UNSUPERVISED MACHINE LEARNING TECHNIQUES FOR MULTI-MODEL COMPARISON: A CASE STUDY ON ANTARCTIC INTERMEDIATE WATER IN CMIP6 MODELS

Ophélie Meuriot[1], **Veronica Nieves**[2], **Yves Plancherel**[1]
[1]Imperial College London, Earth Science and Engineering, London, United Kingdom
[2]Image Processing Laboratory, University of Valencia, Valencia, Spain

ABSTRACT

The Climate Model Intercomparison Project provides access to ensembles of model experiments that are widely used to better understand past, present, and future climate changes. In this study, we use Principal Component Analysis and K-means and hierarchical clustering techniques to guide identification of models in the CMIP6 dataset that are best suited for specific modelling objectives. An example is discussed here that focuses on how CMIP6 models reproduce the physical properties of Antarctic Intermediate Water, a key feature of the global oceanic circulation and of the ocean-climate system, noting that the tools and methods introduced here can readily be extended to the analysis of other features and regions.

1 INTRODUCTION

Climate models play a significant role in advancing the understanding of complex Earth systems and provide important insights on the probable evolution and sensitivity of the climate system to climate change, past and future (IPCC, 2021). The Climate Model Intercomparison Project 6 (CMIP6) provides open access to a wide spectrum of state-of-the-art coupled model experiments. Transparent access to the CMIP6 archive is key as it provides scientists and stakeholders worldwide with the ability to independently compare the output of all climate models and thereby develop a sense of confidence about projections of future climate and quantify uncertainties (Eyring et al., 2016).

Although general climate trends are mostly consistent between models, evaluation of model-to-model differences reveals that substantial differences exist between model simulations on global and especially regional scales, and that these biases also vary through time. For instance, Wang et al., (2021) showed that the sea surface temperature biases differ significantly both regionally and seasonally between CMIP6 models. Tools and advanced methods, such as those from the field of artificial intelligence and machine learning, are needed to help analyse possible biases and extract physically meaningful information from massive and increasing amounts of model output.

Here, we show how a combination of unsupervised machine learning methods can be used to perform a novel objective-oriented multi-model comparison and help identify the models that best represent observed features of interest. This framework first aims to identify key individual variables/features using PCA analysis. Clustering methods are then used to decide how each model or group of models represent properties/processes of interest. The workflow is illustrated using a case-study that examines how CMIP6 models simulate the physical properties (e.g., salinity, temperature, density, outcrop latitude, or extent latitude) of Antarctic Intermediate Water (AAIW) in the Atlantic Ocean. AAIW is a key component of the ocean-climate system and has a major impact on the global ocean circulation. (Sloyan & Rintoul, 2001).

2 DATA AND METHODS

2.1 MULTI-MODEL ENSEMBLES OF CMIP6 CLIMATE MODELS AND OBSERVATIONAL DATA

In this study, we used output from 31 CMIP6 models, obtained from the Centre for Data Analysis (CEDA) Earth System Federation Grid node (https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/). The monthly mean ocean temperatures and salinities of each model were further averaged over the 1991-2010 period and re-gridded to regular 1° resolution products that are used for comparison with the World Ocean Atlas 2018 (Levitus et al. 2012) observational climatology (https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/) (see Appendix).

2.2 PHYSICAL PROPERTIES OF THE AAIW

Antarctic Intermediate Water is a water mass which forms in the Southern Ocean and spreads northward in all ocean basins. It is characterised by a strong mid-depth salinity minimum (the AAIW core). AAIW plays a fundamental role in the global ocean circulation, and acts as an important sink for atmospheric CO2 (Panassa et al., 2018). Despite its significance for the climate system, climate models struggle to capture and simulate the properties, extent, and dynamics of AAIW, as was shown by the analysis of Zhu et al. (2018) in the CMIP5 archive. We here analyse model biases for AAIW in CMIP6 models to illustrate how machine learning-based approaches can be used to help identify the models whose simulations are closest to the observations. Key properties commonly used to describe AAIW in the oceanographic literature were computed. The density was calculated using averaged ocean salinity and temperature. The depth of the AAIW core, as defined by the salinity minimum at 20°S, was identified. Note that 20°S is chosen as it is one of the latitudes where AAIW has subducted to the mid-depth of the oceans (Talley, 1999). The temperature, salinity, and density of the AAIW core at 20°S were then estimated from the re-gridded outputs. Other key variables related to the origin and extension of AAIW were calculated in the Atlantic at 25°W: 1) the outcrop location of AAIW in the Southern Ocean (i.e., the location where the AAIW core reaches the mixed-layer in the upper ocean); 2) the northernmost latitudinal extent of AAIW (maximum northward latitude at which the salinity minimum can be detected); 3) the depth extent at this northernmost latitude location (Zhu et al., 2018; Meuriot et al., 2022). In total we have 7 variables/features associated with AAIW (see Figure 1.a and Table 1 in the Appendix).

