Machine learning emulation of a local-scale UK climate model

IAI CDT Summer Project Report

Henry Addison
Supervised by Peter Watson, Laurence Aitchison, Dann Mitchell

Interactive AI CDT
University of Bristol
UK
September 2021

4 Related work on user requirements for climate data

Working with climate data is not easy. There can be mismatches between the intended and actual uses of projection datasets [9]. There are many sources of uncertainty from the natural variability of climate to issues with a model, both in how it models processes and the setting of parameters and initial conditions. Even expert users who understand these issues do not agree on the best way on to represent and interpret outputs from climate model ensembles [41].

For example, how experts choose to communicate the range of rainfall projections from an ensemble of climate models to less technical policymakers can vary [41]. On the one hand there may be experts who choose to show the full range since "it is the extremes that matter to adaptation policymakers". On the other there are who avoid sharing the most extreme outputs in case it leads to "panic rather than action". Another participant in that study went even further on the matter and refused to give an answer as to what range they would communicate to policymakers without more information about the ensemble than just the visualizations of the full ensemble's rainfall projections.

User requirements gathering work for climate services by Kaspar *et al.* [42] use standards and outputs created by international bodies and projects such as World Meteorological Organization (WMO) and CF Conventions[1] as a source of user requirements. While these may provide a useful starting point these standards will not have considered new approaches to producing climate data and the need for them to be summarized may miss the needs of subsets of users.

User requirements can also be captured through engaging with potential users with surveys and interviews [43]. Skelton *et al.* [43] categorize the users of climate services based on their actual usage of Swiss national climate scenarios. They extend a metaphor of an iceberg created by Braunreiter and Blumer [44] in the context of energy scenarios to three distinct groups of users: divers, sailors and observers. Sailors are interested in the key findings found in the summaries of climate studies, the tip of the iceberg above the water. Divers are also able to work with the raw data, the bulk of the iceberg below the surface. Observers are aware of the iceberg from a distance (they may have skimmed the same summaries) but have not changed their behaviour based on it. Through surveys and interviews a more nuanced view of users is built, rather than working with easy to access characteristics such as whether users are academics or practitioners or highly numerate: an observer may well be a highly numerate academic or practitioner who has so far chosen not to use the raw data.

When it comes to making decisions using climate data, the need may not be for more precise projections coming from bigger and better models but instead the need may be for ways to explore more of the range of plausible futures [45]. More computationally intense, higher resolution projections may not lead to better decision-making. There are groups of users whose needs are not yet met by climate models and downscaling [9]. Weaver *et al.* [45] suggest that it is better to focus less on "overcoming the extremely difficult scientific and technical obstacles to climate prediction" and focus more on a two-way, participatory approach. This balance in focus may be hard to find. Model resolution, both spatial and temporal, can be important in exploring climate scenarios [10]. A faster way to generate future scenarios could allow for the collaborative creation of projection datasets that are tailored to a particular decision context and set of stakeholders to aid their decision-making.

5 User requirements study

To guide the design of the emulator and focus of the research, user requirements for the emulator and the projections it would produce were gathered from a group of potential users.

5.1 Method

Four potential users were recruited using my and my supervisors' professional network with a further two through snowball sampling for a total of six participants. For this report each individual

[1] http://cfconventions.org/

is given an anonymous label, P1 to P7. These participants had roles at government research agencies, universities and private companies and charities interested in flood-risk modelling and management (see Table 1).

The interviews were semi-structured. Information about the user's experience (roles, organizations) working with climate data were gathered to help characterize the participants. Then the topics covered the applications in which participants used climate data and their frustrations working with it, their desires for the outputs of the model (e.g. variables, temporal and spatial resolution, time periods and spatial subsets of interest, file formats), characteristics that it is important the emulator get right (e.g. rainfall profiles) and definitions of extreme events.

A thematic analysis [46] was performed on each interview transcript, with the aim of finding connecting themes, issues and requirements between the different interviews. I coded each interview then examined the codes to inductively identify common themes from which the final set of user requirements were developed.

Approval for this study into user requirements was given by University of Bristol's Engineering Faculty Research Ethics Committee.

In the following sections I discuss the characteristics of the participants, explore the three main themes in user requirements that emerged and consider some recommendations for the output of the emulator.

