

Baseline Climate Variables for Earth System Modelling

Martin Jukes¹, Karl E. Taylor⁵, Fabrizio Antonio⁸, David Brayshaw², Carlo Buontempo³, Jian Cao⁴, Paul J. Durack⁵, Michio Kawamiya⁶, Hyungjun Kim⁷, Tomas Lovato⁸, Chloe Mackallah⁹, Matthew Mizieliński¹⁰, Alessandra Nuzzo⁸, Martina Stockhause¹¹, Daniele Visioni¹², Jeremy Walton¹⁰, Briony Turner¹³, Eleanor O'Rourke¹³, Beth Dingley¹³.

¹Univesrity of Oxford, and UKRI STFC, UK

²University of Reading, UK

³ECMWF, Bonn, Germany

⁴Earth System Modeling Center, Nanjing University of Information Science and Technology, Nanjing, China

⁵PCMDI, Lawrence Livermore National Laboratory, Livermore, USA

⁶JAMSTEC, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

⁷Korea Advanced Institute of Science and Technology, Korea

⁸CMCC Foundation - Euro-Mediterranean Center on Climate Change, Italy

⁹CSIRO, Climate Science Centre, Australia

¹⁰Met Office Hadley Centre, UK

¹¹DKRZ, Germany

¹²Department of Earth and Atmospheric Sciences, Cornell University, USA

¹³CMIP International Project Office, UK

Correspondence to: Martin Jukes (martin.jukes@physics.ox.ac.uk)

Abstract.

The Baseline Climate Variables for Earth System Modelling (ESM-BCVs) are defined as a list of 135 variables which have high utility for the evaluation and exploitation of climate simulations. The list reflects the most heavily used elements of the Coupled Model Intercomparison Project phase 6 (CMIP6) archive. Successive phases of CMIP have supported strong results in science and substantial influence in international climate policy formulation. This paper responds both to interest in exploiting CMIP data standards in a broader range of climate modelling activities and a need to achieve greater clarity about the significance and intention of variables in the CMIP Data Request. As Earth System Modelling (ESM) archives grow in scale and complexity there are emerging problems associated with weak standardisation at the variable collection level. That is, there are good standards covering how specific variables should be archived, but this paper fills a gap in the standardisation of which variables should be archived. The ESM-BCV list is intended as a resource for ESM Model Intercomparison Projects (MIPs) developing requests to enable greater consistency among MIPs, and as a reference for modelling centres to enhance consistency within MIPs. Provisional planning for the CMIP7 Data Request exploits the ESM-BCVs as a core element. The baseline variables list includes 98 variables which have modest or minor data volume footprints and could be generated systematically when simulations are produced and archived for exploitation by the WCRP

community. A further 35 variables are classed as high volume and are only suitable for production when the resource implications are justified.

1 Introduction

1.1 Context and motivation

With the publication of the Baseline Climate Variables for Earth System Modelling (hereafter ESM-BCV; see end of Section 4 for discussion of the name) we aim to address the growing need for climate model data archives to have more consistency between projects and between generations of models. We exploit substantial resources and knowledge that have been developed through the Coupled Model Intercomparison Project (CMIP; see Meehl et al. 1997). CMIP was established to collect data from models that could represent some aspects of the atmospheric, oceanic, land, and cryospheric components of the climate system and has grown over successive phases (Meehl et al., 2000, 2007; Taylor et al. 2012; Eyring et al., 2016) to provide both better representation of those processes and more complete coverage of the Earth system, including chemical, biogeochemical and ecosystem processes. CMIP has also expanded from the initial focus on model evaluation to become "a central element of national and international assessments of climate change" (Eyring et al., op. cit.). The CMIP community has led the way in developing climate model archives as a community resource with a range of users which extends far beyond the modelling centres responsible for developing models and delivering data products. The content of the archive is guided by the CMIP Data Request (CMIPDR; see Figure 1). The latest iteration of this request for CMIP6 (Juckes et al. 2020) contained over 2000 variables, a significant increase over the 970 variables requested for CMIP5 (PCMDI, 2013). The CMIP6 Data Request (CMIP6DR) collated data requirements from dozens of international science projects to create a database of climate variables indexed against priorities, objectives and experimental configurations. The CMIP6DR was seen by many as being too extensive and the mechanisms provided to enable data producers to filter the request down to an appropriate level were not able to compensate for this. A lack of clarity about priorities detracted from consistency of archive content (Section 1.3 below). The ESM-BCVs will provide a clear focus to enable greater consistency both within CMIP and between CMIP and other model intercomparison activities. It is, however, as the name suggests, only a baseline and further variables will generally be needed in many cases. This caveat notwithstanding, the majority of users are interested in a modest subset of the 2000+ variables.

The CMIP6 Data Request

The CMIP6 Data Request **defined all of the variables which should be output** from CMIP6 simulations. It consists of:

- An XML database with the data requirements for all MIPs,
- A browsable online version,
- A python library providing both a programmable and a command line interface.



What
is it?

Why was
it created?

CMIP6 was a global collaborative effort in which dozens of climate models ran a **coordinated set of experiments**. This created an extensive data archive to support **evaluation and intercomparison** of the climate models.

Data requirements were collected from the **23 CMIP6-endorsed MIPs**.

20 of the MIPs represented different scientific areas. Three MIPs represented the interests of communities which do not engage directly with the climate models:

- CORDEX requested data required to drive regional climate models
- ScenarioMIP did not make any data requests, but specified experiments to study climate change scenarios
- VIACS-AB compiled data requirements from climate change impacts and services experts



How was
it created?



Perceived
issues

The **scale** and **complexity** of the data request:

- The diversity of both those proposing requests and the facilities offering services led to a large, and highly complex data request.

Too many **high priority variables**:

- Discussion at WGCM 22 (Barcelona, 2019) highlighted around 50% of variables were designated as priority one (Jukes, 2020)

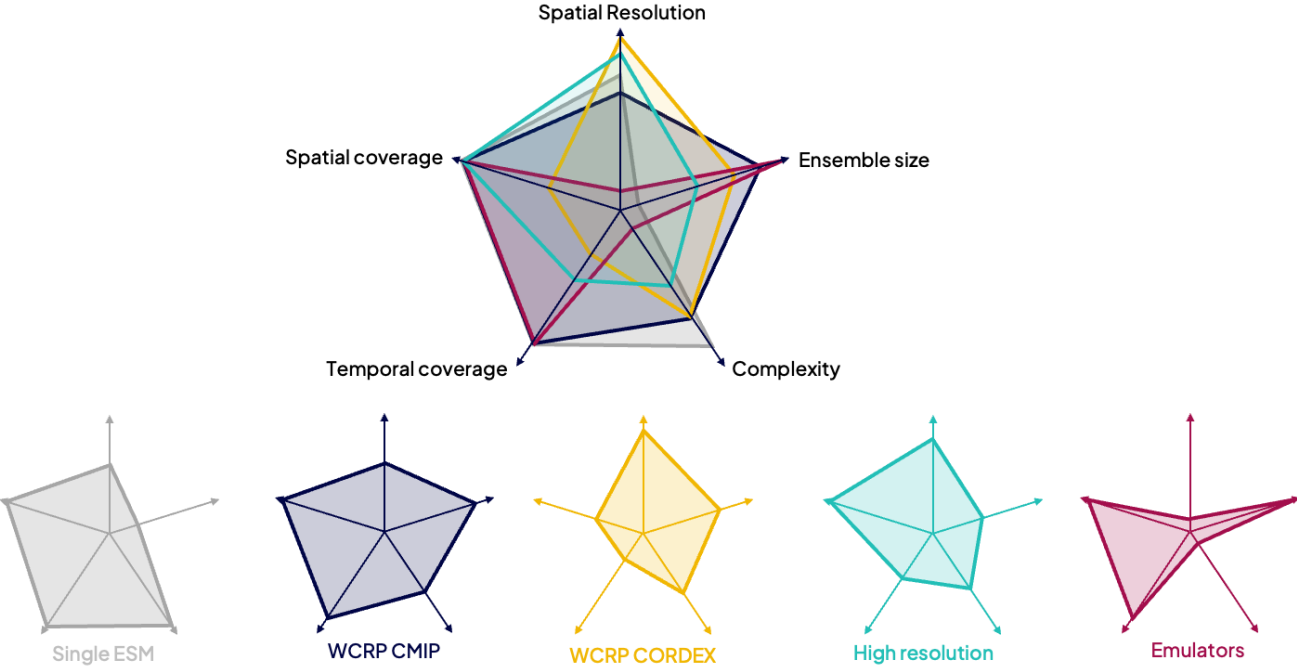
61

62 **Figure 1: CMIP6 Data Request storyboard**

63 **1.2 Expanding scope and impact of Earth system modelling**

64 The scientific scope of the climate models used to analyse the impact of humanity on the global climate is continually
65 expanding (e.g. Flato, 2011) and the community is now experimenting with kilometre resolution models (e.g. Hohenegger et
66 al., 2023) and explicit modelling of human behavioural response to climate (e.g. Tan et al., 2023). A review of this diverse
67 and growing literature is beyond the scope of the current paper, but it is clear that preservation of clarity and interoperability
68 of existing and future data products will be a challenge for this wide-ranging community. As the range of modelling
69 activities has expanded, a diverse range of models and model configurations has emerged to target different areas of climate
70 science, resulting in a multiverse of models (Figure 2).

The WCRP Modelling Multiverse



71
72 **Figure 2: The Modelling Multiverse.** The phase space covered by each climate modelling endeavour within the WCRP. Each type
73 of model or modelling project has a different ability to model over different spatial resolutions, spatial coverage, temporal
74 coverage, model complexities, and ensemble sizes. Each model type or modelling project is exemplified using a different colour.
75 The elements of the radar charts are:

- 76 ● **Spatial resolution:** the ability to resolve fine scale spatial features,
77 ● **Ensemble size:** the ability to resolve details of internal variability,
78 ● **Complexity:** the ability to resolve a wide range of physical and bio-geological climate processes,
79 ● **Temporal coverage:** the ability to cover centennial time scales,
80 ● **Spatial coverage:** the ability to cover the complete globe.

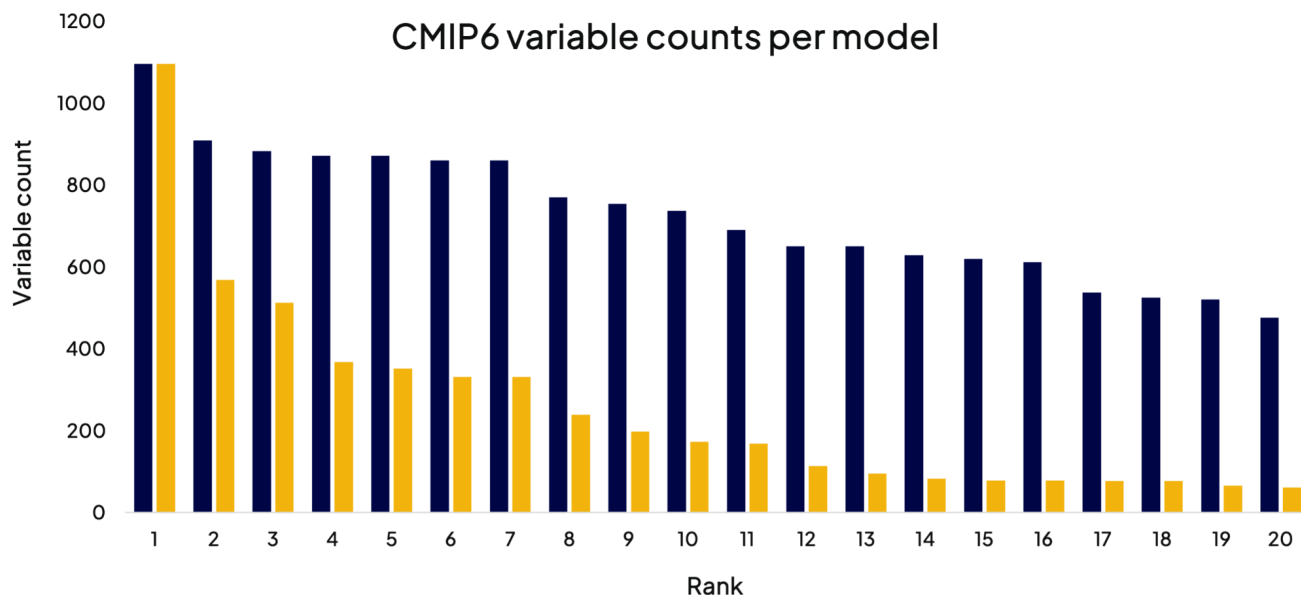
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103

The exchange of interoperable climate model output across multiple Model Intercomparison Projects (MIPs) is now a mainstay of climate science and climate assessment, feeding into the development of policies on climate change mitigation and adaptation. Scientific work supported by CMIP has become the foundation for Intergovernmental Panel on Climate Change (IPCC) assessment reports which are alerting humanity to the risks of catastrophic climate change (Touzé-Peiffer et al., 2020), driving international commitments to decarbonisation of the economy (Paris Agreement, United Nations, 2015; Guterres, 2023).

