

Protokoll vom 1. Jahrestreffen des BMBF-Projekts „Unterstützung der CMIP6⁺-Aktivitäten in Deutschland“

Mi. 11. Okt 2017, DKRZ Hamburg, Bundesstrasse 45a

Verbund 2 CMIP6-Chemie

Vortrag: Markus Kunze (DLR):

„Stand der CMIP6-Aktivitäten mit EMAC“ (Verbund 2 ‚Chemie‘)

Im Vorhaben sollen die DECK-Experimente mit EMAC durchgeführt werden, die für die Teilnahme an AerChemMIP verlangt sind. Gegenwärtig wird ein Spinup-Lauf ohne Chemie durchgeführt. Die Ergebnisse sind plausibel und das Modell befindet sich hinreichend gut im Gleichgewichtszustand. Das piControl-Experiment kann demnächst gestartet werden. Als Atmosphärenmodell wird ECHAM5 genommen.

Diskussion

Die EMAC-Modellierer sind gegenwärtig dabei, das DatenRequest-Web-GUI (s. AP3) auszufüllen. Da ECHAM5 benutzt wird, und der DatenRequest für AerChemMIP große Unterschiede zu denen der anderen MIPs aufweisen dürfte, ist nicht zu erwarten, dass große Synergien mit den anderen Partnern im Vorhaben bei der Anwendung der Infrastruktur (Web-GUI, Processing-Skripte) entstehen. Das DKRZ begrüßt, dass das DLR die Infrastruktur testet.

Die **Universität Bonn** hat keinen Vertreter geschickt.

Verbund 1 CMIP6-DICAD

AP1 CMIP6⁺-Experimente

Vortrag: Johann Jungclaus (MPI-M) i.V. Matthias Bittner:

„DECK-Experimente mit MPIESM-1-2“

Diskussion

CMIP6-Experimente sollen mit 3 verschiedenen Auflösungen von MPIESM-1-2 durchgeführt werden:

low resolution (LR: T63L47, GR15L40)

high resolution (HR: T127L95, TP04L40)

diese Modellversion wird für die im DICAD-Projekt geplanten Szenarien-Experimente benutzt. Im Gegensatz zur LR-Version wird ohne dynamische Vegetation gerechnet. Der Grund liegt hauptsächlich daran, dass der Aufwand die BGC-Zyklen in einen stationären Zustand zu rechnen, nicht geleistet werden kann.

Da in MiKlip und DICAD eine identische Modellversion benutzt werden soll, sollten die Szenarienprojektionen das in MiKlip erstellte Kompilat benutzen. Wie dies koordiniert werden soll, muss noch diskutiert und festgelegt werden. Grundsätzlich sollte erkennbar sein, auf welcher SVN-Revisionsnummer des benutzten Quellcodes das Modell basiert. Wg. der erforderlichen Reproduzierbarkeit der Experimente sollte dies auch an Hand der gesicherten Metadaten (z.B. der Skripte) möglich sein. Wenn die Versionsnummer nur im ausführbaren Programm steht, kann sie später an Hand der Skripte nicht mehr festgestellt werden. Es wurde vorgeschlagen, das Modell als

,module' laden zu können. Mittlerweile ist das geschehen. Nähere Informationen versendete KHW an alle Teilnehmer per Mail (02.11.2017). Zurzeit läuft eine Spinup-Rechnung im ca. 500ten Jahr. Die meisten Diagnostiken, die schon in den Modellen stattfinden sollen, sind mittlerweile eingebaut.

extra high resolution (XR, T255L95, TP04L40)

die Experimente mit dieser Modellversion werden ausschließlich am MPI-M durchgeführt; sie werden auch nicht mit den Ressourcen gerechnet, die im Rahmen des DKRZ-RZ-Projekts beantragt, bzw. bewilligt werden; die Experimente mit XR werden ebenfalls bald starten; es wird aber kein komplettes DECK gerechnet. Die Ergebnisse des höher auflösenden Modells sind signifikant besser.

Die letzte Version der stratosphärischen Aerosoldaten für das historical-Experiment stammt vom 2017-09-15. Daraus werden auch die Daten der DECK-Experimente berechnet. Da der Spinup-Lauf mit MPIESM-1-2-HR zu der Zeit schon gestartet war, werden jetzt alle CMIP6-Experimente mit diesem Modell mit den ,alten' Antriebsdaten gerechnet, d.h. mit den Versionen der externen Daten, die zum Startpunkt der DECK-Experimente vom CMIP6-Projekt freigegebenen waren. Zurzeit sind alle Antriebsdaten für die DECK- und historical-Experimente (1850-2014) für die Benutzung in CMIP6 freigegeben (s. <http://goo.gl/r8up31>, Version vom 6. Oktober 2017 (v6.2.1)). Die Daten für die Szenarienprojektionen (2015-2100) sind in der Mehrheit noch mit 'under preparation' oder 'uncertain' bezeichnet.