2.3 AN UNSUPERVISED LEARNING FRAMEWORK FOR OCEAN MODEL-DATA

Three unsupervised learning methods have here been applied to the analysis of CMIP6 model output for assessing AAIW properties across the Atlantic region: Principal Component Analysis (PCA), K-means clustering, and hierarchical clustering. The Scikit-learn library was used to process the data and apply the unsupervised learning methods. The pre-processing steps involved standardising the variables by centering and scaling each feature and replacing missing data with the mean value over all models. PCA is a highly efficient algorithm for dimensionality reduction, which involves reducing the number of input variables while minimising information loss (Tipping & Bishop, 1999). Clustering techniques help segmenting groups with similar properties and assign them into clusters. In the case of K-means clustering (Lloyd, 1982), it clusters data points based on similarity or closeness between the data points. The number of clusters was selected using the Elbow Curve method (Syakur et al., 2018) (see Figure 5.a in the Appendix). The optimal number was found to be three clusters. The K-means algorithm was run 20 times, with a different random initialization of the centroids (Lloyd, 1982). Hierarchical clustering is an alternative approach that identifies groups in a dataset by using (in this case) the minimum or average linkage clustering method (Miyamoto et al., 2015) (i.e., it essentially minimises the variance of the clusters being merged). The number of clusters was selected based on the analysis of the dendrogram structure (Figure 5.b in the Appendix), which was also three (Sembiring et al., 2010). The three unsupervised learning methods were applied on a matrix composed of 7 columns corresponding to the AAIW features (defined in Section 2.2) with 32 rows (31 CMIP6 models and 1 observational dataset).

3 Results and Discussion

3.1 Assessment of variable importance

The PCA analysis was applied to better interpret the importance of each variable defined to characterise AAIW (Section 2.2). The percentage of explained variance was calculated for each resulting principal component (PC). The first two (or three) PCs explain over 70% (or 85%) of the data variance (Figure 1.b). The scores of the top-ranking PCs were obtained for each variable to evaluate how much each variable contributes to the PC (Figure 1.c). The AAIW depth, density at 20°S, and the extent (in terms of depth) strongly influence PC1 but have limited influence on PC2 and PC3. The temperature at 20°S instead influences both PC1 and PC2. In contrast, the northernmost latitudinal extent, AAIW core salinity and temperature at 20°S dominate PC2. The outcrop position was found to have very little influence on PC1 and PC2 but dominates PC3.

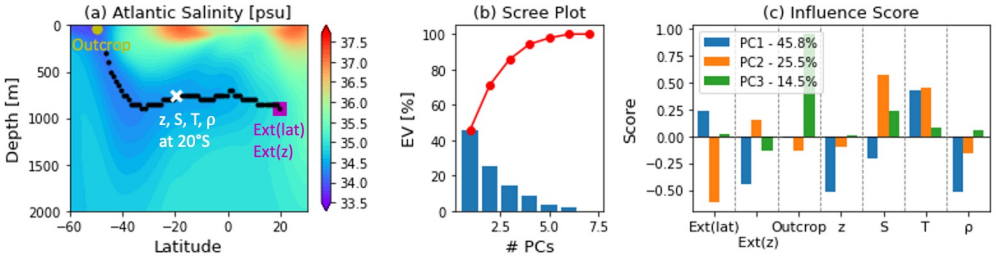

Figure 1: (a) Salinity section in the Atlantic Ocean at 25°W for the WOA18 dataset (observations). The core of AAIW, defined by the mid-depth minimum in salinity, is marked by the black dots. The variables compared in this study are the outcrop location (yellow dot), the temperature T, salinity S, depth z and density ρ at the core of AAIW at 20°S (white cross), and the depth and latitude of the extent (Ext(lat), Ext(z)) of AAIW (magenta square). (b) Scree plot showing the percentage of explained variance (EV) by each principal component (PC) (in red the cumulative sum). (c) Bar plot of the influence score of each variable on the PCs: for PC1 (blue), PC2 (orange) and PC3 (green).