With the growth in the scale and complexity of the models and the intercomparison projects that investigate their behaviour there is a growing interest in multi-variable multi-model analyses. There is an emerging requirement for consistent provision of variable collections across simulations generated by the entire WCRP multiverse of models. For robust simulation and analysis of the climate system on centennial time-scales multi-model ensembles are required. Through multiple phases of CMIP, an open and evolving community approach to creating intercomparisons which span multiple MIPs and all the elements of the WCRP Multiverse has been established. We refer to the collection of simulations generated through these activities as a MultiVerse Ensemble (MVE).

The success of MVEs in creating value which is greater than the sum of the parts has led to a growing ecosystem of MIPs and other community activities coordinating the specification of science goals, experimental configuration and data requirements for MVEs. Data requirements now must serve not only climate researchers but also a diverse community of stakeholders that rely on climate model output. Textual analysis of the 5152 Web of Science publications¹ which, on 24th August 2023, referenced CMIP6, shows two main clusters, one associated with model and climate system analysis and experiments and the other associated with impacts, adaptation, and scenarios (Figure 3). This analysis shows clearly how scenarios and impacts cluster has become of equal significance, in terms of quantity of publications, with the climate science research area.

9 1 The analysis is based on titles and abstracts of 5152 papers identified from Web of Science which either cite Eyring et al.
10 (2016) or mention CMIP6 in the title or abstract. The clustering is based on terms which occur in at least 100 papers.



109

110 **Figure 4: Variable provision in CMIP6.** The number of variables (y-axis) published for the historical simulation by each model (as
 111 represented in the DKRZ Earth System Grid Federation (ESGF) index node August 2023) is shown in blue columns against the
 112 model rank, where models are ranked in order of decreasing variable count. Also shown, in orange, is the number of variables
 113 which are included by all models up to the given rank. For comparison, the total number of variables requested by all MIPs from
 114 the CMIP6 historical simulation was 2301, with 1484 of those assigned priority one by one or more MIPs.

115 1.3 Objectives of the Earth System Modelling Baseline Climate Variables list

116 As the name suggests, the list presented here is intended to define a baseline set of climate variables which can be produced
 117 by ESM activities, and which are of widespread interest. By including a rather limited subset of commonly analysed
 118 variables, we expect that modelling groups should easily be able to routinely provide all variables and that data centres
 119 should be able to accommodate the generated data volumes. For indirect users such as the climate and climate impacts
 120 research communities, the variables in the baseline set will facilitate consistent and efficient comparison of simulations
 121 across multiple intercomparison projects, both within and between existing and future CMIP eras, by enhancing
 122 standardisation at the variable collection level (see Figure 4 and discussion in Section 1.4 for the motivation behind this
 123 objective).

124 Use of the term "Earth System Modelling" in describing the list is meant to convey that these variables should be of interest
 125 from a wide range of models used in studying the climate of the Earth system. This includes, for example, not only models
 126 which have a detailed representation of interactions between the physical climate and the biosphere but also simpler models
 127 which play a role in advancing understanding of critical elements of the Earth system.

128 Although the list serves as a “baseline”, it is not expected to be sufficient in addressing many of the specific science
 129 questions that are the focus of MIPs. Invariably, additional variables will be of value and, in some cases, essential in
 130 interpreting and understanding simulation results. There may also be some model intercomparison experiments that focus on

131 a single aspect of the Earth system where many of the baseline variables will be irrelevant or of little interest. As a trivial
132 example, in the case of an atmospheric model run with prescribed sea surface conditions, all the baseline ocean variables,
133 except sea surface temperature and sea ice fraction, will be irrelevant. On the other hand, none of the variables characterising
134 bio-geochemical cycles and atmospheric chemistry appear in the baseline list even though they would be essential in
135 understanding those aspects of the Earth system.

136 Even if the list cannot meet all the requirements of MIPs, it can be considered the minimal suite of variables to be archived
137 from simulations meant to serve a broad range of WCRP stakeholders. For the climate and climate impacts research
138 communities, the variables in the baseline list will enable consistent and efficient comparison of simulations across multiple
139 intercomparison projects, both within and between existing and future CMIP eras. The baseline list of variables may also
140 nurture development of evaluation tools once there is an expectation that a consistent set of climate variables will be made
141 available from many MIP experiments.

142 The ESM-BCV will also provide a basis for comparison with parameter lists widely used in different communities, such as
143 the variables used for exchange of meteorological observations in the GRIB² protocol, the Essential Climate Variables
144 (ECVs: World Meteorological Organisation 2022a,b), or the Global Climate Indicators (GCIs)³ concept in climate services.

145 **1.4 Variable output by model**

146 The CMIP6 archive contains a comprehensive range of data products, with 72⁴ models contributing to the “all-forcing
147 simulation of the recent past (historical)”⁵, but users looking for data to support multivariable analysis can run into problems
148 because of lack of consistency in the selection of variables which are available for each model. Thus, although there are 25
149 models providing 390 or more variables for the “historical” simulation (see Eyring et al., 2016), the number of variables
150 which those models have in common is only to 57 (see Figure 4)⁶. This lack of consistency can force analysts to be selective
151 about the models included in any analysis and lead to lack of interoperability between derived products. If, for instance, a
152 drought indicator is based on a cluster of models “A” which have a full range of precipitation, runoff and evapotranspiration
153 variables at monthly frequency and the growing season indicator is based on a cluster of models “B” which have daily
154 precipitation, cloud cover and temperature variables, including daily extremes, the differences between clusters “A” and “B”
155 may hamper combined use of the two products. If set A is defined by models which have, for the historical, ssp126 and
156 ssp245 experiments, variables Amon.pr, Lmon.mrro, Lmon.evpspsblveg, Lmon.evpspsblsoi in the CMIP6 naming conventions
157 and set B is defined by models which have day.pr, day.tasmin, day.tasmax, and day.clt for the same experiments, then set A
158 has 34 models from 20 institutions, set B has 27 models from 19 institutions and the intersection is 20 models from 14

17 2GRIB (General regularly distributed information in binary form) is the WMO standard for operational exchange of
18 meteorological data (World Meteorological Organisation, 2023).

19 3 <https://gcos.wmo.int/en/global-climate-indicators>

20 4 This discussion is based on information from the ESGF index, accessed 24 August, 2023.

21 5 The all-forcing experiment of the recent past (historical) in CMIP is designed to enable the evaluation of model
22 simulations against present climate and observed climate change.

23 6 Data publication for CMIP6 is still ongoing, but the pattern of gaps in the archive persists as data volumes expand.

institutions. Publication of CMIP6 data is ongoing and details may evolve, but the patterns of inconsistency seen here represent a snapshot of the data landscape which confronts users dealing with the archive now.

1.5 Stakeholder groups

CMIP, and hence the CMIPDR, has an extensive community of stakeholders. Table 1 lists the main stakeholder groups. Some of these (darker shading) have a direct interest in the specific variables which are requested, archived and disseminated. Others (lighter shading) are more concerned with derived products and messages and with the level of reliability and trust which can be associated with those products and messages. The existence of a set of baseline variables which is available consistently from virtually all models and experiments is of particular importance to this second group because they often use derived products which depend on multiple variables from multiple models and experiments.

2 Process and Methodology

The 2022 CMIP6 Community Survey (O’Rourke, 2023) received over 300 responses. There was very clear appreciation for the coordination effort and the principles behind the CMIP6DR but many respondents did suggest that there were too many variables assigned priority “1” and this placed a burden on the modelling centres⁷. These responses reflected the discussion at the conference held by the WCRP Working Group on Coupled Modeling (2019) in Barcelona at which a community intention to reduce the number of variables at priority “1” from around 50% to a significantly smaller number emerged, with a suggestion to start with those prioritised by AR6 WG1 (see Juckes, 2020). The 2022 CMIP6 community survey also received many responses highlighting a need for additional variables including increased temporal resolution, more ocean variables, variables relevant to extremes, as well as those variables required to support the CORDEX (Gutowski et. al. 2016) regional downscaling community and their downstream users. These requirements for additional variables are not addressed by the baseline list.

180
181
182
183
184

⁷ The prioritisation of variables in the CMIP6 Data Request was always conditional on an objective such as support for a specific MIP. For example, a variable might be priority 1 for SIMIP (Seaice MIP) but of no interest for LUMIP (Land Use MIP).

	Title	Description	Example or reference
DIRECT USERS	Climate System Science	Communities studying the global climate, including those studying geographically localised processes and their role in the global climate.	The endorsed CMIP6 MIPs; research teams and individual researchers at all career stages.
	Science of Climate Impacts and Mitigation	- Communities studying the impact of climate change and variability on environmental systems and socio-economic sectors. - Regional climate modelling.	VIACSAB, ScenarioMIP, CORDEX and GeoMIP.
	Climate modelling	Institutions and networks developing and running climate models	Institutions contributing to CMIP6.
	Climate Research Infrastructure	- Data centres supporting curation, dissemination and analysis. - Software libraries and services, standards, protocols.	
	Climate Service	- Publicly funded organisations providing climate information and related services for public consumption. - Not-for-profit organisations providing climate services. - Commercial organisations providing support for customers.	There is a large and growing ecosystem of climate service providers. Examples include: C3S, European Environment Agency, consulting firms.
INDIRECT/DOWNSTREAM USERS	Providing support for those impacted by climate change	Organisations that work with individuals and communities which are being impacted by climate change. This could be seen as a category of climate service but is included to emphasise the significance of this role.	World Bank, United Nations Environment Programme (UNEP), Adaptation Fund, IPCC and United Nations Framework Convention on Climate Change (UNFCCC).
	Public decision makers	The decision makers often rely on information from the downstream products provided by climate services and consultancies, some of which might be derived in part from CMIP data.	Parties to the UNFCCC, Local and national policy and decision makers.
	Commercial organisations impacted by climate change	Anything from the Panama Canal to a fruit orchard in Normandy, climate change will impact all sectors of society. Most critically, it is starting to impact the habitability of some cities and the security of food supply for many. They may have internal services, or procure services, or be supported by sectoral interest groups/representative bodies.	
	Concerned Public	The public may get their information from news bulletins, but key messages are often derived from CMIP and related activities.	

Table 1: CMIP data request stakeholder groups

2.1 Launch and scoping workshops

The consultation process was launched in April 2022 by the CMIP International Project Office (IPO) with a request for feedback on the proposed process, an invitation to scoping meetings, and a target of establishing “a baseline set of variables for exchange of climate model data” (see Appendix 4). The announcement was sent to modelling centres and Data Request

190 leads, the MIP Chairs and circulated by the World Climate Research Programme. Responses, 32 in all, were received from
191 respondents across Asia, Europe and North America, whose CMIP6 involvement ranged from Data Request leads, modelling
192 centre leads and MIP chairs, to users of CMIP data, for scientific and climate impact modelling as well as climate services
193 provision. The findings from this survey were discussed at two scoping workshops held on May 12th and 17th 2022. The
194 focus of the workshops was on finalising the process for defining the variable list, creating an author team for this paper, and
195 creating an outline of the paper structure.

196 The scoping workshop report includes direction for authors to focus on clarifying the purpose and function of the list and
197 identifying the requirements of user groups.

198 There was also concern about the selection criteria. There is clear agreement on the need for a baseline list and a recognition
199 of the utility of such a list for many user communities, and a high level of support for the process of expert elicitation
200 adopted. Some contributors argued for a process based on defining specific variable selection criteria which could be applied
201 consistently to every variable in the list, but there were no specific proposals for such criteria. Instead, in line with the
202 established approach in the CMIP6DR, the process adopted was to ask experts to consider the list against the agreed
203 objectives (see section 1.2 above).

204 **2.2 Shortlisting from the CMIP6 request**

205 The initial shortlist of baseline variables was arrived at based on the CMIP6 archive's model output statistics, which gauge
206 the willingness of modelling groups to report each variable and the user demand for each variable reported. The resulting
207 shortlist of variables was then edited and augmented, based on community input.

208 Selection of an initial shortlist of variables was based on the variables requested for CMIP6 but excluding all but priority-1
209 variables. Three scores were calculated ranking the variables according to the number of models contributing, volume of
210 data downloaded, and number of files downloaded. The shortlist provided a starting point for the consultation and expert
211 discussion.

212 The formal steps taken were as follows:

213 (1) Extract the list of 1206 variables assigned default priority 1 in CMIP6, out of a total of 2062.

214 (2) For each variable, assign three ranking scores, r1, r2, r3:

215 (a) r1: ranked according to the volume of data downloaded across the entire CMIP6 archive, retrieved from the
216 ESGF dashboard (Fiore et al., 2021)⁸.

217 (b) r2: ranked according to the number of files downloaded across the entire CMIP6 archive, retrieved from the
218 ESGF dashboard.

219 (c) r3: ranked according to the number of models that provided the variable for the CMIP6 historical experiment.

33 8 <http://esgf-ui.cmcc.it/esgf-dashboard-ui/>; the download statistics are from the server log files which record successful
34 responses to requests received over HTTP, including requests from scripts and from browsers. Some usage is not monitored,
35 such as multiple users accessing a shared processing space. Open access has been prioritised at the expense of
36 comprehensive usage information, but the majority of users are still accessing data via the mechanisms which do get tracked.

