Response to anonymous Reviewer #2.

I want to thank anonymous Reviewer #2 (DOI: 10.5194/gmd-2021-176-RC2) for the frank and critical comment on my manuscript (Rasmus E. Benestad 2021). Reviewer #2 thinks it "*is filled with rhetoric and subjective statements, and, surprisingly, does not contain any new scientific advancement*", which is quite different to the opinion of Reviewer #1: reading "*like a project report or a report of a laboratory for an external evaluation*". These are of course very subjective opinions and none of these descriptions are of course true. And exactly what is meant by rhetoric and what is wrong with it? (rhetoric is always present, even in the reviewer's comments)

Benestad (2021) presents a downscaling strategy on an overarching level and cites past papers and work that support the choices made for this strategy. Each cited paper can be thought of as small Lego blocks, and the strategy is the final structure containing all these little building blocks. Such an overarching strategy paper, as far as I know, is not so common in academia. It is perhaps a novel way of presenting a comprehensive approach and may be a reason for prompting two such wildly different impressions.

As is explained to Reviewer #1, very little of this work has made any impression in the international downscaling circles, which is a reason for why it's important to write a paper like this. The described strategy for downscaling has evolved over time, and the reason why we use it is exactly because we think it is superior - if it weren't so, we would of course have chosen a different one. The fact that it differs from approaches adopted elsewhere[1] also implies some criticism of those when we discuss differences. This is an expected part of the scientific discourse and part of the scientific debate. We welcome any critical view and arguments on our strategy.

The paper does not try to pretend that the whole scientific community in Norway is tied to the work described herein, and hence the title 'A Norwegian Strategy to Downscaling'. There is no reason to advance such a misconception. Nevertheless, the title is appropriate since it also describes the strategy adopted by the Norwegian Climate Services and presented in the recent Norwegian national climate reports. One of the aspects that distinguishes Norway from many other countries in terms of downscaling is that we combine both dynamical and empirical-statistical downscaling. We also try to foster a common platform for networking and collaborations: https://sites.google.com/met.no/downscaling. But is this an important issue with this paper? I urge Reviewer #2 to please explain exactly what is rather contradictory and worrisome with it and its title, because this is an unexpected comment which is hard to interpret. For me, this concern seems really far-fetched.

Why this paper?

The world is experiencing rapid climate change and urgently needs to start climate change adaptation, which needs to be based on the best information that we can provide on future risks and opportunities. There is already a motion prompted by the Paris 2015 accord, the Climate Adaptation summit, Copernicus C3S, and the upcoming COP26 (2021). My group has

[1] I think, but I haven't seen any similar paper outlining the general thinking on downscaling and the provision of regional climate information for society.

long experience with climate analysis and downscaling, geared towards climate change adaptation, but my impression is that our progress has until now been ignored outside Norway, as explained to Reviewer #1. Hence, it's important to share our experience, and rather than getting trapped in a forest of details, it's better to take a bird's eye overview to convey the rationale behind the downscaling strategy. All the details are of course available from the cited papers (if some of them are not open-access, please let me know).

What is Mainstream?

Good question, and perhaps there could be a better term for it. In this context, it is the norm within the downscaling community, both for ESD and for RCMs - in addition to the attitudes and beliefs shared within e.g. WCRP, CORDEX and COST-Value. The chapter on regional climate modelling in the recent IPCC AR6 WG1 (2021) reflects some of them. It is true that many countries have their own regional climate simulations with spatial and temporal resolutions that far exceed those of CORDEX. Nevertheless, CORDEX also involves more than spatial and temporal resolutions. For instance, Benestad (2021) refers to protocols for evaluation of the methods. In the revised paper, a definition will be given for the term mainstream: "*The term 'mainstream' is used here when referring to the most common norms within the international circles of downscaling, such as the protocols adopted by CORDEX and COST-Value, however, this notion may not necessarily be acknowledged by everyone since there are many differences between individual research groups*".