3.2 Comparative evaluation of CMIP6 models over the AAIW layer

Two different machine learning clustering approaches (see Section 2.3) were applied to the 7 features of AAIW: hierarchical clustering and k-means clustering (shown projected onto the first two PCs in Figures 2.a and 2.b respectively). General consistent clustering results between methods provide more confidence of the analysis results. Both clustering approaches provide very similar model membership to specific clusters, with the exception of five models (circled in black in Figure 2.b). The clustering results are also consistent with the PC projection where the 3 groups can be visually distinguished. Overall, we found that models from Cluster 0 are closer to the results from observations (marked with a black square in Figure 2).

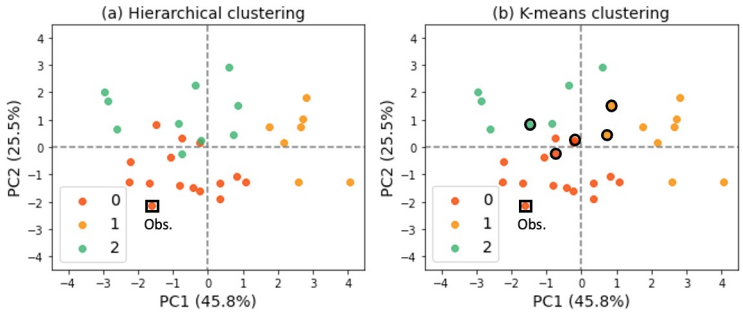

Figure 2: (a) Hierarchical and (b) K-means clustering results projected on PC1 and PC2. The black box highlights the results from the observational dataset. The black circles show the models for which the (hierarchical and K-means) clustering results differ.

A physical attribution of the clusters is often difficult; however, it is possible to distinguish different features. By further investigating each individual cluster projected on the PC1 and PC2 components, we noted that the models in Cluster 1, which have higher PC1 values (Figure 2), are associated with lower depth, higher temperature, and lower density (see Figure 3 for the hierarchical case). This is in agreement with the PCA analysis in Section 3.1 (see Figure 1.c). The models in Cluster 2 show a reduced northward extent and have higher salinity values, whereas Cluster 0 models shows further northward extent and the opposite pattern in terms of salinity. Hence, these results indicate that several characteristics differentiate clusters and AAIW features. Furthermore, we found the group of models (Cluster 0) that reproduces most closely the latitudinal extent, a feature that was not properly captured in CMIP5 models (Zhu et al., 2018).

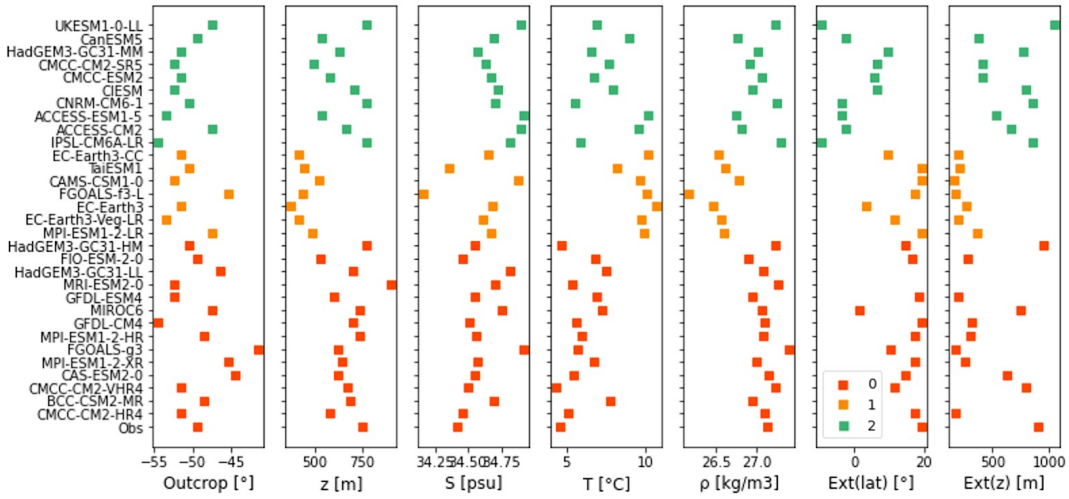

Figure 3: Illustration of the different clustering groups of models for each variable (described in Section 2.2). Results from the hierarchical method. Cluster 0, 1, 2 highlighted in red, orange, and green, respectively.

