet the Paris accord. This is a big problem. The current global warming does not happen naturally, but is a consequence of human activity. The CO_2-concentrations seem to have accelerated after each COP and IPCC report (e.g. the Keeling curve).

For regional climate change, as well as impacts and climate change adaptation, it is important to stress the results from https://doi.org/doi:10.1038/nclimate1562: that decadal variability is both chaotic (non-deterministic) and pronounced on regional scales.

Use the phrase “*carbon capture and storage*” rather than “*negative emissions*” (throughout the paper) - the latter term is meaningless for most people and our usage of terms has a tendency to spill over to the public discourse. The term is also physically meaningless in a literal sense. Perhaps part of the reason why we struggle reaching our objectives with curbing our emissions is that most people don’t understand what we say? Also, “*key warming targets*” should be “*key warming limits*” because we don’t aim at global warming reaching 2.0/1.5°C, but want to stay below.

Section 3.2: Downscaling - use both RCMs and ESD as they are based on *entirely different assumptions* with different strengths and weaknesses, *independent* of each other. Often downscaling is presented involving only RCMs which results in misapprehension of what downscaling entails. There is also a minimum skillful scale associated with numerical models which is the main reason for why downscaling is needed in the first place (also relevant for **L588-595**). Artificial Intelligence, machine learning and knowledge gaps connected to downscaling should also be considered here. E.g. statistical downscaling can be done in many different ways, and different approaches are suitable for different situations.

L415-430: Perhaps mention that we can use model results to explore explanations for biases and improve our understanding of meteorological phenomena such as the jet stream, cyclones, atmospheric rivers, and their role in the hydrological cycle? Through a set of numerical experiments and experimental design. One interesting topic could be to explore so-called tipping points (e.g. thawing permafrost og THC) and how different model set-up affects the results.

Section 4.2: High-resolution modelling can be used to study phenomena and calibrate statistical models (e.g. testing the fractal dimension of precipitation; Doblet et al. in progress). Climate risks can be quantified by downscaling the shape of pdfs describing a climate variable (e.g. temperature, precipitation, wind speed) in a direct manner, rather than downscaling individual outcomes and then fitting a pdf to the resulting data points. It’s probably more robust to estimate risks through downscaling such statistical properties than through downscaling individual outcomes (statistical properties are often very predictable). What we typically seek is the dependency of the shape of pdfs to large-scale conditions.

L513: “Machine Learning (ML) offers the potential to reduce long-standing systematic errors in ESMs and enhance the overall projection capability of these models.” - such a quick-fix is not reassuring and may hide serious shortcomings connected to the models. There is a lot of hype on AI/ML, but ML/AI has been used in downsclaing since 1999 (e.g. DOI:10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2).

L531: One drawback with ultra-high-resolution global storm-resolving models is that we need large ensembles to sample regional decadal non-deterministic chaotic variations (https://doi.org/doi:10.1038/nclimate1562). We can look at seasonal forecasting which typically involves a large ensemble and not one very high-resolution simulation. There are some unresolved issues regarding seasonal forecasting, such as the optimal set-up of the model themselves (e.g. https://doi.org/10.5194/esd-7-851-2016), which also could be mentioned here as well as decadal forecasting.

L598: The phrase “each sampling internal variability (as represented by that *model*).“ should be “each sampling internal variability (as represented by that *simulation*).“ One specific model can provide many different realisations (https://doi.org/doi:10.1038/nclimate1562). It is important to evaluate the models in terms of their representation of the covariance structure (e.g. https://doi.org/10.5194/gmd-16-2899-2023) and the MMEs in terms of their ability to produce statistical properties that are comparable to the observations (e.g. ERA5).

L660: “Such approaches support assessment of potential regional impacts sampling uncertainty in the future driving regional climate, including changes in climate variability and extreme weather.” - exactly how do they support and what about model biases and the model’s minimum skillful scales? Are there examples of this? References?

L695: The phrase ‘cascade of uncertainty’ is commonly used in the research community, but also a bit misleading. If the uncertainties only cascade, then we should stop at the first step. Of course, the uncertainties don’t only cascade, but we also add constraints/information though each step. Hence, it’s a question of whether we add more information or more uncertainty (https://DOI.org/10.1038/NCLIMATE3393).

L702: The climate sceptics *communicate* uncertainty (doubt), but we *explain* uncertainty. In this case, the main message is nevertheless focussed on certainties, knowledge, as well as their limitations and caveats. If I tell a decision-maker that I have a lot of uncertainty, what should we expect that person to do with that?

L723: Gutiérrez et al. (2019) compared a *large number of downscaling methods*, *not wider ESM projections*. However, there are some papers which describe how ESD has been applied to large MMEs (e.g. https://doi.org/10.1175/2010JCLI3687.1, https://doi.org/10.1016/j.cliser.2017.06.013, https://doi.org/10.1175/JAMC-D-18-0179.1).

L725-rest of the paragraph: Not just ML, but also statistical downscaling may provide similar results, e.g. through ‘hybrid downscaling’ (e.g. https://doi.org/10.1175/JAMC-D-20-0013.1).

L768: It may be useful to reconsider traditional ways of archiving data, e.g. https://doi.org/10.1016/j.cliser.2017.06.013. At least as an addition to traditional netCDF/CF repositories.

L782: It's important to keep in mind the dangers of establishing echo-chambres within a research community. Their existence may be visible through clustering of cited work and omission of relevant references (or examine references herein which don't include the authors of the manuscript). It's hard to keep up-to-date with all the progress and literature, and too easy to walk the same circles over and over again. The paper proposes extensive collaboration across approaches, and there is indeed room for improvement as my comments hopefully will allude to. So, is there a point in thinking differently? (scientists can be quite conservative) For climate change adaptation and decision-making, perhaps put more weight on stress testing or sensitivity testing? Many reports on weather-related calamities involve drought and floods, so why not more discussion on the hydrosphere? (the word 'hydrology' is not even mentioned in the manuscript)

The recent report https://doi.org/10.2777/34601 from the EU may be relevant for this discussion paper.