Common EOFs

Reviewer #2 makes a good point that others may use a different term than 'common EOFs', but doesn't suggest what these terms might be. Nevertheless, a read through the IPCC AR6 WG1 on regional climate modelling reveals that common EOFs, or the same concept under different names, are not present in the assessment of the science on regional climate. My own experience is that people often don't understand the concept, and I therefore challenge Reviewer #2 to find more than 63 papers on downscaling that are based on common EOFs - otherwise the claim that "the reasoning behind l. 80 is flawed" is misleading. This particular remark is also a bit weird (and rhetorical), since sentence in question merely observes a fact: "*In spite of the success with utilising common EOFs, a Google scholar search on '"common EOFs" downscaling' only had 63 hits (of which about 40 referred to our own work), despite more than 20 years since they first were introduced in ESD and the widespread need for climate change adaptation and downscaled results.*" Reviewer #1 offered a different opinion on this finding: that our colleagues don't believe the common EOFs have any merit. Nevertheless, this doesn't disqualify a discussion about their merit and appearance. The point with this remark was that the common EOFs appear not to be widely used and we think they are highly underrated. This point is worth sharing within the science community.

Equations

The paper really is about the downscaling strategy, which is a level above the equations. Hence the comment about equations is a bit irrelevant here. The equations themselves are provided in the cited literature with mostly open-access papers (please let me know if they aren't). The headings of sections 2.1 – 2.6 are '*Main differences to the mainstream*', '*Reasons for why a different approaches*', '*Different choices*', '*"Downscaling climate" approach*', '*Common protocols concerning downscaling*', and '*ESD is suitable for downscaling extremes*' - it's not unnatural that they are mostly qualitative. They do, however, cite works which demonstrate the merit of the different choices behind the chosen strategy. There will also be some degree of subjective choices, as there always is in science. When it comes to the "best way", the paper presents nine criteria for evaluating the results, one of them being the use of common EOFs that also provide a quality control on the large-scale conditions simulated by GCMs. The paper argues that this is an improvement over not including such a quality control. This whole comment is a bit surprising, as it seems to disqualify both machine learning and artificial intelligence.

Details of approach

Moreover, the paper really focuses on the overarching strategy on a higher level, rather than lower-level technicalities. The comment "*precise details of his approach in the paper*" is therefore a bit out of place, and suggests that the novel approach of this paper may be confusing for readers with a conserative set of expectations. The paper should be fairly clear about its objective, and the first sentence of the abstract is: "*A description of a comprehensive geoscientific downscaling model strategy is presented outlining an approach that has evolved over the last 20 years*". I struggle to find a clearer way to say it, sorry! This comment, combined with the observation that there have been 740 views of (Benestad 2021) since July, gives a clear indication for the need of such a paper - because there seems to be intellectual barriers between colleagues in the downscaling community.

I must admit that I don't understand the following sentences: *'It is even unclear whether the reported results appear elsewhere. So, what is the purpose of peer review?*' - how do they fit into this context? When the complaints above are about citing previous papers, then the objections seem to be very thin. Critical comments should be based on real substance, and the reviewer has not even discussed the actual strategy presented in (Benestad 2021) - just weird aspects.

Probability distribution functions

The point about describing and explaining the time dependence of pdf parameters is useful and such an explanation will be provided in the revised paper: "*This approach seeks the dependency of the pdf parameters representing local climate statistics on the large scales, and in practise it involves aggregating the parameters on a seasonal or annual basis. Thus, we end up with a shorter time series of such parameters, which are then used as predictand variables in the downscaling methods against large-scale predictors aggregated over the same timescales.*"

The original paper does offer a discussion on spatial dependence which is ensured by PCA - it already states this: "*We use PCA to represent the predictands because they inherently ensure the same spatial covariance as seen in the observations, in addition to reorganising the data to emphasise the large-scale variability* (Benestad et al. 2015)". We have not yet included more than one variable statistics, but it is possible to let the PCA represent both temperature and rainfall statistics - the maths is the same.

Comparison with RCMs.

For comparison with RCMs, the paper cites (Mezghani et al. 2019) and (Hanssen-Bauer et al. 2009), but there is also some grey literature (Met Norway reports) and some presentations in the Norwegian Climate Services. This covers Norway as well as Poland (Mezghani et al. 2019), but has been limited to specific projects and available funding. It would be great to extend this work to other places too, but we have limited resources and time. It is our hope that by sharing our downscaling strategy, with the help of the open-source `esd` (available from https://github.com/metno/esd), our colleagues will become interested and try to reproduce this type of analysis elsewhere independent of us. This hopefully can happen more easily with this paper.