220 (3) Rather than weighting the criteria we take the minimum rank, $r_{\min} = \min(r_1, r_2, r_3)$.
221 (4) Define the shortlist as the first 125 variables ordered by $r_{\min}$, together with their supporting fixed fields (which are
222 necessary for correct interpretation of the data, e.g, grid cell area, or volume).
223 For details of the variables which were included in the shortlist see Appendix 1 and 2.

224 2.3 Community survey and analysis

225 Following the creation of a shortlist, a community survey was designed to elicit expert feedback on the initial list. The
226 survey was targeted at those providing access to and/or utilising the outputs of climate models within commercial, public and
227 voluntary sectors together with academia. The survey was circulated to the CMIP mailing lists for modelling centres, data
228 request, MIP chairs, circulated by WCRP and the author team and promoted through CMIP social media channels and was
229 open to respondents for a period of just over six weeks between 23 August and 8th October, 2022. From the 44 responses
230 received, the majority identified as climate data users, 12 identified as climate model data providers. The shortlisted
231 variables were reviewed in detail by 29 respondents: these respondents were invited to review a selection of variables
232 relevant to their expertise or data usage. 16 respondents reviewed 5 variables or less, the remainder reviewed a larger
233 selection. A scoring methodology was provided to ensure review consistency. A full summary report of the survey responses
234 has now been published (O'Rourke and Turner, 2022; see also the survey announcement in Appendix 4).

235 2.4 Shortlist Revision and Consequences

236 In two further author team meetings in late 2022, the results of the survey were discussed and analysed in depth to consider
237 potential additions and deletions of some shortlisted variables.
238 In early 2024 checks of the ESGF dashboard revealed a previously undetected error in reporting the download statistics that
239 were relied on in arriving at the initial short list of variables. Data transfers associated with unsuccessful requests for partial
240 file downloads over very low capacity networks had been misreported in log files as successful, exaggerating the user
241 demand for some variables. The team at CMCC Foundation (Euro-Mediterranean Center on Climate Change) responsible for
242 the dashboard were able to provide corrected download statistics based on a reanalysis of log files. The corrected download
243 reports were used to re-assess variables in the ESM-BCV list agreed on in 2022, resulting in four variables being removed
244 and a different four being added (see Appendix 2, tables A5 and A6, for details of individual variables).
245 Further discussions by authors and a final meeting in June 2024 led to a review of the criteria for fixed model configuration
246 fields (they were retained if more than 12 models had provided the variable for at least one experiment).

247 3 The form and role and the baseline list

248 The variable list presented here will be a baseline set of variables for global model intercomparison, evaluation, and
249 exploitation projects and programmes. This is intended as a starting point for more comprehensive lists tailored to specific

250 applications. Many of the variable definitions in the list are used in modelling activities across the whole scope of WCRP
251 activities, either through MIPs associated with CMIP (particularly in the Climate and Cryosphere, CliC, and Climate and
252 Ocean Variability, Predictability and Change, CLIVAR, core projects) or for output from activities such as CORDEX
253 (Regional Information for Society, RIfS, and, Global Energy and Water Exchanges, GEWEX, core projects) and the
254 Chemistry-Climate Model Initiative (part of the Atmospheric Processes and their Role in Climate, APARC core project)
255 which are shadowing CMIP data protocols: the ESM-BCV list will support progress towards greater consistency and
256 interoperability in data outputs from this extensive range of activities.

257 **3.1 Form of the list**

258 The baseline variable list should also provide a model for clarity and interoperability. This scope of this paper covers
259 selecting and defining the physical quantities along with their spatio-temporal sampling structures.
260 Some variables are categorised as "high volume" and should be considered as optional when resource constraints apply.
261 These variables have a particularly high value for many users, but are likely to be too resource intensive for many climate
262 simulations. They are included so that they can benefit from the visibility afforded by the baseline list, but are not expected
263 to be systematically produced to the same extent as the other variables in the list.
264 This paper is concerned with the scientific definition of the baseline variables with a simple semantic structure. Each entry is
265 identified by a short name (composed of CMIP6 CMOR table and variable short name), title, a description, a standard name
266 and units, a format that has evolved since CMIP3 (WGCM Climate Simulation Panel, 2007). Syntax rules for list entries are
267 given in Table 2. The identifier will be considered as a registration identifier and is not expected to be used in CMIP7 era
268 products. New naming conventions are under discussion (K. Taylor, personal communication).
269 We have not been able to eliminate redundancy from the list: for instance, there is redundant information in variables on 8
270 atmosphere levels and the same variables on 19 levels. The evidence from CMIP6 usage statistics is that both variations are
271 very highly used.

Item	Syntax Rules
Variable Title	The title should match the rules set out in the "Style Guide for Variable Titles in CMIP6" (Juckes, 2018).
Variable Standard Name	The standard name must be included in the current CF conventions standard name list.
Variable Units	The units must have the same dimensions as canonical units of the standard name. I.e. the same combination of base dimensions as defined by the the International System of Units (International Bureau of Weights and Measures, 2019) ⁹ . The baseline variables list published here uses combinations of four dimensions: time, length, mass, and temperature.
Structure Title	The title should match the rules set out in the "Style Guide for Variable Titles in CMIP6" (Juckes, 2018).

41 9 The University Corporation for Atmospheric Research (UCAR) Udunits package
42 (<https://www.unidata.ucar.edu/software/udunits/>) can be used to check consistency of unit dimensionality.

272 Table 2: Syntax rules for items in the baseline variables list.

273 **3.2 Role from the modeller’s perspective**

274 The list of baseline variables will, in the first place, aid the model development process as a set of diagnostics for which
275 known good output is created by the model. For instance, this set can be used in regression tests when evaluating new model
276 versions, in order to detect significant changes in output.

277 The greater the overlap between what is output by the model and the baseline list, the bigger the contribution the model will
278 be able to make in intercomparison exercises, and the more widely the variables produced by the model will be used. Thus,
279 producing and publishing as many of the listed baseline variables as possible should be an aspiration in the development and
280 use of the model.

281 From the model developer’s perspective, transparency in the process of creating the baseline variables list is important,
282 because it clarifies the purpose of the list. The value of having a list and using it should be well-understood. It is not
283 expected that all models will be able to generate all variables, but the exclusion of specialised variables from the list will
284 ensure that most models can produce most variables. The process for maintaining and extending the list should be equally
285 transparent. If a modelling group is unable to provide a variable (especially one on the baseline list), they should be
286 encouraged to provide a reason using - for example - one of those listed in Table 3. The process for providing feedback
287 should be lightweight and transparent.

288

Reason	Description	Comments
Resource Constraints	The level of resources needed to implement the required workflow and handle the data streams is not available	Resource constraints may be due to staff availability or limits on the data volume or throughput that can be supported.
Workflow Error	An error occurred in the model i/o or post-processing stream and the variable was not generated correctly, and it is not expected to be corrected.	Errors could result from an undetected coding error or disruption of a workflow. Even if the coding error is easy to correct, repeating the workflow to generate the missing variable may be too costly.
Local Priorities	We need to choose between a range of different data requirements, and other variables, not in the baseline request, were found to have a higher priority for our users.	This is similar to Resource Constraints but draws attention to a different prioritisation rather than an overall resource limit.
Model Structure	The model does not support generation of the requested variable.	For example, an ocean ecosystem model can be phosphate-based rather than nitrogen-based (though the latter is more common). In that case the model may not be able to output nitrate concentrations in the ocean, but it is simply a matter of choice of the model developers and not a matter of model quality.

290 **3.3 Role from the infrastructure provider perspective**

291 Data infrastructure such as the Earth System Grid Federation (ESGF, <https://esgf.llnl.gov/>), the Climate and Forecast
292 Conventions (CF, <https://cfconventions.org/>) and the CMIPDR, along with secondary data portals, cloud platforms, and
293 collaborations (e.g. the Copernicus Climate Change Service C3S, <https://climate.copernicus.eu/> and PANGEO
294 <https://pangeo.io/>), together with underlying physical infrastructure, staff, and curation systems provided by host institutions,
295 disseminates climate datasets created by a variety of international modelling centres building on the data standards set by the
296 community. This standardisation ensures that user analysis can be performed across the multimodel ensemble, and facilitates
297 the scaling of data processing systems to provide and work with volumes at the magnitudes involved in CMIP. For
298 automated data processing options, standard compliance is essential (but see comments on incomplete compliance above).
299 For example, ESGF aims to enhance its compute services as part of its future architecture plans (Kershaw et al., 2020).
300 Secondary data evaluation or analysis packages such as ESMValtool (Weigel et al., 2021) and the Program for Climate
301 Model Diagnosis & Intercomparison (PCMDI) metrics package (Lee et al., 2022, 2024) also rely on these data standards.
302 The CMIP approach is founded on the Climate and Forecast (CF) metadata standards for NetCDF data. The CMIP project
303 has built upon these with the Data Reference Syntax (DRS; Taylor et al., 2018) defining file naming and data structure
304 conventions and the Controlled Vocabularies (Durack et al., 2024) defining the terms within these components.
305 A baseline variable list with common variable definitions will furthermore enable portals and indexers to support cross-
306 project data discovery and data analysis. The unique identification of the baseline variables and a consequent versioning and
307 maintenance of the list will ensure traceability of the variable usage in future. The I-ADOPT Framework ontology (Magagna
308 et al., 2021) provides a standard for this, which is implemented by the NERC vocabulary server providing e.g. the CF
309 Standard Names (<http://vocab.nerc.ac.uk/collection/P07/current/>).
310 The quality of data and metadata is of crucial importance. There are certain metadata which must have correct values for the
311 data to be ingestible by applications such as ESMValTool (Weigel et al., 2021).
312 Reliable and maintained software tools for creating standard compliant datasets are available for the modelling centres, but a
313 range of issues associated with implementation workflows has led to incomplete compliance in CMIP archives. A scan of
314 files from the CMIP archive (Petrie et al., 2024) revealed a wide range of technical errors. Some of these are related to
315 mistakes in the specification of cell methods, which might be obviated by improved documentation - particularly for those
316 cell methods which are used by the baseline variables. It should also be noted however, that most of the errors would have
317 little impact on the use of a majority of software applications. Experience shows that time-consuming or resource-intensive
318 data quality checks applied before data publication can reduce the amount of time and energy that has to be invested in
319 correcting issues and replacing datasets. The CMIP6 requirements specified compliance with the CF Conventions and
320 correct implementation of metadata specified in CVs and the data request (the latter can be verified with the PrePARE tool,
321 Nadeau et al., 2023). More detailed data checks such as the World Data Center for Climate quality control approach for

CMIP5 (Stockhouse et al., 2012) or for the C3S Climate Data Store (CDS, Buontempo et al., 2022) include range, outlier, and time axis checks alongside CF compliance.

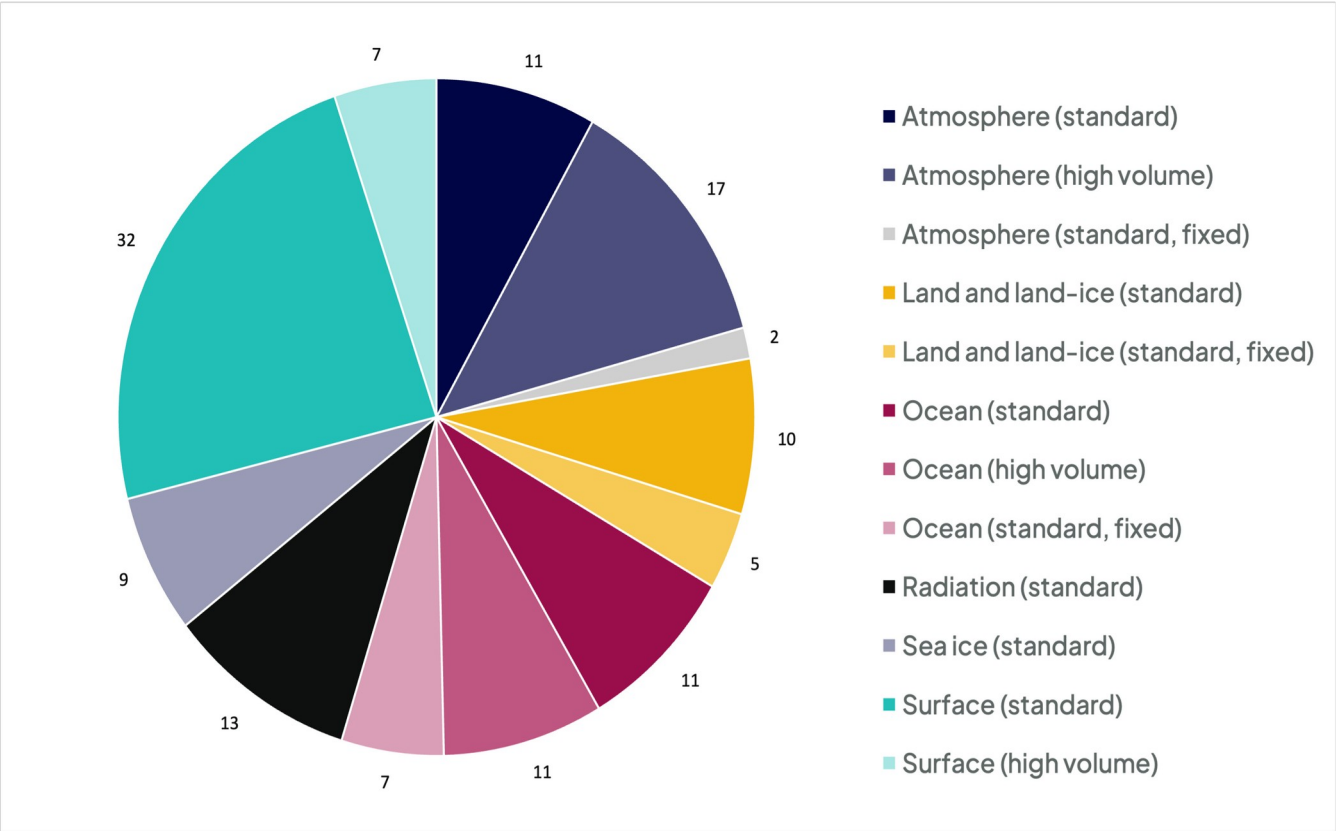
Underlying archive services which host ESGF and other climate data infrastructures will also benefit from greater consistency between different intercomparison projects. Stability of data specifications and data structures will allow archives to develop and maintain systems that exploit these structures with confidence that they will persist and be relevant for the duration of the data exploitation cycle.