In MPIESM-2 (ICON) müssen die meisten Diagnostiken für den CMIP6-Datenrequest noch eingebaut werden. Deshalb wird für nächstes Jahr nicht mit dem Start der CMIP6-Experimente gerechnet.

Es wurde diskutiert, wann für ein CMIP6-Modell ein neuer Name benötigt wird.

Eine neue source_id könnte suggerieren, dass das CMIP5-Modell ein komplett anderes ist als das CMIP6-Modell. Der Modellname des ,alten' (CMIP5) MPI-M-Modells wird von MPI-ESM auf MPIESM-1-2 geändert werden, wie gegenwärtig in CMIP6_CV.json angegeben. Das ,-1-2' suggeriert eine neue Version.

Grundsätzlich müssen Modelle auf verschiedenen Gittern einen eigenen Namen bekommen. Jedes Modell muss die DECK- und das historical-Experiment durchführen, und die Daten im ESGF publizieren, bevor Ergebnisse der von den EndorsedMIPs definierten Experimente publiziert werden können. Im Umkehrschluss müssen Modelle von denen Ergebnisse der DECK- und historical-Experimente publiziert werden sollen, einen eigenen CMIP-Namen haben (andernfalls würden Daten von gleichnamigen Modellen beim Herunterladen überschrieben werden; der Projektname (CMIP5, CMIP6) steht nicht in Dateinamen).

Wenn ein Modellcode nur geringfügig geändert wird, so dass neue DECK- und historical-Experimente nicht erforderlich sind, da es sich nur um andere Realisationen handeln würde, kann das mit der zum Buchstaben ,p' gehörenden Nummer (,N') im Membernamen angezeigt werden (in CMIP6: rLiMpNfK).

Das DKRZ bittet noch mal, frühzeitig Modellrohdaten zur Verfügung zu stellen. Andernfalls kann der volle Funktionsumfang des im Projekt entwickelten cmor-Operators nicht getestet werden. Dasselbe gilt für die Skripte zur 'operationellen' Nachverarbeitung, also der Nachbereitung synchron mit der Modellintegration.

Der Ansprechpartner in allen Belangen der MPI-M-Modelle sind Matthias Bittner oder Karl-Herman Wieners.

Der **DWD** hat keinen Vertreter geschickt.

Vortrag: Dmitri Sein (AWI, assoziierter Partner)
„Stand der AWI-CM Rechnungen“

Diskussion

Zurzeit sind 180 Jahre mit unveränderten Randbedingungen gerechnet. Die Atmosphärenkomponente sollte mit der des MPIESM-1-2-HR identisch sein (die gegenwärtig benutzte Version ist MPIESM.1.2.00p4 mit ECHAM6.3.02p4, veröffentlicht am 09.03.2016). Wenn eine identische Modellversion benutzt wird, können Mapping-Tabellen und entsprechende Teile des Postprocessing-Skripts vom MPI-M und AWI gemeinsam benutzt werden. Mittlerweile (16.11.2017) wurde die aktuellste ECHAM-Version mit FESOM gekoppelt. Die 180-Simulationsjahre, die eigentlich als Teil des piControl-Experiments gelten sollten, wurden zur spin-up-Rechnung umgewidmet.

Wie die Formatierung im CMIP6-Standard durchgeführt werden soll, ist noch nicht klar. Die Kollegen vom AWI haben Befürchtungen, dass die Verarbeitung mit CMOR3 bzw. dem entsprechenden CDO-Operator an dem FESOM-Gitter scheitern wird. Ihnen wurde versichert, dass gegenwärtig keine Hindernisse bekannt sind, und es auf einen Versuch ankommt.

Das AWI überlegt, alle Daten interpoliert auf ein reguläres Gitter abzugeben. Es fragt sich, ob das mit der CMIP6-Datenanforderung verträglich ist. Am AWI wird bereits das ESMValTool benutzt. Der CMIP6-Name (source_id) ist ‚AWI-CM-1-0-HR‘, wie gegenwärtig in CMIP6_CV.json angegeben.