5.2 Participant characteristics

Participant	Role type	Primary organization type
P1	researcher	flood-risk modelling company
P2	science officer	government research agency
P3	hydrology researcher	university
P5	hydrology researcher	university
P6	flood forecaster	environmental risk consultancy
P7	hydrologist	environmental risk research and education charity

Table 1: Overview of participants and their roles working with rainfall projections.

Details of the primary roles of each participant in which they worked with rainfall projections are in Table 1. For five of the six participants (also the five more concerned with consuming rather than producing projections), their interests were in hydrology and flood and drought modelling. It would be interesting in further work to explore if there are potential users of high-resolution precipitation projections in other domains.

The group shows both a diversity in the organizations the participants work for and also demonstrates a difficulty in categorizing a user's organization since there was an overlap between industry, public bodies and academia. Employees at private companies are involved in research, university academics act as consultants and members of government bodies have part-time positions at universities.

To use the Skelton *et al.* [43] iceberg metaphor, the participants were all divers. They did not just use the UKCP summaries and regularly analysed the raw netCDF projection data themselves for their own purposes. Often this was done to support sailors and chains of other users. For example, one participant used the raw data to model floods and droughts for their own research and also worked with insurers interested in complex spatial correlations of storms and flooding. This matches the current state of the project since the initial output of the system will be raw projections but there is more that could be learnt from other user types (sailors and observers) about how these raw projections can be made more useful to them without having to do the work of divers ("????" - P?).

5.3 User requirements

Three main themes emerged from the interviews:

- More data: users want more samples covering more experiments with more ensemble members and using a greater variety of GCMs for the conditioning input.
- Trustworthy data: users need to be able to trust data and need guidance to use it.
- Evaluation depends on the usage: evaluation metrics and definitions of extreme events varies according to application.

5.3.1 More data

Users want more data and there are different ways in which they would like the current local scale dataset expanded. In some cases more samples are desired in order to reduce their uncertainty about extreme events. For P3 "a large ensemble would always be great". According to P2 it is only "2 years out to 100 years return periods we're able to look at with the sort of length simulations that we have". Many wished to study more extreme events (say 1 in 10,000-year return periods), considered shorter time horizons than to 2080 and even questioned whether the 720 years of CPM data (12 lots of three 20 year chunks) allowed for good characterization of events as unlikely as 1 in 100 years. P5 would like to "create synthetic years of climate data or synthetic events and generate many 1000s or even 10,000s of years of synthetic climate" and even 10,000 years would only "allow us to get 10 events with 1 in 1,000 year probability within that. If we're interested in 1,000 year events maybe we need 100,000 years of synthetic climate data."

In other cases it is different experiments that are interesting. An emulator that could generalize to work with different GCMs as inputs would allow for sampling of model uncertainty. P2 describes this as "really critical from an uncertainty perspective" and their "primary interest" for such an emulator. This would allow sampling model uncertainty since the emulator would not be tied to a single GCM implementation. Some participants also felt that the experiments run using the climate models did not match their needs: "climate models are not run with my kind of modelling in mind ... climate scientists are interested in hitting it with a hammer really hard to see what the impact is going to be" - P1. In particular, five of the participants disliked the choice of emission scenario used for the CPM projections. The scenario available is the highest warming one, RCP8.5 (see Section 2.1.1 for more on RCPs), but "it's quite a worst-case scenario" according to P3 and not useful for insurers according to P1 ("do I as a risk modeller selling information to insurers care about RCP8.5? I'll tell you that I don't").

5.3.2 Trustworthy data

Users need to be able to trust the data. Therefore it is important that the rainfall projections generated by the emulator reproduce important characteristics of rainfall, such as intensity-duration frequency mentioned by P2 and P5, "the relative occurrence of high intensity, short duration events versus longer duration events ... a very useful way of indicating the realism of rainfall events within a model" (P2) and seasonal rainfall profiles which should have "more winter frontal rainfall coming in compared to the summer [which should have] more convective storms coming in" (P3).

Bias correction was identified as a problem by three of the participants (P1, P2, and P3). Bias correction attempts to remove systematic bias (such as excessive or too little rainfall) from projections (see Section 2.1 for more). It is something that is hard to do ("GCMs and RCMs require bias correction and that's just horribly messy" — P1) and, to use the words of P2, it is uncertain whether it is "the right thing to do full stop." P3 advised leaving bias correction as something for the user to choose to do. The emulator is not expected to remove bias since it will be present in the training data but the bias of the emulator should be evaluated to see if the problem is made worse.