3.4 Role from the data user's perspective

Data users of CMIP are a diverse community ranging from climate modellers, to scientists from a wide range of disciplines, to private sector product developers, and it is therefore hard to define who a "typical" user is. Historically climate scientists were representing the most important component of the user landscape. Their use of the data was to understand processes and study the future evolution of the climate and its potential impact on the natural system and human activities. There is no obvious boundary between climate impacts scientists and downstream exploitation of CMIP data for climate services (either public or private). CMIP data represents an important source of quantitative information for a large variety of actors and researchers operating well beyond the baseline-remit of the "climate science" community. These users include academics and industry working in areas which could possibly be called the climate adaptation and climate service community. A key need from this community, however, is the access to high-quality quantitative climate projection data, particularly focussing on ECVs (e.g., wind-speeds, insolation, precipitation, surface air temperature) mostly close to the surface. These correspond to a very small subset (~10) of the many variables CMIP makes available, but the existence of high-frequency and high-resolution climate data would enable much deeper integration of climate model output with "downstream impact models" (which often describe highly complex responses to a given set of meteorological time series input). An example of this lies in the energy systems research and applications (Craig et al., 2022; Dubus et al., 2022): the models used to inform electricity system planning typically operate on hourly-timesteps (as many of the fundamental design constraints relate to this timescale) and thus, to couple effectively, hourly gridded climate data (e.g., relating to wind resources at individual sites and timesteps) becomes essential. It would be extremely beneficial for the application community to ensure both widespread output of a small-but-comprehensive set of essential surface climate variables at the highest feasible sub-daily frequency along with a very strict observance of data and metadata standards for them. The contrast between this and previous CMIP archives would be considerable: in the current archives any analyst that wishes to look at more than a few essential surface climate variables must make a choice between having heterogeneous diagnostics, with different multi-model ensembles for each variable, or limiting the number of models involved, or making extreme compromises on the data frequency provided (e.g., daily rather than sub-daily). Neither of these is ideal. By establishing a clear and realistic baseline we hope to ensure that there is a greater level of consistency in the data collections, allowing more robust multivariable analysis and enabling

354 much stronger linking of raw output from climate models with “downstream impact models”, facilitating the translation of
355 climate risk into meaningful and applicable information for end-users and wider society.
356 The ambition is to achieve 90% of models providing 90% of the low volume and configuration baseline variables across
357 major intercomparison programmes such as CMIP7. In fact, in the last CMIP6 exercise only 29% of models provided 29%
358 of priority 1 variables, 90% of models provided 8% or more, and only 2 models¹⁰ provided 50% or more.

359



360

361 **Figure 5: ESM baseline climate variable categories. Distribution of ESM-BCVs across a range of categories (using the data listed**
362 **in Table A7). A variable is considered as “high-volume” (dark shading) if 10 thousand years of simulation generates more than**
363 **1,500 GB of data from a 1-degree resolution model with 60 atmospheric levels and 500 oceanic levels archived (assuming single**
364 **precision data storage without compression).**

51 10 These figures are for August 2023. Figures taken in March 2022 were very similar: 28% of models providing 28% or
52 more of priority 1 variables, 90% or models providing 7.8% or more. 2 models providing more than 50%.

365
366
367
368



Figure 6 Example data volumes expressed as GB per 10,000 years of simulation for a notional 1-degree resolution model with 60 atmospheric levels and 50 oceanic levels (see Table A8 for details). Each rectangle area (both visible and obscured) represents the nominal volume for a specific output category. Single precision data storage without compression are here considered.

369 **4 Results**

370 The ESM-BCV list, after shortlisting and revision, contains 132 variables listed in Appendix 1 (tables A1, A2 and A3). In
371 the final list there are 121 time series and 14 fixed fields. Of the time-varying fields, 35 are classified as high volume (see
372 Figure 6 and Table A8 for the illustrative data volumes that determined the categorization). The high-volume category
373 includes sub-daily data, daily data on 19 pressure levels (see Appendix 5 for details) and monthly data on ocean model
374 levels. The remaining 86 lower volume time-varying fields and the 14 fixed field variables should be considered top priority
375 for most WCRP MIP climate simulations, although it is recognised that in the short term, at least, it may not be possible to
376 provide 100% of them in all cases. More details are given in Figure 5 and Table A7.

377 The shortlisting was based on four criteria: limiting consideration to CMIP6 priority 1 variables, number of files
378 downloaded, volume of data downloaded, and the number of models for which a variable was provided.

379 Although all four criteria were formally included in the shortlisting process, they had differing impacts:

380

- 381 1. Limiting consideration to CMIP6 priority 1 variables prevented only one variable from making the short
382 list (monthly Temperature of Soil, Lmon.tsl).
- 383 2. The criteria based on the number of files downloaded added one variable which would have otherwise not
384 been included (daily Total Cloud Cover Percentage, day.clt).

3. The short list of low-frequency variables (monthly and lower frequency) would have been unaffected had we only considered the number of contributing models.

4. If download volumes were used as the only criterion, the resulting list of higher-frequency variables (daily and higher) would have been the same as considering all four criteria (apart from “day.clt”).

Thus, for fixed and monthly-mean fields, the shortlist was largely based on the model participation statistic, and for the high frequency fields it was based on volume of data downloaded. This process resulted in a shortlist of 147 variables.

During the subsequent community consultation, 27 variables were removed from the shortlist (see Table A5) and 15 were added (see Table A6), resulting in the 135 variables listed in Appendix 1.

We can support the reasonableness of the ESM-BCV list by pointing out that it is not dissimilar to past lists of CMIP high priority standard output. Modelling groups participating in MIPs have been producing many of these variables for over two decades. It is informative to compare the ESM-BCV list with the 118 high priority variables specified for CMIP3 (WGCM Climate Simulation Panel, 2007). Some variables in the CMIP3 list were dropped prior to CMIP6 because they were designed to monitor model limitations which are no longer relevant (e.g., imposed ocean “flux corrections” are no longer needed). Eliminating such variables, we find 80% of the variables remaining are also included in the BCV list. This indicates that although list development followed different procedures in the past, nevertheless, there is a high degree of continuity in the perceived value of these variables.

The process of consultation, in defining the short list as well as agreeing subsequent revisions, has helped to spread awareness of the scope and impact of the CMIP variable metadata and has driven new engagement in the process. There was strong support for the utility of the list (80% of survey respondents rating usefulness four or five out of five). There was also support for the process, but with caveats raised about the possible bias towards past requirements rather than future needs (O'Rourke et al. 2023). The author discussions leading to finalisation of the list went beyond evaluation of the community consultation. The name of the list, which started as "Baseline Climate Variables" changed twice. Firstly to "Baseline Climate Variables for Earth System Models" in order to avoid any appearance of detracting from the Global Climate Observing System (GCOS) Essential Climate Variable work (World Meteorological Organisation, 2022c) by clearly emphasising the focus here on model data and finally to Baseline Climate Variables for Earth System Modelling to avoid the potentially restrictive interpretation of Earth System Models only being those with a comprehensive range interactions between the biosphere and physical climate components.

4.1 Provenance

The CMIP6 variables derive from many sources. Many variables were inherited from CMIP5 Standard Output (PCMDI, 2013). Revisions and extensions in CMIP6 came from Griffies et al. (2016) for Omon variables, van den Hurk (2016) for land surface variables, Notz et al. (2016) for sea-ice variables, Gerber and Manzini (2016) for daily atmospheric fields, Haarsma et al. (2016), Ruane et al. (2016) for 6hrPlev.hurs, Jones et al. (2016) for carbon cycle and terrestrial biosphere.

417 4.2 Limitations, extensions, and revision

418 As noted in the introduction, the ESM-BCV list is deliberately limited in scope so that it can be implemented across a wide
419 range of modelling activities without incurring unreasonable cost. The need for additional variables in many important use-
420 cases was discussed, such as the need for more variables to close the carbon budget, accurately reflect ocean heat content, or
421 monitor ocean currents. Coverage of these use cases is deliberately omitted here and is being picked up in the CMIP AR7
422 Fast Track Data Request¹¹. The latter request contains, in the version 1.0 release, over 1800 variables which are associated
423 with specified scientific or climate impacts use cases.

424 It has also been noted that the use of model levels for ocean variables results in high volume datasets which can be difficult
425 for some users to exploit because of the complexity of the vertical coordinates used in models. Discussions about a potential
426 shift to an agreed set of layers in a standard coordinate system are taking place within the ocean theme of the CMIP AR7
427 Fast Track Data Request. Some overlapping spatial dimensions (P8 and P19 atmospheric levels) and temporal frequencies
428 (3hr and 1hr) are still part of the list to enable a certain flexibility, e.g. for MIPs proposal, to either retain or avoid such
429 redundancies in the compilation of a specific experiment request.

430 Revision of the list, that is to say changing the variables included in the baseline rather than constructing a larger list which
431 builds on the baseline as is being done for the CMIP AR7 Fast Track Data Request, has also been discussed. It is clear that
432 revision will be needed to accommodate changes in scientific focus, but this need for revision needs to be balanced against a
433 need for stability associated with the central aim of enhancing interoperability between distinct activities and distinct phases
434 of CMIP.

435 5 Conclusion

436 The set of 132 ESM-BCVs presented here provides a reference collection of variables for MIPs which will facilitate greater
437 consistency in data requests. By identifying variables which have high utility in many applications, the ESM-BCV list will
438 also enable modelling centres to develop, standardise, and rationalise workflows.

439 The baseline list presents a standardised set which should be within reach of any modelling centre aspiring to generate data
440 for community evaluation and exploitation. There will always be circumstances in which variables need to be omitted,
441 especially the high-volume subset of 35 variables, but we expect this baseline set to lead to enhanced consistency in the
442 expanding WCRP climate projections archive.

443 The ESM-BCV list should be considered a snapshot of variables currently considered by modelling groups and users to be of
444 general high value. Its similarities with earlier CMIP lists of high priority variables attests to its likely continued relevance
445 long into the future, but the expectation is that reassessment of community priorities will result in modifications to the list.

59 11The process for extending the list was launched by a CMIP Panel decision “G1 [Gateway 1] DR Strategic Approach”
60 (<https://airtable.com/shrIAHOuVw8ktdoe1>, items 9 and 10, approved July 24, 2023) and announced in Dec. 2023
61 (<https://wcrp-cmip.org/cmip7-data-request-harmonised-thematic-variables/>) [both links accessed 26th June, 2034]).

446 The baseline variable list has grown out of the CMIP6DR (Juckes et al., 2020) and is dominated by variables already present
447 in earlier requests (PCMDI, 2013). It has been shaped by feedback about the problems caused for users by inconsistencies in
448 the CMIP6 archive and for providers by late finalisation of the request (see Figure 3 for more details). The baseline variable
449 list will reduce the workload for data providers, service providers and users by providing a reusable and reliable basic set of
450 variables. For users in the climate services and other communities outside the research community, the baseline variables
451 will promote greater consistency and transparency in the derived products used by these communities, which typically
452 depend on multiple variables and multiple climate models.

453 Although the baseline set includes little over 7% of the variables found in the CMIP6 Data Request, the consultation process
454 revealed that most climate-service users tend to use an even smaller subset of the variables. A more detailed analysis of the
455 needs of the user and stakeholder landscape is required and may call for a further differentiation of the baseline variables
456 portfolio.

457 There has been considerable interest in the creation and sharing of standard indices of climate variability (e.g. Klein Tank et
458 al., 2009). The level of standardisation of definitions of these indices is not sufficiently advanced to support reliable direct
459 collection through the data request. The underlying challenge providing a central reference for these indices will, however,
460 be picked up by the CMIP7 Rapid Evaluation Framework (REF) project¹².