Extremes

It is true that when qq-plots deviate from the 1-1 line it shows that the Gaussian distribution is unsuitable for extremes. A careful read of section 2.6 '*ESD is suitable for downscaling extremes*' conveys that the Gaussian distribution ***is not used*** for extremes, but ESD can deal with them in a different fashion that nevertheless is well-suited for extremes: in terms of how often e.g. heatwaves occur, assuming Poisson process with low probability, or how long they last, assuming that the duration follow a geometric distribution with a particular success probability p. The original paper explains it this way: "*To estimate the probability of long-lasting events, the geometric distribution needs to be combined with the probability of the events occurring (e.g. Bayesian approach where the probability of the event can be modelled as a Poisson process)*". More details are provided in Benestad et al. (2018), but there are only a few scientific papers on use of statistics in the field of climate research (cited in the cited papers).

Specifics

- I don't understand what is meant by a biased *statement of "state-of-the-art", sorry*!
- *'The goal of the paper is not stated'* - the abstract says it is to present a description of a comprehensive geoscientific downscaling model strategy.
- 'New weather-related hazards' are those that come in the future. Some may even be a new type for a given region. This comment is rhetorical.
- The paper does not aim to present an exhaustive list of types of downscaling approaches. Hopefully, other research groups will present their strategy and compare it with others/the norm/what-ever.

- “Mainstream” will be defined in the revised version.
- In this context, linear algebra involves EOFs and PCA in addition to vectors holding the data and ways of dealing regression on computers. There is a book with the title 'Linear algebra’ (Strang 1988) that provides a good background.
- No problem that Reviewer #2 doubts that CORDEX can be considered as “mainstream” for downscaling. This is a minor issue.
- The passage “*and some will say it gives the right answers for the wrong reasons*” will be dropped in the revised version.
- Modelling temperature by a Gaussian distribution is not advocate for the extremes. See the point on extremes above.
- The question whether the variables are iid - it depends. There is little year-to-year dependence between the parameter estimates, as explained in the original paper. Within a season, there is autocorrelation for temperature, which means that the real degrees of freedom for a season are less than 90. Still, testing them for normality (qq-plot) tends to give reasonable results, and the mean value is also mostly non-problematic. For all intents and purposes, the results are useful for society.
- Modelling the precipitation by an exponential distribution will be more carefully explained in the revised paper: “*It is well-known that 24-hr precipitation doesn't follow the exponential distribution, and especially for the more extreme rainfall amounts. It can nevertheless provide a useful framework for the analysis of 24-hr rainfall (Benestad et al. 2012a; 2012b) and by introducing a 'scaling factor' into this framework, it is possible to to get a more accurate representation (Benestad et al. 2019).*”
- The details about the way the parameters of the statistical laws are “downscaled” are provided in the cited papers, but also explained in this paper during the discussion of the use of regression on aggregated parameters (this prompted som comments from Reviewer #1). Also, more details will be provided in the revised paper.
- Extremes of temperature and precipitation ***are not*** modelled with any assumption of the Gaussian distribution, and for 24-hr rainfall, the exponential distribution is merely used as a starting point (a kind of framework) with a modification to capture the deviations from it. These details are discussed in the cited papers. Benestad (2021) should be fairly clear on these points if the manuscript is read carefully.
- The point regarding ‘*The strategy of “keeping things simple and elegant (mathematical)”’* referred to the merits of using common EOFs and downscaling the parameters of the pdfs directly, but there is still a need for thorough evaluation (nine levels). The remark of *“spherical cows in vacuum”* is, using Reviewer 2’s own words, merely *rhetorical.*
- GCM grids are the grid meshes on which the GCMs represent variables such as temperature and rainfall. Isn’t that obvious?
- Common EOFs deal with differences in mean values by combining respective anomalies from reanalyses and GCM results, and the mean climatology from the station observations is subsequently added to the output of the ESD results to provide a downscaled record. If PCA are used as predictand, then the PCA is estimated on the anomalies of the station data, the PCA is projected through ESD, and results corresponding to the original data are derived after the PCA have been ‘reversed’ to recover the original data structure with the mean values subsequently added. Keeping