A more complex model is not necessarily the right answer to provide this trust:

> Are we interested in one single insanely precise answer that's definitely wrong or are we going to relax that decision a little bit to understand parameter and model structure uncertainty? Always the latter I'd say. — P1

> Do I want a very precise answer that's almost likely to be wrong or will I settle for something that's more approximate but I understand the uncertainty better? There's a trade-off there. There's no right or wrong answer. — P5

Trust also rests on having the appropriate guidance and support for users of climate projection datasets. Participants talked about the importance of working with the producers of the data to understand its limitations and avoid misuse. However, not all users are able to do so. They may not have the resources or they may not even have the knowledge to know what questions to ask or which dataset is best to use: "How the different products relate to each other, or the different climate projections all relate to each other is not an easy question" (P2).

5.3.3 Evaluation metrics

There is no one set of metrics to use for evaluating the emulator or even to define an extreme event. Different use cases require different properties of the rainfall projections to be correct and depend on different extreme events.

A drought might be the result of having "10% less than the mean annual rainfall for winter ... 3 years in a row" or "4 months of very little rain, 30 degree plus [temperatures] so lots of the water gets evaporated from the reservoirs". Similarly extreme floods are not linked to extreme rainfall in a straight-forward manner (e.g. P5 shared a work by Sharma *et al.* [47] on the matter).

Hazard maps summarize flood-risk across a large region (e.g. the inundation of the same probability flood across an entire nation or the probability flooding in a year across a continent). They are useful for risk zoning and for P2 "the duration [of rainfall] at a point is more relevant [than the spatial structure of a rainfall system] to stakeholders on the ground who are interested in flooding". However, an insurer may want to know about long-range correlations:

> to do a full blown risk analysis you can't just simply say I'm interested in the 1 in 100 year 24 hour rainfall everywhere because storms don't look like that. They only hit a small part of the country and the return period of the rainfall varies in space. So we want to be able to characterize that and build that into our loss calculations - P5.

If the model can reproduce the characteristics of rainfall as discussed in the previous point about trusting the data then it should be able to support a range of use cases. The important thing to communicate to the user will be at what extremes the emulator can confidently predict. Participants commonly talked about return periods of the orders of 1 in 10 years, 1 in 100 years and 1 in 10,000 years. The latter return period is linked to maximal probable precipitation which is of interest for critical infrastructure such as nuclear power plants and reservoirs.

5.4 Output recommendations

Participants gave no reason not to stick to the same conventions as the UKCP18 Local dataset for the files of projections: use netCDF files with the same horizontal grid[2]. There were some complaints about the grids used by the UKCP18 datasets ("the grids are different between 12km and the 2.2" — P3). The 2.2km data in particular was "awkward" according to P6 and used "weird spherical rotated lat-long" (P3). However, these problems had been overcome to work with this data so it is best to stick with this grid ("I would recommend not changing it to another lat-long system or something different because that would be annoying" — P3).

For the spatial resolution, the extra resolution provided by the 2.2km dataset is useful and provides benefits over lower resolution datasets. In coarser models, rainfall looks very different. P2 noted "you get lots of light rain, not enough heavy rain, the diurnal cycle is wrong, there's a whole host of deficiencies". For P1, "60km doesn't really work, 25 is still pushing it but some of the recent 2k stuff is perfect." P3 explained that many catchment areas in the UK are too small to be well represented in coarser projections:

[2] CF Conventions provide guidance on creating netCDF files suitable for sharing — http://cfconventions.org/

> 25 by 25km is 625km^2 [the resolution of older RCMs]. Now I would say about $\frac{3}{4}$ of the gauged catchments in the UK are less than 625km^2. Most of the catchments, most of the flood generating processes that go on are in small head water catchments of the order of 100km^2. Now if you just have a single massive grid cell with a rainfall mean, it's not, particularly in high elevation areas where the rainfall varies within metres, 10s of metres, not kilometres, a 625km^2 rainfall grid is not going to cut it.

2.2km resolution may not even be small enough for some users. P5 reckoned "you would probably see an improvement if you went to resolution higher than 30m". Unfortunately such scales are not feasible at the moment, not least because the observation record won't support it ("national gridded rainfall is 1km" — P3) let alone the computational cost of such a model.