461 The ESM-BCV list is intended to address issues associated with the rapid expansion and relatively weak prioritisation of the
462 CMIP6 data request (around 50% of variables were classified as top priority, more than most models provided). The list
463 provides a starting point for any model workflows which are intended to support community multi-ESM ensembles.

464 The list falls well short of the scope needed to support scientific analysis or detailed climate impacts assessment. In either of
465 those cases additional variables will need to be defined for MVEs which target specific science goals or climate impacts
466 work. For instance, work on dynamical processes in the atmosphere will require high resolution models and specialised
467 atmospheric variables to capture details of those processes. Work on the terrestrial biosphere will typically use lower
468 resolution models and a broad range of land surface variables. Work on climate impacts will require a range of surface and
469 near-surface variables archived at sufficient frequency to support analysis of impact on social, economic, and biological
470 systems. For CMIP7, the baseline list will form the core of the data request and be complemented by a set of topic themed
471 papers to be developed through a process which is based on that described here for the baseline variables¹³.

472 We have not made a detailed comparison of the EM-BCVs and the GCOS Essential Climate Variables in this paper. This
473 work is part of a wider set of changes to the way that climate model data standards are supported within WCRP. Care needs
474 to be taken when comparing model output with observations: for instance the cloud cover variables used to compare models
475 with each other are not directly comparable with observations.

64 12 The figure for the number of ocean levels here is based on what was submitted to the CMIP6 archive. Some modelling
65 centres submitted data at lower resolution than the full model grid.

66 13 Launched November 2024: <https://wcrp-cmip.org/event/ref-project-launch/> [accessed 2024/12/03].

476 **Appendix 1: The baseline climate variables**

477 There are 135 ESM-BCVs. Of these, 35 are flagged as “high volume”and 14 are fixed model configuration fields. They are
478 listed in Tables A1,2,3 below.

CMIP6 Identifier	Realm (frequency)	Title	Units	CF Standard Name
Model Configuration Field				
fx.rootd	Land (f)	Maximum Root Depth	m	root_depth
fx.orog	Grid (f)	Surface Altitude	m	surface_altitude
fx.sftgif	Grid (f)	Land Ice Area Percentage	%	land_ice_area_fraction
fx.sftlf	Grid (f)	Percentage of the Grid Cell Occupied by Land (Including Lakes)	%	land_area_fraction
Model Configuration Surface Field, Area Mean, Masked (Land)				
fx.mrsofc	Land (f)	Capacity of Soil to Store Water (Field Capacity)	kg m-2	soil_moisture_content_at_field_capacity
Model Configuration Surface Field, Area Mean, Ocean Grid				
Ofx.sftof	Ocean (f)	Sea Area Percentage	%	sea_area_fraction
Model Configuration Field on Sea Floor, Area Mean, Ocean Grid				
Ofx.deptho	Ocean (f)	Sea Floor Depth Below Geoid	m	sea_floor_depth_below_geoid
Ofx.hfgeou	Ocean (f)	Upward Geothermal Heat Flux at Sea Floor	W m-2	upward_geothermal_heat_flux_at_sea_floor
Model Configuration Field, Area Sum (No Height Dependence)				
fx.areacella	Grid (f)	Grid-Cell Area for Atmospheric Grid Variables	m2	cell_area
Model Configuration Field, Area Sum, Ocean Grid (No Height Dependence)				
Ofx.areacello	Ocean (f)	Grid-Cell Area for Ocean Variables	m2	cell_area
Model Configuration Integer Field of Flag Values, Ocean Grid				
Ofx.basin	Ocean (f)	Region Selection Index	1	region
Model Configuration Field on Ocean-Model Levels and Grid, Area Mean				
Ofx.masscello	Ocean (f)	Ocean Grid-Cell Mass per Area	kg m-2	sea_water_mass_per_unit_area
Ofx.thkcello	Ocean (f)	Ocean Model Cell Thickness	m	cell_thickness
Model Configuration Field on Soil-Model Levels, Masked (Land)				
Efx.slthick	Land (f)	Thickness of Soil Layers	m	cell_thickness

479 **Table A1: ESM Baseline Climate Variables: 14 fixed model configuration fields.** These ESM-BCVs are listed under 10 different
480 structures. For masked fields, the nature of the unmasked points is indicated in brackets, e.g. "Masked (Land)" implies that
481 only land points are included. All are global fields. The "Model Configuration" fields have no temporal dimension. Area
482 means and sums are taken over grid cells, time means are taken over the sampling period, e.g. a day or a calendar month.
483 The frequency in column 2 indicates the frequency of stored data points, which may be time means or instantaneous values.
484 The abbreviation is `f` for fixed.

CMIP6 Identifier	Realm (frequency)	Title	Units	CF Standard Name
Temporal Maximum, Near-Surface Field (2m)				
day.tasmax	Surface (d)	Daily Maximum Near-Surface Air Temperature	K	air_temperature
Temporal Minimum, Near-Surface Field (2m)				
day.tasmin	Surface (d)	Daily Minimum Near-Surface Air Temperature	K	air_temperature
Time-Mean on 19 Pressure Levels				
Amon.hur	Atmosphere (m)	Relative Humidity	%	relative_humidity
Amon.hus	Atmosphere (m)	Specific Humidity	1	specific_humidity
Amon.ta	Atmosphere (m)	Air Temperature	K	air_temperature
Amon.ua	Atmosphere (m)	Eastward Wind	m s-1	eastward_wind
Amon.va	Atmosphere (m)	Northward Wind	m s-1	northward_wind
Amon.wap	Atmosphere (m)	Omega (=dp/dt)	Pa s-1	lagrangian_tendency_of_air_pressure
Amon.zg	Atmosphere (m)	Geopotential Height	m	geopotential_height
Time and Area Mean on Single Level				
Amon.prw	Atmosphere (m)	Water Vapor Path	kg m-2	atmosphere_mass_content_of_water_vapor
Amon.clivi	Atmosphere (m)	Ice Water Path	kg m-2	atmosphere_mass_content_of_cloud_ice
Amon.clt	Atmosphere (m)	Total Cloud Cover Percentage	%	cloud_area_fraction
Amon.clwvi	Atmosphere (m)	Condensed Water Path	kg m-2	atmosphere_mass_content_of_cloud_condensed_water
Amon.hfss	Surface (d)	Surface Upward Sensible Heat Flux	W m-2	surface_upward_sensible_heat_flux
Amon.rlds	Radiation (m)	Surface Downwelling Longwave Radiation	W m-2	surface_downwelling_longwave_flux_in_air
Amon.rldscs	Radiation (m)	Surface Downwelling Clear-Sky Longwave Radiation	W m-2	surface_downwelling_longwave_flux_in_air_assuming_clear_sky
Amon.rlus	Radiation (m)	Surface Upwelling Longwave Radiation	W m-2	surface_upwelling_longwave_flux_in_air
Amon.rluscs	Radiation (m)	Surface Upwelling Clear-Sky Longwave Radiation	W m-2	surface_upwelling_longwave_flux_in_air_assuming_clear_sky
Amon.rlut	Radiation (m)	TOA Outgoing Longwave Radiation	W m-2	toa_outgoing_longwave_flux
Amon.rlutcs	Radiation (m)	TOA Outgoing Clear-Sky Longwave Radiation	W m-2	toa_outgoing_longwave_flux_assuming_clear_sky
Amon.rsds	Radiation (m)	Surface Downwelling Shortwave Radiation	W m-2	surface_downwelling_shortwave_flux_in_air
Amon.rsds	Radiation (m)	Surface Downwelling Clear-Sky Shortwave Radiation	W m-2	surface_downwelling_shortwave_flux_in_air_assuming_clear_sky
Amon.rsdt	Radiation (m)	TOA Incident Shortwave Radiation	W m-2	toa_incoming_shortwave_flux
Amon.rsus	Radiation (m)	Surface Upwelling	W m-2	surface_upwelling_shortwave_flux_in_air

		Shortwave Radiation		
Amon.rsuscs	Radiation (m)	Surface Upwelling Clear-Sky Shortwave Radiation	W m-2	surface_upwelling_shortwave_flux_in_air_assuming_clear_sky
Amon.rsut	Radiation (m)	TOA Outgoing Shortwave Radiation	W m-2	toa_outgoing_shortwave_flux
Amon.rsutcs	Radiation (m)	TOA Outgoing Clear-Sky Shortwave Radiation	W m-2	toa_outgoing_shortwave_flux_assuming_clear_sky
day.clt	Atmosphere (d)	Total Cloud Cover Percentage	%	cloud_area_fraction
day.rsds	Radiation (d)	Surface Downwelling Shortwave Radiation	W m-2	surface_downwelling_shortwave_flux_in_air
Amon.pr	Surface (m)	Precipitation	kg m-2 s-1	precipitation_flux
Amon.evspsb_l	Surface (m)	Evaporation Including Sublimation and Transpiration	kg m-2 s-1	water_evapotranspiration_flux
Amon.hfls	Surface (m)	Surface Upward Latent Heat Flux	W m-2	surface_upward_latent_heat_flux
Amon.prc	Surface (m)	Convective Precipitation	kg m-2 s-1	convective_precipitation_flux
Amon.prsn	Surface (m)	Snowfall Flux	kg m-2 s-1	snowfall_flux
Amon.ps	Surface (m)	Surface Air Pressure	Pa	surface_air_pressure
Amon.psl	Surface (m)	Sea Level Pressure	Pa	air_pressure_at_mean_sea_level
Amon.tauu	Surface (m)	Surface Downward Eastward Wind Stress	Pa	surface_downward_eastward_stress
Amon.tauv	Surface (m)	Surface Downward Northward Wind Stress	Pa	surface_downward_northward_stress
Amon.ts	Surface (m)	Surface Temperature	K	surface_temperature
day.pr	Surface (m)	Precipitation	kg m-2 s-1	precipitation_flux
CFday.ps	Surface (d)	Surface Air Pressure	Pa	surface_air_pressure
day.psl	Surface (d)	Sea Level Pressure	Pa	air_pressure_at_mean_sea_level
LImon.snc	Land (m)	Snow Area Percentage	%	surface_snow_area_fraction
Time and Area Mean on Single Level, Masked (Land)				
LImon.snw	Land (m)	Surface Snow Amount	kg m-2	surface_snow_amount
Lmon.evspsb_lsoi	Land (m)	Water Evaporation from Soil	kg m-2 s-1	water_evaporation_flux_from_soil
Lmon.evspsb_lveg	Land (m)	Evaporation from Canopy	kg m-2 s-1	water_evaporation_flux_from_canopy
Lmon.lai	Land (m)	Leaf Area Index	1	leaf_area_index
Lmon.mrfso	Land (m)	Soil Frozen Water Content	kg m-2	soil_frozen_water_content
Lmon.mrro	Land (m)	Total Runoff	kg m-2 s-1	runoff_flux
Lmon.mrros	Land (m)	Surface Runoff	kg m-2 s-1	surface_runoff_flux
Lmon.mrso	Land (m)	Total Soil Moisture Content	kg m-2	mass_content_of_water_in_soil

Weighted Time-Mean on Single Level, Ocean Grid, Masked (Sea Ice)				
SImon.sithick	Seaice (m)	Sea Ice Thickness	m	sea_ice_thickness
SImon.sitemptop	Seaice (m)	Surface Temperature of Sea Ice	K	sea_ice_surface_temperature
Time and Area Mean on Single Level, Ocean Grid				
Oday.sos	Ocean (d)	Sea Surface Salinity	0.001	sea_surface_salinity
Oday.tos	Ocean (d)	Sea Surface Temperature	degC	sea_surface_temperature
Oday.zos	Ocean (d)	Sea Surface Height Above Geoid	m	sea_surface_height_above_geoid
Omon.hfds	Ocean (m)	Downward Heat Flux at Sea Water Surface	W m-2	surface_downward_heat_flux_in_sea_water
Omon.mlotst	Ocean (m)	Ocean Mixed Layer Thickness Defined by Sigma T of 0.03 kg m-3	m	ocean_mixed_layer_thickness_defined_by_sigma_t
Omon.sos	Ocean (m)	Sea Surface Salinity	0.001	sea_surface_salinity
Omon.tos	Ocean (m)	Sea Surface Temperature	degC	sea_surface_temperature
Omon.zos	Ocean (m)	Sea Surface Height Above Geoid	m	sea_surface_height_above_geoid
SImon.simass	Seaice (m)	Sea-Ice Mass per Area	kg m-2	sea_ice_amount
SImon.sitimefrac	Seaice (m)	Fraction of Time Steps with Sea Ice	1	fraction_of_time_with_sea_ice_area_fraction_above_threshold
Weighted Time-Mean on Single Level, Masked (Snow on Sea Ice)				
SImon.sisnthick	Seaice (m)	Snow Thickness	m	surface_snow_thickness
Time-Mean Surface Field, Ocean Grid (Area Mean or Vertices)				
Omon.tauuo	Ocean (m)	Sea Water Surface Downward X Stress	N m-2	downward_x_stress_at_sea_water_surface
Omon.tauvo	Ocean (m)	Sea Water Surface Downward Y Stress	N m-2	downward_y_stress_at_sea_water_surface
Time-Mean Near-Surface Field (10m)				
Amon.uas	Surface (m)	Eastward Near-Surface Wind	m s-1	eastward_wind
Amon.vas	Surface (m)	Northward Near-Surface Wind	m s-1	northward_wind
day.sfcWind	Surface (d)	Near-Surface Wind Speed	m s-1	wind_speed
day.uas	Surface (d)	Eastward Near-Surface Wind	m s-1	eastward_wind
day.vas	Surface (d)	Northward Near-Surface Wind	m s-1	northward_wind
Amon.sfcWind	Surface (m)	Near-Surface Wind Speed	m s-1	wind_speed