4 Conclusions

The work presented here illustrates the potential of unsupervised learning techniques in the context of model inter-comparison, where a large number of models are used. A case-study analysis of the AAIW physical properties presented here shows that PCA analysis of CMIP6 multi-model datasets can help identify those features that contribute the most in terms of explained variance. The clustering techniques can help diagnose and select models which are more in line with observations (or with any reference). They can also be used to identify the dominant differences between the clusters. In the example discussed here, each resulting cluster appears to be associated with different physical properties, suggesting different physical underlying processes or mechanisms, information that is useful to guide further research. This study provides a workflow for unsupervised learning techniques that can be easily generalized to other model inter-comparison experiments.

REFERENCES

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev, 9. https://doi.org/10.5194/gmd-9-1937-2016

IPCC, Masson-Delmotte, V., Zhai, P., Chen, Y., Goldfarb, L., Gomis, M. I., Matthews, J. B. R., Berger, S., Huang, M., Yelekçi, O., Yu, R., Zhou, B., Lonnoy, E., Maycock, T. K., Waterfield, T., Leitzell, K., & Caud, N. (2021). Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change Edited by. https://doi.org/10.1017/9781009157896

Lloyd, S. P. (1982). Least Squares Quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Meuriot, O., Lique, C., & Plancherel, Y. (2022). Properties, sensitivity, and stability of the Southern Hemisphere salinity minimum layer in the UKESM1 model. Climate Dynamics, 1, 1–21. https://doi.org/10.1007/S00382-022-06304-2/

Miyamoto, S., Abe, R., Endo, Y., & Takeshita, J. I. (2015). Ward method of hierarchical clustering for non-Euclidean similarity measures. Proceedings of the 2015 7th International Conference of Soft Computing and Pattern Recognition, SoCPaR 2015, 60–63. https://doi.org/10.1109/SOCPAR.2015.7492784

Panassa, E., Santana-Casiano, J. M., González-Dávila, M., Hoppema, M., van Heuven, S. M. A. C., Völker, C., Wolf-Gladrow, D., & Hauck, J. (2018). Variability of nutrients and carbon dioxide in the Antarctic Intermediate Water between 1990 and 2014. Ocean Dynamics, 68(3), 295–308. https://doi.org/10.1007/s10236-018-1131-2

Sembiring, R. W., Zain, J. M., & Embong, A. (2010). A Comparative Agglomerative Hierarchical Clustering Method to Cluster Implemented Course. 2. http://arxiv.org/abs/1101.4270

Sloyan, B. M., & Rintoul, S. R. (2001). Circulation, renewal, and modification of antarctic mode and intermediate water. Journal of Physical Oceanography, 31(4), 1005–1030. https://doi.org/10.1175/1520-0485(2001)031¡1005:CRAMOA¿2.0.CO;2

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. IOP Conference Series: Materials Science and Engineering, 336(1). https://doi.org/10.1088/1757-899X/336/1/012017

Talley, L. D. (1999). Some aspects of ocean heat transport by the shallow, intermediate and deep overturning Circulations. Geophysical Monograph Series, 112, 1–22. https://doi.org/10.1029/GM112P0001

Tipping, M. E., & Bishop, C. M. (1999). Mixtures of probabilistic principal component analyzers. Neural Computation, 11(2), 443–482. https://doi.org/10.1162/089976699300016728

Wang, Y., Heywood, K. J., Stevens, D. P.,& Damerell, G. M. (2021). Seasonal extrema of sea surface temperature in CMIP6 models. Ocean Science, 18(3), 839–855. https://doi.org/10.5194/os-18-839-2022

Zhu, C., Liu, Z., & Gu, S. (2018). Model bias for South Atlantic Antarctic intermediate water in CMIP5. Climate Dynamics, 50(9–10), 3613–3624. https://doi.org/10.1007/s00382-017-3828-1