Participants did note a few downsides. The size of the dataset does present a data management problem. The increased size is also a problem when working with the raw data — "they're so massive which can cause problems when you're trying to hold them all in memory" (P2). Thus there are dangers with producing another high-resolution dataset as users may choose to use it for purposes which do not merit the extra computational and storage costs. However, because it has clear benefits to some users it is better instead to rely on guidance to steer users who do not need the extra resolution to a more suitable dataset.

For the temporal resolution projections at sub-daily (e.g. hourly) rates was mentioned by five participants so there is value in providing them as well as daily rates. It is particularly important for flooding. According to P3 "many of our catchments in the UK are very small and they have a travel time of less than a day" (that is the time is takes for rainfall to hit the surface and travel to a point of interest such as a particular part of a river) so as P2 says "when you look at flooding you're not just interested in the daily rainfall". Despite UKCP18 Local dataset including hourly projections, not all the users were able to make use of it ("we generally only work with daily because that's all we have access to. We know that sub-daily is important." — P1). Some participants even identified sub-hourly use cases such as water companies using 2-minute rain gauges, according to P6. UKCP18 Local dataset is available daily and hourly while the GCM data is only available daily so I will initially target daily predictions and follow later with hourly.

Even in domains where rainfall is important, other outputs are useful too. For the purposes of flood and drought modelling participants would like temperature, longwave radiation, shortwave radiation, humidity, wind speed in order to use evaporation as well as rainfall. For working with storms, wind outputs are also useful.

To keep down computational costs for this investigation, it may be necessary to use only a geographical subset of the data. The main advice from participants was to "cover quite contrasting regions of the UK" (P2) where different processes determine rainfall and it has different characteristics — from convective summer rainfall of the South and East of England to the peak winter rainfall further North and West and uplands and intricate coast of North-West Scotland.

Participant P2 was even able to provide advice on which inputs from a GCM should be used. In particular, while many coarse projections include a precipitation variable, it may be wise to not use it as it may be a parameterized value not capturing all the rainfall like a CPM. Instead it may be preferable to rely on the underlying physical processes and "get it to work with the actual physical predictors [of rainfall]" (P2).

5.5 Conclusion

The requirements from interviews with participants show a desire for more data like the 2.2km hourly projections of rainfall (and more) in the UKCP18 Local dataset. However, this desire for more data needs to be balanced with ensuring the data is trustworthy and comes with sufficient guidance to help users understand the limitations of the data. The description of a minimum viable product (MVP) is a way to summarize these requirements. An MVP emulator from these requirements would:

- allow for fast generation of samples

- generate samples of precipitation rates at 2.2km and daily resolutions on the same grid as the UKCP 2.2km projections.
- output data in netCDF format following CF Conventions.
- be backed up with an evaluation of its ability to re-create well common rainfall patterns and extreme events with a return period of at least 10 years.

Further desirable extensions to this include:

- a flexible interface to allow different inputs from different GCMs running with different GHG emission assumptions.
- generating samples at hourly resolution.
- guidance on how reliably it can represent varying levels of extreme events and long range correlations of rainfall events across the UK.

The next section describes the in-progress method for creating an emulator of the Met Office's CPM capable of meeting these needs.

References

[1] M. G. Donat, A. L. Lowry, L. V. Alexander, P. A. O'Gorman and N. Maher, "More extreme precipitation in the world's dry and wet regions," *Nature Climate Change*, vol. 6, pp. 508–513, 2016. DOI: 10.1038/nclimate2941.

[2] E. J. Kendon, G. Fosser, J. Murphy, S. Chan, R. Clark, G. Harris, A. Lock, J. Lowe, G. Martin, J. Pirret *et al.*, "Ukcp convection-permitting model projections: Science report," 2019. [Online]. Available: https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP-Convection-permitting-model-projections-report.pdf.

[3] J. M. Gutiérrez, D. Maraun, M. Widmann, R. Huth, E. Hertig, R. Benestad, O. Roessler, J. Wibig, R. Wilcke, S. Kotlarski, D. San Martín, S. Herrera, J. Bedia, A. Casanueva, R. Manzanas, M. Iturbide, M. Vrac, M. Dubrovsky, J. Ribalaygua, J. Pórtoles, O. Räty, J. Räisänen, B. Hingray, D. Raynaud, M. J. Casado, P. Ramos, T. Zerenner, M. Turco, T. Bosshard, P. Štěpánek, J. Bartholy, R. Pongracz, D. E. Keller, A. M. Fischer, R. M. Cardoso, P. M. M. Soares, B. Czernecki and C. Pagé, "An intercomparison of a large ensemble of statistical downscaling methods over europe: Results from the value perfect predictor cross-validation experiment," *International Journal of Climatology*, vol. 39, no. 9, pp. 3750–3785, 2019. DOI: 10.1002/joc.5462.