Monthly-Mean Daily Maximum, Near-Surface Field (2m)				
Amon.tasma x	Surface (m)	Daily Maximum Near-Surface Air Temperature	K	air_temperature
Monthly-Mean Daily Minimum, Near-Surface Field (2m)				
Amon.tasmin	Surface (m)	Daily Minimum Near-Surface Air Temperature	K	air_temperature
Time-Mean Near-Surface Field (2m)				
Amon.hurs	Surface (m)	Near-Surface Relative Humidity	%	relative_humidity
Amon.huss	Surface (m)	Near-Surface Specific Humidity	1	specific_humidity
day.hurs	Surface (d)	Near-Surface Relative Humidity	%	relative_humidity
day.huss	Surface (d)	Near-Surface Specific Humidity	1	specific_humidity
day.tas	Surface (d)	Near-Surface Air Temperature	K	air_temperature
Amon.tas	Surface (d)	Near-Surface Air Temperature	K	air_temperature
Time-Mean on Single Soil-Model Level, Masked (Land)				
Lmon.mrsos	Land (m)	Moisture in Upper Portion of Soil Column	kg m-2	mass_content_of_water_in_soil_layer
Time-Mean on Single Level, Ocean Grid				
SImon.siconc	Seaice (m)	Sea-Ice Area Percentage (Ocean Grid)	%	sea_ice_area_fraction
SIday.siconc	Seaice (d)	Sea-Ice Area Percentage (Ocean Grid)	%	sea_ice_area_fraction
Time-Mean Weighted by Sea-Ice Area, Single Level, on Ocean Grid Vertices, Masked (Sea Ice)				
SImon.siu	Seaice (m)	X-Component of Sea-Ice Velocity	m s-1	sea_ice_x_velocity
SImon.siv	Seaice (m)	Y-Component of Sea-Ice Velocity	m s-1	sea_ice_y_velocity
Time and Global Mean on a Single Level				
Omon.zostoga	Ocean (m)	Global Average Thermosteric Sea Level Change	m	global_average_thermosteric_sea_level_change

Table A2: ESM Baseline Climate Variables: 86 low volume variables. These ESM-BCVs are listed under 17 different structures. For masked fields, the nature of the unmasked points is indicated in brackets, e.g. "Masked (Land)" implies that only land points are included. All are global fields. Area means and sums are taken over grid cells, time means are taken over the sampling period, e.g. a day or a calendar month. The frequency in column 2 indicates the frequency of stored data points, which may be time means or instantaneous values. The abbreviations are: `m' for monthly, `d' for daily.

CMIP6 Identifier	Realm	Title	Units	CF standard Name
------------------	-------	-------	-------	------------------

Synoptic Field on Three Pressure Levels, Cell Mean				
6hrPlevPt.ta	Atmosphere (6)	Air Temperature	K	air_temperature
6hrPlevPt.ua	Atmosphere (6)	Eastward Wind	m s-1	eastward_wind
6hrPlevPt.va	Atmosphere (6)	Northward Wind	m s-1	northward_wind
Synoptic Near-Surface Field (10m)				
3hr.uas	Surface (3)	Eastward Near-Surface Wind	m s-1	eastward_wind
3hr.vas	Surface (3)	Northward Near-Surface Wind	m s-1	northward_wind
Synoptic Near-Surface Field (2m)				
3hr.huss	Surface (3)	Near-Surface Specific Humidity	1	specific_humidity
3hr.tas	Surface (3)	Near-Surface Air Temperature	K	air_temperature
Time-Mean on 19 Pressure Levels				
Eday.hus	Atmosphere (d)	Specific Humidity	1	specific_humidity
Eday.ua	Atmosphere (d)	Eastward Wind	m s-1	eastward_wind
Eday.va	Atmosphere (d)	Northward Wind	m s-1	northward_wind
Eday.ta	Atmosphere (d)	Air Temperature	K	air_temperature
Eday.zg	Atmosphere (d)	Geopotential Height	m	geopotential_height
Time-Mean on 8 Pressure Levels				
day.hur	Atmosphere (d)	Relative Humidity	%	relative_humidity
day.hus	Atmosphere (d)	Specific Humidity	1	specific_humidity
day.ta	Atmosphere (d)	Air Temperature	K	air_temperature
day.ua	Atmosphere (d)	Eastward Wind	m s-1	eastward_wind
day.va	Atmosphere (d)	Northward Wind	m s-1	northward_wind
day.wap	Atmosphere (d)	Omega (=dp/dt)	Pa s-1	lagrangian_tendency_of_air_pressure
Time and Area Mean on Single Level				
3hr.pr	Surface (3)	Precipitation	kg m-2 s-1	precipitation_flux
E1hr.pr	Surface (1)	Precipitation	kg m-2 s-1	precipitation_flux
Time-Mean Near-Surface Field (2m)				
6hrPlev.hurs	Surface (6)	Near-Surface Relative Humidity	%	relative_humidity
Time-Mean on Atmosphere-Model Levels				
Amon.cl	Atmosphere (m)	Percentage Cloud Cover	%	cloud_area_fraction_in_atmosphere_layer
Amon.cli	Atmosphere (m)	Mass Fraction of Cloud Ice	kg kg-1	mass_fraction_of_cloud_ice_in_air
Amon.clw	Atmosphere (m)	Mass Fraction of Cloud Liquid Water	kg kg-1	mass_fraction_of_cloud_liquid_water_in_air
Time-Mean, Area Sum, Field on Ocean-Model Levels				

Omon.wmo	Ocean (m)	Upward Ocean Mass Transport	kg s-1	upward_ocean_mass_transport
Time-Mean Field on Ocean-Model Levels				
Omon.thkcello	Ocean (m)	Ocean Model Cell Thickness	m	cell_thickness
Omon.masscello	Ocean (m)			
Omon.so	Ocean (m)	Sea Water Salinity	0.001	sea_water_salinity
Omon.thetao	Ocean (m)	Sea Water Potential Temperature	degC	sea_water_potential_temperature
Omon.bighetao	Ocean (m)	Sea Water Conservative Temperature	degC	sea_water_conservative_temperature
Omon.umo	Ocean (m)	Ocean Mass X Transport	kg s-1	ocean_mass_x_transport
Omon.uo	Ocean (m)	Sea Water X Velocity	m s-1	sea_water_x_velocity
Omon.vmo	Ocean (m)	Ocean Mass Y Transport	kg s-1	ocean_mass_y_transport
Omon.vo	Ocean (m)	Sea Water Y Velocity	m s-1	sea_water_y_velocity
Omon.wo	Ocean (m)	Sea Water Vertical Velocity	m s-1	upward_sea_water_velocity

Table A3: Baseline Climate Variables, high volume list of 35 variables. Abbreviations for frequency are as for Table A1, extended to include The abbreviations are: `3' and `6' for 3- and 6-hourly respectively.

Variable	Condition
Ofx.thkcello	To be provided if ocean grid cells have a fixed thickness
Omon.thkcello	To be provided if ocean grid cells have a time varying thickness
Omon.masscello	To be provided if ocean grid cells have a time varying mass
Omon.bighetao	Contributed only for models using conservative temperature as the prognostic field. Note that in this case bighetao is needed to enable analysis of model processes but Omon.thetao still needs to be provided in order to enable comparison with observations.
Eday.ua, va, ta, hus day.ua, va, ta, hus	The "day" versions of these fields provide data on 8 pressure levels which are a subset of the 19 levels used in the "Eday" versions, so, in general, only one version should be archived. Both options are included so that modelling centres can provide the greater detail afforded by 19 levels when resources permit.

Table A4: Variables which are only provided under specific conditions.

Appendix 2: Variables removed from and added to the shortlist

CMIP6 Identifier	Realm	Title	Units	CF standard Name	R
6hrLev.ta	Atmosphere (6)	Air Temperature	K	air_temperature	S
6hrLev.ua	Atmosphere (6)	Eastward Wind	m s-1	eastward_wind	S
6hrLev.va	Atmosphere (6)	Northward Wind	m s-1	northward_wind	S
6hrLev.hus	Atmosphere (6)	Specific Humidity	1	specific_humidity	S
CFday.hur	Atmosphere (d)	Relative Humidity	%	relative_humidity	D
CFday.hus	Atmosphere (d)	Specific Humidity	1	specific_humidity	D

CFday.ta	Atmosphere (d)	Air Temperature	K	air_temperature	D
CFday.ua	Atmosphere (d)	Eastward Wind	m s-1	eastward_wind	D
CFday.va	Atmosphere (d)	Northward Wind	m s-1	northward_wind	D
CFday.zg	Atmosphere (d)	Geopotential Height	m	geopotential_height	D
CFday.wa p	Atmosphere (d)	Omega (=dp/dt)	Pa s-1	lagrangian_tendency_of_air_pressure	D
CFday.cl	Atmosphere (d)	Percentage Cloud Cover	%	cloud_area_fraction_in_atmosphere_layer	D
LImon.snd	Land (m)	Snow Depth	m	surface_snow_thickness	E
Efx.rld	Radiation (f)	Downwelling Longwave Radiation	W m-2	downwelling_longwave_flux_in_air	O
Efx.rlu	Radiation (f)	Upwelling Longwave Radiation	W m-2	upwelling_longwave_flux_in_air	O
Efx.rsu	Radiation (f)	Upwelling Shortwave Radiation	W m-2	upwelling_shortwave_flux_in_air	O
Efx.rsd	Radiation (f)	Downwelling Shortwave Radiation	W m-2	downwelling_shortwave_flux_in_air	O
Efx.fldcap acity	Land (f)	Field Capacity	%	volume_fraction_of_condensed_water_in_soil_at_field_capacity	O
Efx.siltfrac	Land (f)	Silt Fraction	1	volume_fraction_of_silt_in_soil	O
Efx.wilt	Land (f)	Wilting Point	%	volume_fraction_of_condensed_water_in_soil_at_wilting_point	O
fx.areacellr	Grid (f)	Grid-Cell Area for River Model Variables	m2	cell_area	O
fx.zfull	Grid (f)	Altitude of Model Full-Levels	m	height_above_reference_ellipsoid	O
day.rlut	Radiation (d)	TOA Outgoing Longwave Radiation	W m-2	toa_outgoing_longwave_flux	X
day.rlds	Radiation (d)	Surface Downwelling Longwave Radiation	W m-2	surface_downwelling_longwave_flux_in_air	X
E3hr.sfcW ind	Surface (3)	Near-Surface Wind Speed	m s-1	wind_speed	X
Oyr.o2	Ocean (y)	Dissolved Oxygen Concentration	mol m-3	mole_concentration_of_dissolved_molecular_oxygen_in_sea_water	X
Ofx.volcell o	Ocean (f)	Ocean Grid-Cell Volume	m3	ocean_volume	D

Table A5: Variables which were included in the shortlist and removed in the revision process. Reasons: S: specialist variables of use for a limited range of applications; D: duplicate or near duplicate of another variable in the list; E: included in the shortlist as a result of a clerical error, these variables do not meet shortlisting criteria; O: These variables were included following an initial decision to include all fixed variables, but as they have extremely low usage, only being published for 12 or fewer models, they were subsequently removed.; X: Low usage in corrected download statistics.

Variable	Reason
Oday.sos	This variable was considered to be of high importance for characterising the ocean state.
Oday.tos	This variable was considered to be of high importance for characterising the ocean state.
CFday.ps	This variable is important for models which have vertical coordinates which are defined in terms of surface pressure, such as the sigma coordinate. Where needed, it should be included as an auxiliary variable, not as an independently requested

	variable.
Ofx.thkcello	To ensure full information about the ocean model grid
Omon.thkcello	To ensure full information about the ocean model grid
Omon.masscello	To ensure full information about the ocean model grid
Omon.bigthetao	This variable is of fundamental importance for those models that use conservative potential temperature as a prognostic variable, but appeared low in the shortlisting because this was a minority of models on CMIP6.
SIday.siconc	To provide basic information about sea-ice cover
LImon.snc	High usage in corrected download statistics (see Section 2.4)
Omon.zostoga	High usage in corrected download statistics (see Section 2.4)
Lmon.evpsblveg	High usage in corrected download statistics (see Section 2.4)
SImon.sitimefrac	High usage in corrected download statistics (see Section 2.4)
Eday.ta	Added in response to review comments
Oday.zos	Added in response to review comments
Amon.rluscs	Added in response to review comments

Table A6: Variables added in the review process

Appendix 3: Summary Tables

	Standard	High Volume	Fixed
Atmosphere	11	17	2
Land and Landice	10		5
Ocean	11	11	7
Radiation	13		
Sea ice	9		
Surface	32	7	
Total	86	35	14

Table A7. The counts of baseline climate variables in different categories. For explanation of the “high volume” category, see table A8.

