

Protokoll DICAD 3-Jahrestreffen

Stephan Kindermann - ESGF

- Eingehende Daten müssen frühzeitig angemeldet werden um die Veröffentlichung planen zu können. Ansonsten können Verzögerungen auftreten (Urlaub, größere Datenmengen in der Queue).
- Versionierung: Die Veröffentlichung einer zweiten korrigierten Version der Simulationsdaten ist nur möglich wenn die errata-Informationen vorliegen. Dies kann unkompliziert mittels einer Web-Form vorgenommen werden: <https://errata.es-doc.org/static/index.html>
Alternativ gibt es den ES-DOC Errata Command Line Client: <https://es-doc.github.io/esdoc-errata-client/client.html>
Alle Informationen können dem "CMIP6 Participation Guidance for Modelers" entnommen werden: <https://pcmdi.llnl.gov/CMIP6/Guide/modelers.html>
- Die Publikation nach abgeschlossener QA dauert etwa 1-2 Tage.
- DOIs werden separat vergeben: die Autoren haben diese Informationen bereits bereitgestellt. Bei einer Publikation der Daten im ESGF erscheinen diese automatisch.

Christian: Autorenliste für MPI-ESM1-2-HR ist bereits festgelegt. Am DKRZ und vom DWD finden aber auch Simulationen damit statt. Kann diese Liste aktualisiert werden?

Katharina: Diese Liste kann aktualisiert werden.

ToDo: Liste aktualisieren -> DWD meldet Autoren an MPI (oder DKRZ?)

Christian: Können die ES-DOC Modellbeschreibungen vom MPI-M übernommen werden?

-> Martin/Johann: Das MPI-M wird keine Modellbeschreibung bereitstellen.

-> Katharina: Die Modellbeschreibung wird dann standardmäßig aus den netCDF-Headern erstellt. Es gibt keine Pflicht für die Autoren eine Modellbeschreibung zu erstellen.

Fabian: Werden die ES-DOC Modell-/Experimentinformationen unabhängig von der Errata-Information gesammelt?

Katharina/Stephan: Ja.

Eine Modelldokumentation in ES-DOC ist nicht erforderlich. Das MPI-M wird das auch nicht machen

HDH - Qualitätsanalyse (QA)

- Es zeigte sich, dass die Nutzung von CMOR bzw. cdo CMOR schneller und bessere Ergebnisse liefert als die Standardisierung "von Hand".
- Für Informationen zur Replikation, siehe angehängte Vortragsfolien.

Stephan: In Europa haben wir am meisten Speicherplatz für die Replikation zur Verfügung. 2 Petabyte im Vgl. zu 200-300 Terabyte in Frankreich und UK.

Lisa: Gibt es eine Priorisierung bei der Replikation?

HDH: Es gibt zwei Maschinen mit denen repliziert wird:

- 1) Hier haben die CMIP-Läufe (DECK+historicals) sowie die Szenarien Priorität.
- 2) Hier wird ein "Selectionfile" angewendet, das die Wünsche der IPCC-Autoren berücksichtigt und über das WIP verteilt wurde. In dieser Liste sind Variablen aus allen MIPs enthalten in Form von Table-IDs und Variablennamen.

Stephan: Sobald der Speicher voll ist müssen wir entweder über ein Aufstocken verhandeln oder weniger wichtige Daten entfernen.

Johann: Die von Veronika Eyring benannte Deadline (August 2019) für die Veröffentlichungen von Variablen ist nicht offiziell. Daten können auch nach diesem Termin noch für Auswertungen für den AR6 genutzt werden. Allerdings steht die "Paper-Submission"-Deadline Ende dieses Jahres.

Christian: Wie läuft die Veröffentlichung der DWD-Daten am ESGF-Datenknoten des DWD ab?

Stephan/Katharina: Nach QA am Rechner des DWD kann direkt am ESGF-Datenknoten des DWD veröffentlicht werden. Da er Teil des DKRZ-Indexknotens ist, werden die Daten dort über die Suchmaske abrufbar sein. Um sie später auch auf mistral zu haben müssen sie repliziert werden.