[4] Met Office Hadley Centre, *UKCP18 Local Projections at 2.2km Resolution for 1980-2080*, Centre for Environmental Data Analysis, 2019. [Online]. Available: https://catalogue.ceda.ac.uk/uuid/d5822183143c4011a2bb304ee7c0baf7 (visited on 09/06/2021).

[5] T. Vandal, E. Kodra and A. R. Ganguly, "Intercomparison of machine learning methods for statistical downscaling: The case of daily and extreme precipitation," *Theoretical and Applied Climatology*, vol. 137, no. 1-2, pp. 557–570, 2018. DOI: 10.1007/s00704-018-2613-3.

[6] P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi and G. Yacalis, "Could machine learning break the convection parameterization deadlock?" *Geophysical Research Letters*, vol. 45, no. 11, pp. 5742–5751, 2018, ISSN: 0094-8276 1944-8007. DOI: 10.1029/2018gl078202.

[7] S. Rasp, M. S. Pritchard and P. Gentine, "Deep learning to represent subgrid processes in climate models," *Proc Natl Acad Sci U S A*, vol. 115, no. 39, pp. 9684–9689, 2018, ISSN: 1091-6490 (Electronic) 0027-8424 (Linking). DOI: 10.1073/pnas.1810286115.

[8] D. Maraun, M. Widmann and J. M. Gutierrez, "Statistical downscaling skill under present climate conditions: A synthesis of the value perfect predictor experiment," *International Journal of Climatology*, vol. 39, no. 9, pp. 3692–3703, 2019. DOI: 10.1002/joc.5877.

[9] O. Rössler, A. M. Fischer, H. Huebener, D. Maraun, R. E. Benestad, P. Christodoulides, P. M. M. Soares, R. M. Cardoso, C. Pagé, H. Kanamaru, F. Kreienkamp and D. Vlachogiannis, "Challenges to link climate change data provision and user needs: Perspective from the cost-action value," *International Journal of Climatology*, vol. 39, no. 9, pp. 3704–3716, 2019. DOI: 10.1002/joc.5060.

[10] N. Schaller, J. Sillmann, M. Muller, R. Haarsma, W. Hazeleger, T. J. Hegdahl, T. Kelder, G. van den Oord, A. Weerts and K. Whan, "The role of spatial and temporal model resolution in a flood event storyline approach in western norway," *Weather and Climate Extremes*, vol. 29, p. 100 259, 2020, ISSN: 2212-0947. DOI: ARTN10025910.1016/j.wace.2020.100259.

[11] F. J. Tapiador, A. Navarro, R. Moreno, J. L. Sánchez and E. García-Ortega, "Regional climate models: 30 years of dynamical downscaling," *Atmospheric Research*, vol. 235, p. 104 785, 2020, ISSN: 01698095. DOI: 10.1016/j.atmosres.2019.104785.

[12] J. H. Christensen, F. Boberg, O. B. Christensen and P. Lucas-Picher, "On the need for bias correction of regional climate change projections of temperature and precipitation," *Geophysical Research Letters*, vol. 35, no. 20, 2008, ISSN: 0094-8276. DOI: ArtnL2070910.1029/2008gl035694.

[13] Met Office Hadley Centre, "UKCP18 Guidance: Representative Concentration Pathways," Tech. Rep., 2018. [Online]. Available: https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance---representative-concentration-pathways.pdf.

[14] Z. Hausfather and G. P. Peters, "Emissions - the 'business as usual' story is misleading," *Nature*, vol. 577, no. 7792, pp. 618–620, 2020, ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/d41586-020-00177-3.

[15] H. J. Fowler, S. Blenkinsop and C. Tebaldi, "Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling," *International Journal of Climatology*, vol. 27, no. 12, pp. 1547–1578, 2007. DOI: 10.1002/joc.1556.

[16] S. Ravuri, K. Lenc, M. Willson, D. Kangin, R. Lam, P. Mirowski, M. Fitzsimons, M. Athanassiadou, S. Kashem, S. Madge, R. Prudden, A. Mandhane, A. Clark, A. Brock, K. Simonyan, R. Hadsell, N. Robinson, E. Clancy, A. Arribas and S. Mohamed, "Skillful precipitation nowcasting using deep generative models of radar," 2021. DOI: arxiv:2104.00954.