AP2 Kompatibilität der CDOs mit CMIP6⁺-Datenstandards

Vortrag: Fabian Wachsmann (DKRZ)
‘Der CDO cmor-Operator‘

Die CMOR2-Library, die am PCMDI für CMIP5 entwickelt wurde, ist in den im Vorhaben entwickelten cmor-Operator der CDOs eingebunden, und wird schon an verschiedenen Instituten und für verschiedene Projekte (DKRZ für HAPPI, KIT für CMIP6, MPI-M für Paläodaten) benutzt. Eine Installation ist auf der mistral am DKRZ bereitgestellt. Im Vorhaben wurde die Re-Formatierung der mit den CMIP5-Modellen MPIESM-HR/-MR/-P des MPI-M am DKRZ erzeugten Variablen getestet. Auch CORDEX-Daten können mit dieser Version standardgemäß formatiert werden. Allerdings ist für CORDEX-Daten noch eine Umbenennung und Änderung der Verzeichnisstruktur (DRS) nötig. CMOR2 kennt nur die DRS von CMIP5.

Die Anbindung der CMOR3-Library, die für die Re-Formatierung der CMIP6-Variablen benutzt werden soll, wird aktuell getestet. Eine Installation ist auf der mistral vorhanden. Allerdings müssen beta-Tester sich auf Anpassungen einstellen, die für neuere Versionen von CMOR3 nötig werden können.

Der cmor-Operator verwendet die CDIs um die Informationen, die sich in den Eingabedaten befinden, auszuwerten, falls sie nicht vom Anwender überschrieben werden.

Die Überschreibung von Angaben im Datenfile ist u. a. mit Hilfe von Textdateien möglich, die über ein Schlüsselwort dem Operator bekannt gemacht werden.

Diskussion

Es wurde diskutiert, inwieweit Modellgitter bei der Anwendung vom cmor-Operator ein Problem darstellen können. Im Prinzip sollte der Operator in der Lage sein, Daten zumindest auf CF-Gittern standardgemäß aufzubereiten. Es wird allerdings eine CF-gemäße Gitterbeschreibung verlangt. Gegebenenfalls muss sie mit dem WIP abgestimmt werden. Ansonsten ist die Möglichkeit der Gitterbeschreibung allgemein gehalten: zu jedem Gitterpunkt müssen geographische Koordinaten

abgegeben werden, sowie die maximale Anzahl von Gitterzellecken (vertices) und zugehörige Koordinaten. Fehlende Gitterinformation kann in einer Gitterspezifikationsdatei bereitgestellt werden (cdo cmor spezifisch). Allerdings ist schon angedacht, die Funktionsfähigkeit des Operator mit Beispiel-Datensätzen sicherzustellen. Als Beispiel wurden polar-stereographische Gitter der Arktis/Grönland- und Antarktis genannt. Das MPI-M plant zurzeit keine Abgabe von Variablen auf diesen Gittern, so dass hier kein Test stattfinden wird.

Im Zusammenhang mit den Modellgittern wurde auch die für CMIP6 eingeführte ‚nominale‘ Auflösung diskutiert. Dabei handelt es sich um ein NetCDF-Attribut, für das im CMIP6-ControlledVocabulary eine Liste von erlaubten Werten bereitgestellt wird. Für die Auswahl der richtigen Vokabel wird ein Algorithmus zur Verfügung gestellt (siehe <http://goo.gl/v1drZl>). Die ‚nominale‘ Auflösung ist nicht zu verwechseln mit der DRS-Komponente <grid_label>. Die erste ist keine DRS-Komponente, also nicht Teil des Dateipfades noch des Dateinamens. Dagegen ist das grid_label als DRS-Komponente Teil des Dateinamens als auch des Dateipfades. Es wird aus einer CV-Liste (CMIP6_CV.json) gewählt. Beide Terme sind gitterabhängig, d.h. in einem ESM mit mehreren Gittern werden jedem Gitter sowohl eine nominale Auflösung als auch ein grid_label zugeordnet. Sollte eine Modellierungsgruppe der Ansicht sein, dass das Verfahren unangemessen ist, kann sie vom WIP eine Ausnahme verlangen (siehe <http://goo.gl/v1drZl>).

Es wurde gefragt, ob der Funktionsumfang der CDIs ausreicht, um alle gewünschten Informationen aus den Datenfiles auszulesen. Zurzeit sind keine Defizite bekannt. Die CMOR3-Library erlaubt auch Angaben über JSON-Dateien. Dieses Format wurde in früheren Versionen des cmor-Operators unterstützt, ist aber zurzeit deaktiviert, da kein Bedarf vorhanden zu sein scheint. Mit dem Text-Format besteht auch eine größere Ähnlichkeit zwischen der Benutzung der mit CMOR2 und CMOR3 verlinkten Versionen des cmor-Operators. Eine Eingabe mit JSON-Formaten wird mit CMOR2 nicht möglich sein.