Stephan: Der Datenpool wird auch zur Analyse genutzt (ESMVal). Wenn nötig muss den Wissenschaftlern ein Konto am DKRZ/mistral eingerichtet werden. Ebenso müssen sie ein Projekt zugeordnet bekommen. Der Zugriff auf den Datenpool mit den (replizierten) CMIP6-Daten ist dann gegeben.

Lisa Bock - ESMVal

- Vorstellung des ESMValTools in v1.1 und 2.0, sowie der ESMVal Quicklooks zum Monitoring von Modellsimulationen. Siehe angehängte Vortragsfolien.
- Für die Anwendung des Quicklooks-Monitorings ist es nötig, dass die Daten standardisiert sind. Für Version 2 des ESMValTools wird das überarbeitet und Schnittstellen für andere Modelle ermöglicht.

Christopher: Einbindung von ESMValTool in Freva muss abgestimmt werden.

Martin: Gab es Probleme mit dem sich ändernden Datenstandard?

Lisa: Es gab keine Probleme.

Christopher: Wann wird der ESMVal Resultbrowser für die Öffentlichkeit freigegeben? Wird das Wasserzeichen auf den Plots entfernt?

Lisa: Geplant ist die Freischaltung für Ende Juli. Das Wasserzeichen wird vorerst bleiben und evtl. nach eigenen Veröffentlichungen entfernt.

Lisa: Bei Anwendung des ESMValTools werden viele Metadaten in die Header der Bilddateien geschrieben. Dies ist ein großer Pluspunkt, da so die Anforderung des IPCC, die Nachverfolgbarkeit und Reproduzierbarkeit der Plots, erfüllt wird.

Bianca Wentzel - Freva:

- Vorstellung der Zugriffsstatistik auf die Freva-Seite. Siehe angehängte Vortragsfolien.
- Großteil des Systems wird auf einer VM betrieben, nur ein Teil befindet sich auf Lustre-Filesystem des DKRZ Rechners mistral. Freva hat trotz des Lustre-Ausfalls tadellos funktioniert.

Christopher: Es ist geplant die Auswertung wieder mit Google Analytics oder ähnlichen Tools vorzunehmen. Was verwendet das ESGF, was spricht gegen Google Analytics, da keine personenbezogenen Daten erfasst werden?

Stephan/Katharina: Wir verwenden ElasticStack/Kibana. Aber die neue VM dafür ist noch nicht benutzbar. Gegen Google Analytics spricht nichts.

Stephan: Die SolR-Search-Version die in Freva verwendet wird ist als unsicher eingestuft.

Christopher: Wir planen ein Update von Django und SolR. Erst ist aber geplant von Python2 auf Python3 umzusteigen. Es ist zu erwarten, dass die neue SolR-Version auch deutliche bessere Performance und neue Nutzungsmöglichkeiten haben wird.

Lisa: Der Exif-Header entspricht nun dem W3-Prof-Standard und soll bleiben.

Christopher: Wir wollen diesen neuen Header auch unterstützen, haben aber noch eine Übergangslösung.

Christian Steger - DWD

- Vortrag zu Updates im WP des DWD. Siehe Vortragsfolien.
- Keine Neuigkeiten bei der Entwicklung der Nesting Option in ICON. Vera wird sich erst ab Herbst wieder damit beschäftigen können.
- Rechnungen für CMIP6: Die zwei Realisationen der Szenarien RCP2.6 und RCP8.5 müssen so bald wie möglich gestartet werden.

Johann Jungclaus - MPI-M

- Es gab Verzögerungen im Projekt durch Krankheitsfall und den Abgang von Matthias Bittner.
- Bei Anfragen die ans MPI-M bzgl. CMIP6 gerichtet werden bitte öfters nachhaken, da sie innerhalb der Abteilungen verteilt werden und auch mal der Überblick fehlt ob dem nachgegangen wird.
- Mit ICON-ESM sollen ab Herbst die DECK Experimente gerechnet werden. Die Daten sollen auch im ESGF veröffentlicht werden.