[17] D. B. Walton, F. P. Sun, A. Hall and S. Capps, "A hybrid dynamical-statistical downscaling technique. part i: Development and validation of the technique," *Journal of Climate*, vol. 28, no. 12, pp. 4597–4617, 2015. DOI: 10.1175/Jcli-D-14-00196.1.

[18] Met Office Hadley Centre, *UKCP18 Global Projections at 60km Resolution for 1900-2100*, Centre for Environmental Data Analysis, 2018. [Online]. Available: https://catalogue.ceda.ac.uk/uuid/97bc0c622a24489aa105f5b8a8efa3f0 (visited on 09/06/2021).

[19] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention, Pt Iii*, vol. 9351, pp. 234–241, 2015, ISSN: 0302-9743. DOI: 10.1007/978-3-319-24574-4_28.

[20] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen and J. Hickey, "Machine learning for precipitation nowcasting from radar images," 2019. DOI: arxiv:1912.12132.

[21] G. Ayzel, T. Scheffer and M. Heistermann, "Rainnet v1.0: A convolutional neural network for radar-based precipitation nowcasting," *Geoscientific Model Development*, vol. 13, no. 6, pp. 2631–2644, 2020, ISSN: 1991-9603. DOI: 10.5194/gmd-13-2631-2020.

[22] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2014.

[23] T. Karras, T. Aila, S. Laine and J. Lehtinen, *Progressive Growing of GANs for Improved Quality, Stability, and Variation*, 2018. arXiv: 1710.10196 [cs.NE].

[24] A. Van Den Oord, N. Kalchbrenner and K. Kavukcuoglu, "Pixel recurrent neural networks," in *33rd International Conference on Machine Learning, ICML 2016*, vol. 4, 2016, pp. 2611–2620.

[25] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Foundations and Trends in Machine Learning*, vol. 12, no. 4, pp. 4–89, 2019, ISSN: 1935-8237. DOI: 10.1561/2200000056.

[26] K. Sohn, H. Lee and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, Eds., vol. 28, Curran Associates, Inc., 2015.

[27] D. J. Rezende, S. Mohamed and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proceedings of the 31st International Conference on Machine Learning*, P. X. Eric and J. Tony, Eds., vol. 32, PMLR, 2014, pp. 1278–1286.

[28] K. Gregor, F. Besse, D. Jimenez Rezende, I. Danihelka and D. Wierstra, "Towards conceptual compression," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon and R. Garnett, Eds., vol. 29, Curran Associates, Inc., 2016.

[29] Z.-S. Liu, W.-C. Siu and Y.-L. Chan, “Photo-realistic image super-resolution via variational autoencoders,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1351–1365, 2021, ISSN: 1051-8215 1558-2205. DOI: 10.1109/tcsvt.2020.3003832.

[30] T. D. Kulkarni, W. F. Whitney, P. Kohli and J. Tenenbaum, “Deep convolutional inverse graphics network,” in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, Eds., vol. 28, Curran Associates, Inc., 2015.

[31] T. White, “Sampling generative networks,” *arXiv preprint*, 2016. DOI: arXiv:1609.04468.

[32] A. B. L. Larsen, S. K. Sønderby, H. Larochelle and O. Winther, “Autoencoding beyond pixels using a learned similarity metric,” in *Proceedings of The 33rd International Conference on Machine Learning*, B. Maria Florina and Q. W. Kilian, Eds., vol. 48, PMLR, 2016, pp. 1558–1566.

[33] I. Gulrajani, K. Kumar, F. Ahmed, A. A. Taïga, F. Visin, D. Vázquez and A. C. Courville, “Pixelvae: A latent variable model for natural images,” in *International Conference on Learning Representation*, 2017.

[34] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, “Generative adversarial networks an overview,” *Ieee Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018, ISSN: 1053-5888. DOI: 10.1109/Msp.2017.2765202.

[35] J. Leinonen, D. Nerini and A. Berne, “Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network,” *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2020.

[36] A. Grover, M. Dhar and S. Ermon, “Flow-gan: Combining maximum likelihood and adversarial learning in generative models,” *Thirty-Second Aaai Conference on Artificial Intelligence / Thirtieth Innovative Applications of Artificial Intelligence Conference / Eighth Aaai Symposium on Educational Advances in Artificial Intelligence*, vol. 32, no. 1, pp. 3069–3076, 2018.