APPENDIX

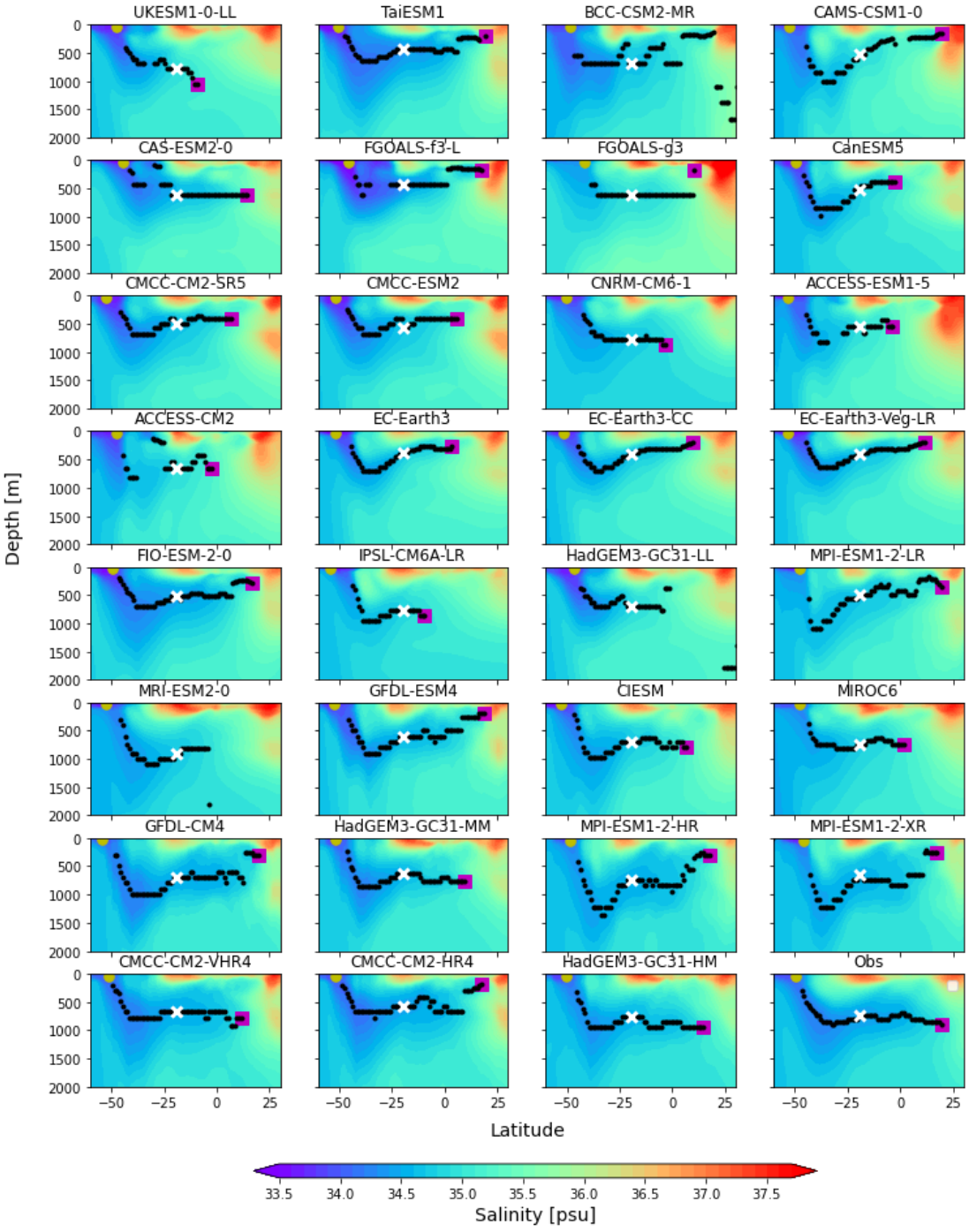

Figure 4: (Same as in Figure 1, but for the 31 CMIP6 models. In the bottom right panel, the salinity section at 25°W from observations to facilitate visual comparison of patterns.