Die CMOR-Library führt eine Konvertierung von Daten in die vorgeschriebenen Einheiten durch. Dabei muss allerdings berücksichtigt werden, dass die übergebene Einheit ein von der UDUNITS-Library akzeptiertes Format hat (<http://ferret.pmel.noaa.gov/Ferret/documentation/udunits.dat>). Außerdem kann mit dem cmor-Operator auch die CMOR-Append-Funktionalität benutzt werden. Dabei werden Daten an schon existierende NetCDF-Dateien angehängt. Mit dem cmor-Operator besteht zudem die Möglichkeit eine maximale Dateigröße anzugeben, ab der automatisch von einem vorgegebenen Append-Modus auf den Replace-Modus (eine neue Datei wird angefangen) umgeschaltet wird.

Im Antrag ist geplant, auch die Formatierung von CORDEX-Daten zu unterstützen. Daneben sind noch ana4MIPs- und obs4MIPs-Daten in der Diskussion. Da CMOR3 flexibler in Hinblick auf Dateinamen und -pfade ist, sollte dies besser als mit CMOR2 möglich sein.

Für CORDEX-Belange ist eine Task-Force im Entstehen, die dafür sorgen soll, dass CMIP-Belange an die CORDEX-Community weitergegeben werden, und eine stabile Version des cmor-Operator für CORDEX-Daten, die auf CMIP6-Daten basieren zur Verfügung gestellt wird. HDH meldet sich freiwillig.

Es wurde diskutiert, wie aus dem ESGF stammende Variable unterschieden werden können. Dazu gibt es zum einen die Projektangabe (z.B. CMIP5, PMIP, CMIP era) und IDs für das Modell und das Experiment. Versionsnummern werden beim Kopieren vom ESGF nicht automatisch mitgeliefert, so dass eventuell nur die in den Dateien enthaltene UUID (tracking_ID) zur Unterscheidung verschiedener Versionen herangezogen werden kann. Deshalb ist es auch nicht erlaubt, Daten zur ESGF-Publikation mit einer tracking_ID abzuliefern, die schon im ESGF vorhanden ist. Die Unterscheidung von Dateien via model_id (CMIP5) oder source_ID (CMIP6) sollte unproblematisch sein. Grundsätzlich wird von einem neuen Modell (neuer Modellname) verlangt, Daten für alle DECK- (piControl, 1pctCO2, abrupt4xCO2, amip) und CMIP6-Experimente (historical) abzuliefern. Modelle auf unterschiedlichen Gittern werden grundsätzlich als verschiedene Modelle betrachtet, und müssen eine eigene model_id (CMIP5) bzw. source_id (CMIP6) (Modellname) haben. Es besteht außerdem die Möglichkeit, Daten von Modellen, deren Ergebnisse statistisch zu demselben Ensemble gehören sollen, über die

Memberkennung, und darin speziell die dem Buchstaben ‚p‘ (wie Physik) zugeordnete Nummer zu unterscheiden (RIP in CMIP5, variant_label in CMIP6).

Es wurde zudem darauf hingewiesen, dass eine Installation der CDOs existiert, die u. a. durch Verwendung eines anderen (als zurzeit in der freigegebenen Version verwendeten) Compilers sehr viel effektiver ist (im Vorhaben wurde an Hand von MPIESM-1-2-HR-Daten für die Interpolation auf Druckflächen eine Beschleunigung mit Faktor 2 beobachtet).

Zum Schluss wurde diskutiert, ob und in welcher Form eine Nutzer-Unterstützung gewünscht ist. Mehr Unterstützung für lokale Installationen wird zugesagt. Verschiedene Formate (e-learning, youtube, Dokumentation, en vivo Schulungen) wurden angesprochen. Eventuell sind auch hands-on-Sitzungen im Rahmen von Treffen einzelner endorsedMIPs sinnvoll, an denen auch andere technische Aspekte direkt auf die MIP-Aktivitäten bezogen angesprochen werden können (z. B. Datenrequest, cmor-Operator, automatische Erzeugung von Postprocessing-Skriptcode). Es sollte unterschieden werden, zwischen Fragen technischer Art und solchen, die durch die MIP-Koordination geklärt werden müssen. Zu letzteren gehören z.B. die Fragen nach 2 Zeitachsen (wie in SPECS erlaubt), oder wann dekadische Experimente gestartet, und ab welchem Zeitraum Daten abgeliefert werden sollen.