[37] I. Kobyzev, S. Prince and M. Brubaker, “Normalizing flows: An introduction and review of current methods,” *IEEE Trans Pattern Anal Mach Intell*, vol. PP, 2020, ISSN: 1939-3539 (Electronic) 0098-5589 (Linking). DOI: 10.1109/TPAMI.2020.2992934.

[38] J. Ho, X. Chen, A. Srinivas, Y. Duan and P. Abbeel, “Flow++: Improving flow-based generative models with variational dequantization and architecture design,” in *Proceedings of the 36th International Conference on Machine Learning*, C. Kamalika and S. Ruslan, Eds., vol. 97, PMLR, 2019, pp. 2722–2730.

[39] A. van den Oord, N. Kalchbrenner, L. Espeholt, k. kavukcuoglu koray, O. Vinyals and A. Graves, “Conditional image generation with pixelcnn decoders,” in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon and R. Garnett, Eds., vol. 29, Curran Associates, Inc.

[40] S. Reed, A. Oord, N. Kalchbrenner, S. G. Colmenarejo, Z. Wang, Y. Chen, D. Belov and N. Freitas, “Parallel multiscale autoregressive density estimation,” in *Proceedings of the 34th International Conference on Machine Learning*, P. Doina and T. Yee Whye, Eds., vol. 70, PMLR, 2017, pp. 2912–2921.

[41] J. Daron, S. Lorenz, A. Taylor and S. Dessai, “Communicating future climate projections of precipitation change,” *Climatic Change*, vol. 166, no. 1-2, p. 23, 2021, ISSN: 0165-0009 1573-1480. DOI: 10.1007/s10584-021-03118-9.

[42] F. Kaspar, F. Kratzenstein and A. K. Kaiser-Weiss, “Interactive open access to climate observations from germany,” *Advances in Science and Research*, vol. 16, pp. 75–83, 2019, ISSN: 1992-0636. DOI: 10.5194/asr-16-75-2019.

[43] M. Skelton, A. M. Fischer, M. A. Liniger and D. N. Bresch, “Who is ‘the user’ of climate services? unpacking the use of national climate scenarios in switzerland beyond sectors, numeracy and the research–practice binary,” *Climate Services*, vol. 15, p. 100 113, 2019, ISSN: 2405-8807. DOI: 10.1016/j.cliser.2019.100113.

[44] L. Braunreiter and Y. B. Blumer, "Of sailors and divers: How researchers use energy scenarios," *Energy Research and Social Science*, vol. 40, pp. 118–126, 2018, ISSN: 2214-6296. DOI: 10.1016/j.erss.2017.12.003.

[45] C. P. Weaver, R. J. Lempert, C. Brown, J. A. Hall, D. Revell and D. Sarewitz, "Improving the contribution of climate model information to decision making: The value and demands of robust decision frameworks," *WIREs Climate Change*, vol. 4, no. 1, pp. 39–60, 2012, ISSN: 1757-7780 1757-7799. DOI: 10.1002/wcc.202.

[46] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006. DOI: 10.1191/1478088706qp063oa.

[47] A. Sharma, C. Wasko and D. P. Lettenmaier, "If precipitation extremes are increasing, why aren't floods?" *Water Resources Research*, vol. 54, no. 11, pp. 8545–8551, 2018, ISSN: 0043-1397. DOI: 10.1029/2018wr023749.

[48] S. C. Chan, E. J. Kendon, N. Roberts, S. Blenkinsop and H. J. Fowler, "Large-scale predictors for extreme hourly precipitation events in convection-permitting climate simulations," *Journal of Climate*, vol. 31, no. 6, pp. 2115–2131, 2018, ISSN: 0894-8755. DOI: 10.1175/Jcli-D-17-0404.1.

[49] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society: Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972. DOI: 10.2307/2344614.

[50] G. Bürger, T. Q. Murdock, A. T. Werner, S. R. Sobie and A. J. Cannon, "Downscaling extremes—an intercomparison of multiple statistical methods for present climate," *Journal of Climate*, vol. 25, no. 12, pp. 4366–4388, 2012, ISSN: 0894-8755 1520-0442. DOI: 10.1175/jcli-d-11-00408.1.

[51] J. Walker, A. Razavi and A. v. d. Oord, "Predicting video with vqvae," 2021. DOI: arxiv:2103.01950.