	Examples	GB/10k years of simulation
Monthly single level	Amon.tas, Omon.tos	62
Monthly, 19 levels	Amon.ua	1182
Daily single level	day.tas	1893
Monthly, atmosphere levels	Amon.clw	3733
Monthly, ocean levels	Omon.uo	3110

Table A8: Example data volumes. Based on a 1-degree resolution model with 60 atmospheric levels and 50 oceanic levels¹⁴. Single precision data storage without compression. A variable is considered as “high-volume” (dark shading) if 10 thousand years of simulation generates more than 1,500 GB of data.

¹⁴ <https://wcrp-cmip.org/cmip7-data-request-harmonised-thematic-variables/> [accessed 2024/12/04]

508 **Appendix 4: Invitation to Participate**

509 Invitation to participate in a DATA REQUEST exercise on variable prioritisation (April 13th, 2022)

510 Greetings from the newly established CMIP International Project Office. As part of the CMIP community, you are invited to
511 participate in a DATA REQUEST exercise on variable prioritisation. We are supporting the WGCM Infrastructure Panel
512 (WIP) to implement this activity.

513 If you would like to participate in this activity, please complete this form <https://forms.office.com/r/qCNtTfywqN> <see
514 below> before 11am UTC 21 April 2022. This will enable you to:

515 ● Express interest in attending an online workshop in May

516 ● Express interest in being a paper author or reviewer

517 ● Contribute your thoughts on methodological approach (the questions are based on reviewing this list of parameters,
518 indicating how you feel about the number prioritised, the methodology proposed, any additional quantitative criteria you feel
519 should be taken into account in short-listing, any science/impact based prioritisation issues for consideration and any
520 thoughts you have on alternative methodological approaches to prioritisation)

521
522 If you have any questions about this, or would like to reach out to the new CMIP IPO about anything else, please do contact
523 myself or CMIP-IPO Director, Eleanor O'Rourke eleanor.orourke@ext.esa.int.

524
525 Form Introduction

526 **CMIP DATA REQUEST variable prioritisation: Event registration, input and author EoI**

527 CMIP has expanded and now has a substantial range of communities, all with their own specialised requirements. The
528 WGCM Infrastructure Panel (WIP) are aware that there are too many variables being listed as top priority and that conflicts
529 are emerging between what the data centres and data users (including intermediary platforms such as C3S) would consider
530 highest priority.

531
532 The Data Request function of the WIP wish to address the immediate challenge of establishing an agreed variable
533 prioritisation methodology from the CMIP modelling community and some means of giving authority to “priority = 1”
534 statements; a community intention discussed at WGCM 2019 in Barcelona. It is envisaged that these prioritised variables can
535 form a baseline set of variables for exchange of climate model data, in following FAIR data and Open Science principles.
536 The intention is to publish these as a Geoscientific Model Development (GMD) paper.

537
538 The CMIP community are therefore invited to provide input to, and consider self-nomination for paper authorship of, a paper
539 setting out an appropriate methodology for prioritising variables that could be considered as a baseline set of variables for
540 exchange of climate model data, in any intercomparison project, in accordance with FAIR data and Open Science principles.
541 There are three sections to this survey, it will take you 5-10 minutes to complete, longer if you wish to provide detailed
542 responses.

543
 544 Section 1- Your details (required)
 545 Section 2 - Workshop preference and EOI for paper roles (author/reviewer) (required)
 546 Section 3 - Your thoughts on methodological approach (optional) -these will be used to underpin workshop discussions
 547
 548 This participation form has been developed by the CMIP International Project Office, hosted by ESA Climate Office, in
 549 consultation with the WCRP WGCM Infrastructure Panel. This workstream is being led by Martin Jukes, UKRI-STFC
 550 working alongside Charlotte Pascoe NCAS/CEDA and Alison Parent, CEDA. If you have any problems completing this
 551 form/accessing the links please contact: briony.turner@ext.esa.int
 552
 553 This participation form has been issued by CMIP IPO to the Modelling Centre leads, data request leads and the MIP Chairs
 554 and can be shared more widely if you are aware of others that would wish to input into this activity.
 555
 556 Please note this Registration & Author Expression of Interest form closed 18:00 UTC 26 April 2022 however you can still
 557 share your thoughts on the methodological approach and indicate which workshop you'd like to attend until 18:00 UTC 6
 558 May 2022.
 559
 560 This activity is supported by the CMIP IPO and is made possible by funding from IS-ENES3 part of the European Union's
 561 Horizon 2020 research and innovation programme under grant agreement No 82408.
 562

563 **Appendix 5: Pressure levels for atmospheric variables**

564 The pressure levels defined in the CMIP6 Data Request and brought into the ESM-BCV list are listed below.
 565
 566 • 19 pressure levels (plev19): 100,000., 92500., 85,000., 70,000., 60,000., 50,000., 40,000., 30,000., 25,000., 20,000.,
 567 15,000., 10,000., 7,000., 5,000., 3,000., 2,000., 1,000., 500., and 100.hPa
 568 • 8 pressure levels (plev8): 100,000., 85,000., 70,000., 50,000., 25,000., 10,000., 5,000., and 1,000.hPa
 569
 570 The usage and the range of levels used may be modified in CMIP7 following detailed discussion of scientific requirements
 571 being led by the atmosphere theme of the CMIP AR7 Fast Trask Data Request¹⁵.
 572

573 **Data Availability**

574 The prioritisation data is available as an Excel workbook in Jukes, M., (2023) (<https://doi.org/10.5281/zenodo.8263526>).

92 15 See <https://wcrp-cmip.org/cmip7/cmip7-data-request/public-consultation/> [accessed 2024/12/04]

575 **Author contributions**

576 Conceptualisation: MJ. Funding acquisition: na. Methodology: MJ, BT, EO. Project administration: BT, EO, BD. Writing
577 and original draft preparation: all. Revision and verification of ESGF download statistics: FA, AN. Writing, review and
578 editing: all.

579 **Competing interests**

580 The contact author has declared that none of the authors has any competing interests

581 **Acknowledgements**

582 The work developing the baseline variables was coordinated and led by Martin Juckes, UKRI-STFC. Its implementation was
583 supported by the CMIP International Project Office, hosted by the European Space Agency. All figures have been
584 commissioned by the CMIP International Project Office and are under a Creative Commons Attribution 4.0 International
585 licence. The ESGF download statistics are provided by the Euro-Mediterranean Center on Climate Change, CMCC
586 Foundation. Thanks to John Dunne for chairing the discussions leading to final consensus on the published list. We
587 gratefully acknowledge the valuable feedback provided in reviews by Claire Macintosh and Young Ho Kim, and in
588 comments by Anne Marie Treguier, Isla Simpson, Alistair Adcroft, Baylor Fox-Kemper, Nathan Gillett, Christopher Danek,
589 Gaëlle Rigoudy, and Gavin A. Schmidt.

590 **Financial Support**

591 The work developing the baseline variables has been made possible by funding from IS-ENES3 part of the European Union's
592 Horizon 2020 research and innovation programme under grant agreement No 824084 and through NERC funding for the UK
593 National Centre of Atmospheric Science. The work of PJD and KET was performed under the auspices of the U.S.
594 Department of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract DE-AC52-07NA27344. It is a
595 contribution to the science portfolio of the U.S. Department of Energy, Office of Science, Earth and Environmental Systems
596 Sciences Division, Regional and Global Model Analysis Program. LLNL Release Number: LLNL-JNRL-853910

- 598 Buontempo, C., Burgess, S. N., Dee, D., Pinty, B., Thépaut, J.-N., Rixen, M., Almond, S., Armstrong, D., Brookshaw, A.,
 599 López-Alos, A., Bell, B., Cedric Bergeron, C., Chiara Cagnazzo, Edward Comyn-Platt, Eduardo Damasio-Da-Costa,
 600 Anabelle Guillory, Hans Hersbach, András Horányi, Julien Nicolas, Andre Obregon, Eduardo Penabad Ramos, Baudouin
 601 Raoult, Joaquín Muñoz-Sabater, Adrian Simmons, Cornel Soci, Martin Suttie, Freja Vamborg, James Varndell, Stijn
 602 Vermoote, Xiaobo Yang, and Juan Garcés de Marcilla. "The Copernicus Climate Change Service: Climate Science in
 603 Action", Bulletin of the American Meteorological Society 103, 12 (2022): E2669-E2687, doi:
 604 <https://doi.org/10.1175/BAMS-D-21-0315.1>, 2022.
- 605 Craig, M. T., Wohland, J., Stoop, L. P., Kies, A., Pickering, B., Bloomfield, H. C., Browell, J., De Felice, M., Dent, C. J.,
 606 Deroubaix, A., Frischmuth, F., Gonzalez, P. L., Grochowicz, A., Gruber, K., Härtel, P., Kittel, M., Kotzur, L., Labuhn, I.,
 607 Lundquist, J. K., Pflugradt, N., van der Wiel, K., Zeyringer, M., and Brayshaw, D. J.: Overcoming the disconnect between
 608 energy system and climate modeling, *Joule*, 6, 1405–1417, <https://doi.org/10.1016/j.joule.2022.05.010>, 2022.
- 609 Dubus, L., Brayshaw, D. J., Huertas-Hernando, D., Radu, D., Sharp, J., Zappa, W., and Stoop, L. P.: Towards a future-proof
 610 climate database for European energy system studies. *Environmental Research Letters*, 17(12), 121001.
 611 <https://doi.org/10.1088/1748-9326/aca1d3>, 2022.
- 612 Durack, P. J., Taylor, K. E., Mizielinski, M., Doutriaux, C., Nadeau, D., and Juckes, M.: WCRP-CMIP/CMIP6_CVs:
 613 6.2.58.68, <https://doi.org/10.5281/zenodo.12197151>, 2024.
- 614 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled
 615 Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, 9, 1937–1958,
 616 <https://doi.org/10.5194/gmd-9-1937-2016>, 2016.
- 617 Fiore, S., Nassisi, P., Nuzzo, A., Mirto, M., Cinquini, L., Williams, D., and Aloisio, G.: "A climate change community
 618 gateway for data usage & data archive metrics across the earth system grid federation." CEUR workshop proceedings. Vol.
 619 2975. CEUR-WS, <https://iris.unitn.it/handle/11572/329932>, 2021.
- 620 Flato, G. M.: Earth system models: An overview. *WIREs Climate Change*, 2(6), 783–80, <https://doi.org/10.1002/wcc.148>,
 621 2011.
- 622 Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., Chassignet, E. P., Curchitser, E.,
 623 Deshayes, J., Drange, H., Fox-Kemper, B., Gleckler, P. J., Gregory, J. M., Haak, H., Hallberg, R. W., Heimbach, P., Hewitt,
 624 H. T., Holland, D. M., Ilyina, T., Jungclaus, J. H., Komuro, Y., Krasting, J. P., Large, W. G., Marsland, S. J., Masina, S.,
 625 McDougall, T. J., Nurser, A. J. G., Orr, J. C., Pirani, A., Qiao, F., Stouffer, R. J., Taylor, K. E., Treguier, A. M., Tsujino, H.,
 626 Uotila, P., Valdivieso, M., Wang, Q., Winton, M., and Yeager, S. G.: OMIP contribution to CMIP6: experimental and
 627 diagnostic protocol for the physical component of the Ocean Model Intercomparison Project, *Geosci. Model Dev.*, 9, 3231–
 628 3296, <https://doi.org/10.5194/gmd-9-3231-2016>, 2016.
- 629 Guterres, A.: *Planet Hurtling towards Hell of Global Heating, Secretary-General Warns Austrian World Summit, Urging*
 630 *Immediate Emissions Cuts, Fair Climate Funding*. UN Press Release, <https://press.un.org/en/2023/sgsm21799.doc.htm>,
 631 2023.
- 632 Gerber, E. P. and Manzini, E.: The Dynamics and Variability Model Intercomparison Project (DynVarMIP) for CMIP6:
 633 assessing the stratosphere–troposphere system, *Geosci. Model Dev.*, 9, 3413–3425, [https://doi.org/10.5194/gmd-9-3413-](https://doi.org/10.5194/gmd-9-3413-2016)
 634 [2016](https://doi.org/10.5194/gmd-9-3413-2016), 2016.

635 Gutowski Jr., W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin,
636 G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., and Tangang, F.: WCRP COordinated Regional Downscaling
637 EXperiment (CORDEX): a diagnostic MIP for CMIP6, *Geosci. Model Dev.*, 9, 4087–4095, [https://doi.org/10.5194/gmd-9-](https://doi.org/10.5194/gmd-9-4087-2016)
638 [4087-2016](https://doi.org/10.5194/gmd-9-4087-2016), 2016.