Fabian: Die vertikale Koordinate war in Testdaten von Vera nicht ordnungsgemäß definiert. Welche Koordinate wird es werden?

Johann: z-Achse mit 47 Schichten.

Fabian: Soll hier der CMIP-Standard erweitert werden um diese Achse zuzulassen oder soll interpoliert werden?

Christian: Der Erfahrung nach werden komplexe Gitter, die einen deutlichen Mehraufwand in der Auswertung bedeuten, gerne mal außen vor gelassen. Eine Interpolation ist daher sinnvoll, auch für das horizontale Gitter von ICON.

Fabian: Der CMIP-Standard unterstützt zumindest das horizontale Gitter. Das sollte keine Probleme machen, da hier eine Gitterbeschreibung beigegeben wird.

Christian: Der netCDF-Header der ICON-Modellausgabedaten muss optimiert werden. Gemäß Aufgabenverteilung in der ICON-Community ist das DKRZ dafür zuständig, daher sollte in jedem Fall ein Abstimmung erfolgen. Innerhalb der CLM-Community wird derzeit daran gearbeitet es ist allerdings unklar bis wann das fertig gestellt werden kann. Eine AG wäre hier sinnvoll.

Martin: Wurden die DECK-Simulationen, die aus MiKlip stammen, bereits dort archiviert?

Kamesh: Nein, sie müssen noch archiviert werden.

Martin: In DICAD haben wir dafür keine Ressourcen beantragt, die Archivierung (DECK) muss also noch im Projekt MiKlip stattfinden. (Anmerkung: Die historicals können in DICAD archiviert werden. Hierfür wurden Ressourcen für 2019 beantragt.)

Diskussion

Allocation Request

ToDo: Neue Ressourcenverteilung im Projekt muss bekanntgegeben werden. Einhaltung desselbigen ist nun zwingend notwendig um Planungssicherheit für die einzelnen Modellierungsgruppen zu gewährleisten.

Post-prozessing Workshop

- Bedarf akut vorhanden von MPI-M Seite.
- Angedachter Zeitraum: Vor Wintersemesterbeginn (Ende September, Anfang Oktober)

CORDEX Daten

- Mail an Grisha, Christopher Kadow, Katharina Bülow, ... um Datenanfrage einzuholen. Insbesondere was sich seit CMIP5 geändert hat.
- Deutschlandweite Einigung (REMO, WRF, CCLM, ...). Rest der Welt nur Deadline abwarten.
- Wenn keine Rückmeldung kommt: Gleiches Prozedere wie bei CMIP5: Veröffentlichung in CERA.

Paperveröffentlichungen Workflow (Vorschlag Fabian)

- Von Projektseite aus gibt es keine Vorbehalte/Vorzüge bestimmter Journals wenn es um Veröffentlichungen von einzelnen Projektmitglieder geht.
- Mögliche Journals für Infrastruktur-Software:
 - GMD.
 - J open research software. Hier scheint es Wartezeiten auf Grund fehlender Reviewer von über 6 Monaten zu geben.

Paperveröffentlichung Szenarien (Vorschlag Johann)

- AWI-CM und MPI-ESM1-2-HR/LR rechnen die Szenarien.
- Vergleich in Paperform möglich
- Anhand des Zeitdrucks eher unter einfacheren Gesichtspunkten
- Lisa: Es wird Ende August ein Leadautorentreffen geben. Davor werden die Ergebnisse zusammengetragen.

Kostenneutrale Verlängerung (Vorschlag Vera/Christian)

- (ToDo Christian) Gibt es Ausnahmen bei Elternzeit?
- jeder Projektpartner kann sowieso fünf Monate über das Projektende hinaus verlängern. Eine Verlängerung des Gesamtprojekt ist nur relevant, wenn darüber hinaus noch Mittel verfügbar wären, was beim DWD der Fall ist.
- FUB und DWD haben größeres Interesse an und Möglichkeiten für eine Verlängerung
- (ToDo Martin) Muss ein Verbund-Abschlussbericht erstellt werden oder ist es ausreichend für jedes Teilprojekt einen zu erstellen (und evtl. einen knappen Gesamtbericht)?