Models	Ext(lat) [°]	Ext(z) [m]	outcrop [°]	z [m]	S [psu]	T [°C]	ρ [kg/m3]	Cluster Hiearchical	Cluster K-means
BCC-CSM2-MR			-48.5	689	34.69	7.7	26.94	0	0
CAS-ESM2-0	14.5	621	-44.5	621	34.55	5.4	27.14	0	0
CMCC-CM2-HR4	17.5	179	-51.5	576	34.45	5.1	27.1	0	0
CMCC-CM2-VHR4	11.5	789	-51.5	675	34.5	4.4	27.22	0	0
FGOALS-g3	10.5	178	-41.5	621	34.92	5.7	27.4	0	0
FIO-ESM-2-0	16.5	285	-49.5	527	34.46	6.8	26.9	0	0
GFDL-CM4	19.5	313	-54.5	700	34.51	5.6	27.09	0	0
GFDL-ESM4	18.5	200	-52.5	600	34.55	6.9	26.94	0	0
HadGEM3-GC31-HM	14.5	947	-50.5	773	34.55	4.7	27.23	0	0
HadGEM3-GC31-LL	NaN	NaN	-46.5	697	34.82	7.5	27.08	0	0
MIROC6	1.5	740	-47.5	740	34.75	7.3	27.06	0	2
MPI-ESM1-2-HR	17.5	310	-48.5	740	34.56	6	27.08	0	0
MPI-ESM1-2-XR	17.5	263	-45.5	645	34.57	6.8	26.99	0	0
MRI-ESM2-0	NaN	NaN	-52.5	905	34.7	5.4	27.27	0	0
Obs	19.5	900	-49.5	750	34.42	4.6	27.13	0	0
CAMS-CSM1-0	19.5	165	-52.5	522	34.88	9.7	26.78	1	1
EC-Earth3	3.5	271	-51.5	371	34.68	10.7	26.46	1	1
EC-Earth3-CC	9.5	200	-51.5	412	34.66	10.2	26.53	1	1
EC-Earth3-Veg-LR	11.5	200	-53.5	412	34.62	9.7	26.57	1	1
FGOALS-f3-L	17.5	178	-45.5	433	34.16	10	26.17	1	1
MPI-ESM1-2-LR	19.5	363	-47.5	485	34.68	9.9	26.6	1	1
TaiESM1	19.5	210	-50.5	443	34.36	8.2	26.61	1	1
ACCESS-CM2	-2.5	665	-47.5	665	34.89	9.6	26.82	2	2
ACCESS-ESM1-5	-3.5	537	-53.5	537	34.92	10.1	26.74	2	2
CanESM5	-2.5	382	-49.5	534	34.7	9	26.76	2	1
CIESM	6.5	787	-52.5	708	34.72	7.9	26.94	2	2
CMCC-CM2-SR5	6.5	416	-52.5	490	34.63	7.7	26.91	2	1
CMCC-ESM2	5.5	416	-51.5	576	34.67	6.8	27.06	2	0
CNRM-CM6-1	-3.5	857	-50.5	773	34.71	5.5	27.25	2	2
HadGEM3-GC31-MM	9.5	773	-51.5	628	34.57	6.5	27.02	2	0
IPSL-CM6A-LR	-9.5	857	-54.5	773	34.81	5.9	27.29	2	2
UKESM1-0-LL	-9.5	1046	-47.5	773	34.89	6.9	27.22	2	2

Table 1: AAIW variables for each CMIP6 model. The last two columns show the clustering results from the hierarchical and K-means clustering methods, respectively.

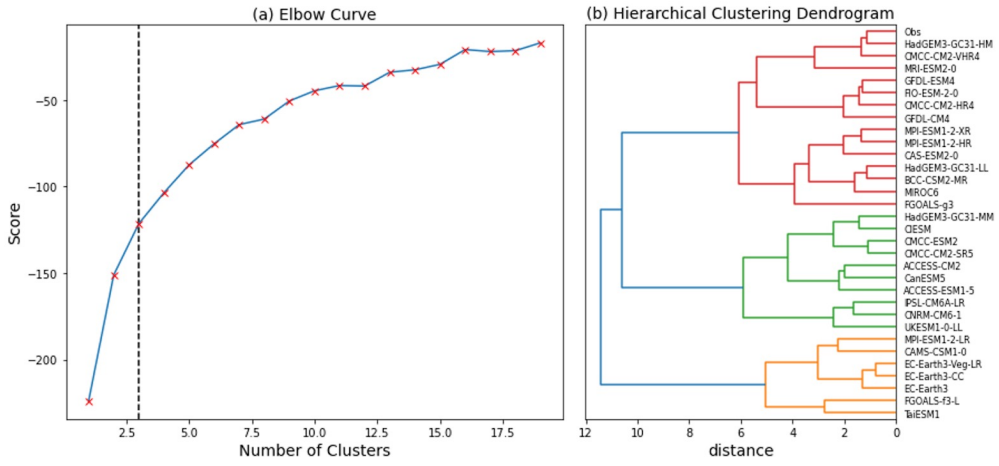

Figure 5: (a) Elbow curve showing the score against the number of clusters for the K-means clustering method. The number of clusters is chosen by finding the position where the curve starts flattening out. (b) Dendrogram, a tree like structure showing the main three clusters as a function of the distance between the models.