639 Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S.,
640 Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J.,
641 Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High
642 Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, *Geosci. Model Dev.*, 9, 4185–4208,
643 <https://doi.org/10.5194/gmd-9-4185-2016>, 2016.

644 Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behraves, M.,
645 Bergemann, M., Biercamp, J., Bockelmann, H., Brokopf, R., Brüggemann, N., Casaroli, L., Chegini, F., Datseris, G., Esch,
646 M., George, G., Giorgetta, M., Gutjahr, O., Haak, H., Hanke, M., Ilyina, T., Jahns, T., Jungclaus, J., Kern, M., Klocke, D.,
647 Kluff, L., Kölling, T., Kornblueh, L., Kosukhin, S., Kroll, C., Lee, J., Mauritsen, T., Mehlmann, C., Mieslinger, T.,
648 Naumann, A. K., Paccini, L., Peinado, A., Praturi, D. S., Putrasahan, D., Rast, S., Riddick, T., Roeber, N., Schmidt, H.,
649 Schulzweida, U., Schütte, F., Segura, H., Shevchenko, R., Singh, V., Specht, M., Stephan, C. C., von Storch, J.-S., Vogel, R.,
650 Wengel, C., Winkler, M., Ziemann, F., Marotzke, J., and Stevens, B.: ICON-Sapphire: simulating the components of the Earth
651 system and their interactions at kilometer and subkilometer scales, *Geosci. Model Dev.*, 16, 779–811,
652 <https://doi.org/10.5194/gmd-16-779-2023>, 2023.

653 International Bureau of Weights and Measures: The International System of Units, 9th edition
654 <https://www.bipm.org/documents/20126/41483022/SI-Brochure-9.pdf> (last access: 1 November 2024), 2019.

655 Intergovernmental Panel on Climate Change: Summary for Policymakers. In: Climate Change 2023: Synthesis Report.
656 Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate
657 Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1-34, doi: 10.59327/IPCC/AR6-
658 9789291691647.001, 2023.

659 Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G.,
660 Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled
661 Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6, *Geosci. Model Dev.*, 9, 2853–
662 2880, <https://doi.org/10.5194/gmd-9-2853-2016>, 2016.

663 Juckes, M.: *CMIP Data Request Schema 2.0*, <https://doi.org/10.5281/ZENODO.4287148>, 2020.

664 Juckes, M.: Style Guide for Variable Titles in CMIP6 (0.01). Zenodo. <https://doi.org/10.5281/zenodo.2480853>, 2018.

665 Juckes, M., Taylor, K. E., Durack, P. J., Lawrence, B., Mizielinski, M. S., Pamment, A., Peterschmitt, J.-Y., Rixen, M., and
666 Sényesi, S.: The CMIP6 Data Request (DREQ, version 01.00.31), *Geosci. Model Dev.*, 13, 201–224,
667 <https://doi.org/10.5194/gmd-13-201-2020>, 2020.

668 Juckes, M.: WCRP Baseline Variables - MIP Prioritisation raw data [Data set]. CMIP IPO.
669 <https://doi.org/10.5281/zenodo.8263526>, 2023.

670 Kershaw, P., Abdulla, G., Ames, S., and Evans, B.: *ESGF Future Architecture Report*.
671 <https://doi.org/10.5281/ZENODO.3928223>, 2020.

672 Klein Tank, A., Zwiers, F., and Zhang, X.: Guidelines on Analysis of Extremes in a Changing Climate in Support of
673 Informed Decisions for Adaptation. World Meteorological Organization, 2009.

674 Lee, J., Gleckler, P., Ordonez, A., Ahn, M.-S., Ullrich, P., Vo, T., Boutte, J., Doutriaux, C., Durack, P., Shaheen, Z.,
675 Muryanto, L., Painter, J., and Krasting, J.: *PCMDI/pcmdi_metrics: PMP Version 2.5.1* (v2.5.1) [Computer software].
676 Zenodo. <https://doi.org/10.5281/ZENODO.7231033>, 2020.

677 Lee, J., Gleckler, P. J., Ahn, M.-S., Ordonez, A., Ullrich, P. A., Sperber, K. R., Taylor, K. E., Planton, Y. Y., Guilyardi, E.,
678 Durack, P., Bonfils, C., Zelinka, M. D., Chao, L.-W., Dong, B., Doutriaux, C., Zhang, C., Vo, T., Boutte, J., Wehner, M. F.,
679 Pendergrass, A. G., Kim, D., Xue, Z., Wittenberg, A. T., and Krasting, J.: Systematic and objective evaluation of Earth
680 system models: PCMDI Metrics Package (PMP) version 3, *Geosci. Model Dev.*, 17, 3919–3948,
681 <https://doi.org/10.5194/gmd-17-3919-2024>, 2024.

682 Magagna, B., Schinder, S., Stoica, M., Moncoiffe, G., Devaraju, A., and Pamment, A.: *I-ADOPT Framework*
683 *ontology* [Computer software]. <https://w3id.org/iadopt/ont/1.0.3> (accessed 1 November 2024) 2023.

684 Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: Intercomparison makes for a better climate model. *Eos,*
685 *Transactions American Geophysical Union*, 78(41), 445. <https://doi.org/10.1029/97EO00276>, 1997.

686 Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: The Coupled Model Intercomparison Project (CMIP), *B.*
687 *Am. Meteorol. Soc.*, 81, 313–318, 2000.

688 Meehl, G. A., Covey, C., Taylor, K. E., Delworth, T., Stouffer, R. J., Latif, M., McAvaney, B., and Mitchell, J. F. B.: The
689 WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research, *B. Am. Meteorol. Soc.*, 88, 1383–1394, 2007.

690 Nadeau, D., Doutriaux, C., & Taylor, K. E.: *Climate Model Output Rewriter (CMOR3)* (3.7) [Computer software].
691 <https://github.com/PCMDI/cmor>, 2022.

692 Notz, D., Jahn, A., Holland, M., Hunke, E., Massonnet, F., Stroeve, J., Tremblay, B., and Vancoppenolle, M.: The CMIP6
693 Sea-Ice Model Intercomparison Project (SIMIP): understanding sea ice through climate-model simulations, *Geosci. Model*
694 *Dev.*, 9, 3427–3446, <https://doi.org/10.5194/gmd-9-3427-2016>, 2016.

695 O'Rourke, E.: CMIP6 Community Survey Results. Zenodo. <https://doi.org/10.5281/zenodo.8113057>, 2023.

696 O'Rourke, E. and Turner, B. (eds): *Priority variables for evaluation and exploitation of WCRP climate simulations workshop*
697 *report*. Top priority variables community workshop - 12 and 17 May 2022. Available from:
698 <https://doi.org/10.59555/TUOC4428>, 2022.

699 O'Rourke, E., Turner, B., and Dingley, B. Baseline variables for climate model intercomparison and evaluation - Community
700 survey summary report. Zenodo. <https://doi.org/10.5281/zenodo.8248691>, 2023.

701 Pascoe, C., Lawrence, B. N., Guilyardi, E., Jukes, M., and Taylor, K. E.: Documenting numerical experiments in support of
702 the Coupled Model Intercomparison Project Phase 6 (CMIP6), *Geosci. Model Dev.*, 13, 2149–2167,
703 <https://doi.org/10.5194/gmd-13-2149-2020>, 2020.

704 Petrie, R., Eggleton, F., and Jukes, M.: CF Compliance errors in the CMIP6 archive. Zenodo.
705 <https://doi.org/10.5281/zenodo.12820690>, 2024.

706 Program for Climate Model Diagnosis and Intercomparison (PCMDI): Standard Output [Excel file], Lawrence Livermore
707 National Laboratory, https://pcmdi.llnl.gov/mips/cmip5/docs/standard_output.xls [accessed July 3rd, 2024], 2013.

708 Petrie, R., Denvil, S., Ames, S., Levavasseur, G., Fiore, S., Allen, C., Antonio, F., Berger, K., Bretonnière, P.-A., Cinquini,
709 L., Dart, E., Dwarakanath, P., Druken, K., Evans, B., Franchistéguy, L., Gardoll, S., Gerbier, E., Greenslade, M., Hassell, D.,
710 Iwi, A., Jukes, M., Kindermann, S., Lacinski, L., Mirto, M., Nasser, A. B., Nassisi, P., Nienhouse, E., Nikonov, S., Nuzzo,
711 A., Richards, C., Ridzwan, S., Rixen, M., Serradell, K., Snow, K., Stephens, A., Stockhause, M., Vahlenkamp, H., and

Wagner, R.: Coordinating an operational data distribution network for CMIP6 data, *Geosci. Model Dev.*, 14, 629–644, <https://doi.org/10.5194/gmd-14-629-2021>, 2021.

Ruane, A. C., Teichmann, C., Arnell, N. W., Carter, T. R., Ebi, K. L., Frieler, K., Goodess, C. M., Hewitson, B., Horton, R., Kovats, R. S., Lotze, H. K., Mearns, L. O., Navarra, A., Ojima, D. S., Riahi, K., Rosenzweig, C., Themessl, M., and Vincent, K.: The Vulnerability, Impacts, Adaptation and Climate Services Advisory Board (VIACS AB v1.0) contribution to CMIP6, *Geosci. Model Dev.*, 9, 3493–3515, <https://doi.org/10.5194/gmd-9-3493-2016>, 2016.

Stockhouse, M., Höck, H., Toussaint, F., and Lautenschlager, M.: Quality assessment concept of the World Data Center for Climate and its application to CMIP5 data, *Geosci. Model Dev.*, 5, 1023–1032, <https://doi.org/10.5194/gmd-5-1023-2012>, 2012.

Stockhouse, M. and Lautenschlager, M.: CMIP6 Data Citation of Evolving Data, *Data Science Journal*, <https://doi.org/10.5334/dsj-2017-030>, 2017.

Tan, J., Duan, Q., Xiao, C., He, C., and Yan, X.: A brief review of the coupled human-Earth system modeling: Current state and challenges, *The Anthropocene Review*, 10, 664–684, <https://doi.org/10.1177/20530196221149121>, 2023.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, *B. Am. Meteorol. Soc.*, 93, 485–498, 2012.

Taylor, K. E., Jukes, M., Balaji, V., Cinquini, L., Denvil, S., Durack, P. J., Elkington, M., Guilyardi, E., Kharin, S., Lautenschlager, M., Lawrence, B., Nadeau, D., and Stockhouse, M.: CMIP6 Model Output Metadata Requirements, Data Reference Syntax (DRS) and Controlled Vocabularies (CVs), <https://doi.org/10.5281/zenodo.12768887>, 2024.

Touzé-Peiffer, L., Barberousse, A., and Le Treut, H.: The Coupled Model Intercomparison Project: History, uses, and structural effects on climate research. *WIREs Climate Change*, 11(4). <https://doi.org/10.1002/wcc.648>, 2020.

United Nations: Paris Agreement, United Nations Treaty Collection, Chapter XXVII 7. d, adopted 2015-12-12 and inforce 2016-11-04, 2015.

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M., Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Materia, S., Law, R. M., and Sheffield, J.: LS3MIP (v1.0) contribution to CMIP6: the Land Surface, Snow and Soil moisture Model Intercomparison Project – aims, setup and expected outcome, *Geosci. Model Dev.*, 9, 2809–2832, <https://doi.org/10.5194/gmd-9-2809-2016>, 2016.

WCRP Working Group on Coupled Modeling: Report of the 22nd session of the WCRP Working Group on Coupled Modeling (WGCM), vol. 14 of WCRP Report, World Climate Research Programme (WCRP), Geneva, Switzerland, 2019.

Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., Adeniyi, K., Andela, B., Arnone, E., Berg, P., Caron, L.-P., Cionni, I., Corti, S., Drost, N., Hunter, A., Lledó, L., Mohr, C. W., Paçal, A., Pérez-Zanón, N., Predoi, V., Sandstad, M., Sillmann, J., Sterl, A., Vegas-Regidor, J., von Hardenberg, J., and Eyring, V.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – diagnostics for extreme events, regional and impact evaluation, and analysis of Earth system models in CMIP, *Geosci. Model Dev.*, 14, 3159–3184, <https://doi.org/10.5194/gmd-14-3159-2021>, 2021.

WGCM Climate Simulation Panel: IPCC Standard Output from Coupled Ocean-Atmosphere GCMs, Program for Climate Model Diagnosis & Intercomparison, <https://pcmdi.llnl.gov/mips/cmip3/variableList.html> (accessed 5th December 2024), 2007.

749 World Meteorological Organisation: The 2022 GCOS Implementation Plan, WMO GCOS 244, World Meteorological
750 Organisation, <https://library.wmo.int/idurl/4/58104> (last access: 5 December 2024), 2022a.

751 World Meteorological Organisation: GCOS-22: The 2022 GCOS ECVs Requirements, WMO GCOS-245, World
752 Meteorological Organisation, <https://library.wmo.int/idurl/4/58111> (last access: 5 December 2024), 2022b.

753 World Meteorological Organisation:WMO (2023). Manual on Codes – International Codes, Volume I.2, Annex II to the
754 WMO Technical Regulations, World Meteorological Organisation, <https://library.wmo.int/idurl/4/35625> (last accessed 5
755 December 2024), 2023.