all these details in this overview paper would get any reader lost in the forest of details, but they are available from the cited literature. And again, the fact that I need to explain how common EOFs are used may suggest that the choices in our strategy, for instance using common EOFs and PCA, are unfamiliar for most of our colleagues. This is a bit ironic, since this question suggests that the use of common EOFs are not so common and that the results from the Scholar Google search actually do reflect the situation.

- The strategy for storing large multivariate data also involves common EOFs, and it may be a bit tricky to explain to people who are not familiar with them. It's explained in the cited paper (Benestad et al. 2017). The point with this section is also explained to Reviewer #1: It deals with concerns of making regional climate information accessible for the society.
- It is true that Figure 6 does not demonstrate anything on the storage method - but it only took seconds to produce it from a large multi-model ensemble. Try the cited URL instead.
- *What makes the author believe of "silo thinking"?* Separate CORDEX white papers for RCMs and ESD, for instance.

References

Benestad, Rasmus E. 2021. "A Norwegian Approach to Downscaling." Preprint. Climate and Earth system modeling. https://doi.org/10.5194/gmd-2021-176.

Benestad, Rasmus E., Deliang Chen, Abdelkader Mezghani, Lijun Fan, and Kajsa Parding. 2015. "On Using Principal Components to Represent Stations in Empirical-Statistical Downscaling." *Tellus A* 67 (0). https://doi.org/10.3402/tellusa.v67.28326.

Benestad, Rasmus E., Bob van Oort, Flavio Justino, Frode Stordal, Kajsa M. Parding, Abdelkader Mezghani, Helene B. Erlandsen, Jana Sillmann, and Milton E. Pereira-Flores. 2018. "Downscaling Probability of Long Heatwaves Based on Seasonal Mean Daily Maximum Temperatures." *Advances in Statistical Climatology, Meteorology and Oceanography* 4 (1/2): 37–52. https://doi.org/10.5194/ascmo-4-37-2018.

Benestad, Rasmus E, Kajsa M Parding, Helene B Erlandsen, and Abdelkader Mezghani. 2019. "A Simple Equation to Study Changes in Rainfall Statistics." *Environmental Research Letters* 14 (8): 084017. https://doi.org/10.1088/1748-9326/ab2bb2.

Benestad, Rasmus, Kajsa Parding, Andreas Dobler, and Abdelkader Mezghani. 2017. "A Strategy to Effectively Make Use of Large Volumes of Climate Data for Climate Change Adaptation." *Climate Services*, June. https://doi.org/10.1016/j.cliser.2017.06.013.

Benestad, R.E., D. Nychka, and L. O. Mearns. 2012a. "Spatially and Temporally Consistent Prediction of Heavy Precipitation from Mean Values." *Nature Climate Change* 2 (doi: 10.1038/NCLIMATE1497): 544–47.

———. 2012b. "Specification of Wet-Day Daily Rainfall Quantiles from the Mean Value." *Tellus A* 64 (14981). https://doi.org/10.3402/tellusa.v64i0.14981.

Hanssen-Bauer, I., H. Drange, E. J. Førland, L. A. Roald, K. Y. Børsheim, H. Hisdal, D. Lawrence, et al. 2009. "Klima i Norge 2100." september 2009. Oslo: Norsk klimasenter.

Mezghani, Abdelkader, Andreas Dobler, Rasmus Benestad, Jan Erik Haugen, Kajsa M. Parding, Mikołaj Piniewski, and Zbigniew W. Kundzewicz. 2019. "Sub-Sampling Impact on the Climate Change Signal over Poland Based on Simulations from Statistical and Dynamical Downscaling." *Journal of Applied Meteorology and Climatology*, March. https://doi.org/10.1175/JAMC-D-18-0179.1.

Strang, G. 1988. *Linear Algebra and Its Application*. San Diego, California, USA: Harcourt Brace

& Company.