- (ToDo Martin) Kann ein (Teil-)Projekt IMMER bis zur Abgabe des Abschlussberichts (+6 Monate) verlängert werden?

Nächstes Projektmeeting

- Wunsch: Ende Januar.
- Es wird ein DICAD Teil im Postprocessing-Workshop miteingeplant. Dort werden:
 - Feinabstimmungen getroffen falls nötig
 - Ein Update beschlossen bzgl. gemeinsames Szenarienpaper (s.h. Absprache)

Im Anschluss wurden in kleinerer Runde die nötigen Schritte besprochen um zügig mit den Szenarienrechnungen am DKRZ und DWD beginnen und die CMIP-Läufe des MPI-ESM1-2-HR veröffentlichen zu können:

Abprache für Szenarienrechnungen mit dem MPI-ESM1-2-HR von DKRZ, MPI und DWD

Modellversion, -installation und -konfiguration

- (ToDo Kalle) Kalle erstellt eine Veröffentlichung P6 vom MPI-ESM1-2-HR mit neuer Konfiguration...
 - .. und mit "Hilfsskripten" von Martin
 - Mit überprüfung von JSBACH
 - Veg Stream
 - Yasso Stream
 - Überprüfung MPIOM (Meereisvariablen) u. ECHAM (tdiag-stream)
- (ToDo Christian) Kalle empfiehlt Christian zu überprüfen, wo und ob die Config-Dateien der MPI-ESM "Installation" von mistral und DWD-Server unterschiedlich sind.
- Das Forcing für das Jahr 2100 ist valide für MPI-ESM1-2-HR (nicht für LR)
- (ToDo Kalle) Die Restart-Dateien vom Ende der historicals werden in /pool/data gelegt. Kalle macht das und schickt den genauen Link per Mail.

Simulation

- Szenarien werden parallel zur DECK-Cmorisierung gerechnet.
- DWD startet auf historical Realisation 2 und heißt auch r2
- Die restart-Dateien der historicals liegen unter
/work/bk0988/m300237/CMIP6PROD/mpiesm-1.2.01-release-9157/experiments/historical_r?i1p1-HR/restart
- Die restart-Dateien werden zusätzlich in den CMIP Datenpool kopiert um sie länger und über das Projekt hinaus verfügbar zu haben.
- Monitoring: Verschieben der Monitoring-Daten nach /work/bk0988/web-monitoring/

Postprocessing

- Die Szenarien-Cmorisierung wird separat nachgeschaltet.

- (ToDo Fabi) Alle Institute nehmen die gleiche CDO und CMOR Version (Mail von Fabi)
- ggf. wird die Standardisierung für die Variablen die mit Hilfe von CDO expr berechnet werden nicht funktionieren; sollten aber sehr wenige sein
- Gleiche Infrastruktur?
- (ToDo ?) Evtl. braucht es einen zweiten Gang Postprocessing wegen fehlender "TEM" Diagnostiken
- Kalle: Die DECK-Läufe aus MiKlip sowie die 10 historischen Läufe aus DICAD mit dem MPI-ESM1-2-HR können nun standardisiert werden. Die Plausibilitätsprüfung ist abgeschlossen.

Analyse

- Soll ein gemeinsames Paper mit einem DICAD-Modellvergleich der Szenarienrechnungen erstellt werden?
 - (ToDo Johan) Johann fragt Tido
 - Genauere Absprache beim Postprocessingworkshop (s.h. X)

Randbedingungen:

- Urlaub: Kalle Ende Juli, Christian Mitte August bis Anfang September; dann Dienstreisen bis 23.09, HDH Ende Juli, Fabi Anfang Juli