Dagegen müssen Fragen nach Inkonsistenzen im Projektstandard vom WIP geklärt werden. Angesprochen wurden in dieser Hinsicht saisonale Frequenzen, die für CORDEX/CMIP5-Daten verlangt, aber nicht von CMOR2 unterstützt wurden. Auch die Frage, ob klimatologische Datensätze eine eigene Tabelle haben sollten, gehört dazu. Der gegenwärtige CMOR-Standard hat sich nicht festgelegt (es gibt klimatologische Daten in Amon, während monatliche Klimatologien von Ozean-daten eine eigene MIP-Tabelle (Oclim) haben. Fabian hat beim PCMDI nachgefragt, aber noch keine Antwort erhalten. In diese Kategorie gehören auch Fragen nach der CF-Kompatibilität vom CMIP6-Standard.

AP3 Infrastruktur zur Anpassung an CMIP6 Projektdatenstandards

Vortrag: Martin Schupfner (DKRZ)

CMIP6-Variablenanforderung und Anpassung der Modelldaten an den CMIP-Standard

Das DataRequest-Web-GUI wurde vorgestellt, sowie der Workflow mit dem aus den eingegebenen Informationen Skripte zur CMIP6-konformen Datenformatierung hergestellt werden können. Es wird kein vollständiges Postprocessing-Skript erstellt, sondern Skriptcode (ksh), sogenannte Header-Files, die in existierende Postprocessing-Skripte eingebaut werden können. Die Erzeugung der Header-Files soll über die Website möglich sein.

Diskussion

Auf Grund der Entwicklung der Zusammenarbeit mit den Projektpartnern wurden einige Meilensteine neu formuliert. So wird z. B. kein Skript für die diagnostische Nachbereitung und standard-konforme Formatierung erstellt, sondern eine Infrastruktur entwickelt, die abhängig von aktivitäts-bezogenen CMIP6-Datenanforderungen Skriptcode erzeugt, der von den Modellierern in ihre haus-eigenen Skripte eingepflegt werden kann. Dabei wird besonders Wert auf die Modularität der Datenaufbereitung gelegt. Aus der Modularität ergibt sich, dass die Postprocessing-Skripte nicht für Performance optimiert werden. So werden teilweise Berechnungen doppelt durchgeführt oder auf Pipen (I/O Minimierung) verzichtet. Der Vorteil ist, dass die Variablen unabhängig voneinander berechnet werden können, und ohne Seiteneffekte die Berechnung von einigen Variablen herausgenommen oder die von anderen hinzugefügt werden kann. Außerdem ist es sinnvoll, bis auf den Aufruf des CDO-cmor-Operators, alle Verarbeitungsschritte parallel mit der Modellintegration durchzuführen und die Re-Formatierung mit CMOR3 nach Beendigung des Experiments nachzuziehen.

Skripte und GUI sind soweit es zurzeit möglich ist, d.h. mit dem CMIP5-Variablenumfang getestet.

Da der Datenrequest sehr umfangreich ist, wurde bei der Entwicklung des Web-GUIs und dem erzeugtem Skriptcode Wert darauf gelegt, Teamarbeit zu unterstützen. So ist es z.B. möglich, ‚Editor’s notes‘ im GUI abzulegen, nach denen von anderen Nutzern im GUI oder im erzeugten Skriptcode gesucht werden kann.

Es wurde noch mal darauf hingewiesen, dass dringend Tests mit den Tools durchgeführt werden müssen. Auf Grund des hohen Grades an Parallelität (Backgroundprozesse) ist das Debuggen nicht ganz einfach.

AP4 Datenqualitätsprüfung

Vortrag Heinz-Dieter Hollweg (DKRZ)
‘Datenqualitätsprüfung‘

Das QA-Tool wurde vorgestellt und erschiedene Installationswege wurden aufgezeigt. Es wurde vorgeführt, wo die Konfiguration vorgenommen wird und wie der Output aussieht (YAML Format).

Diskussion

Es muss sichergestellt werden, dass die Nutzer miteinander verträgliche Versionen der verschiedenen im Vorhaben entwickelten Tools benutze. Wobei bedacht werden muss, dass auch die verschiedenen Teile der externen Software nicht immer zusammenpassen müssen. Z. B. werden die MIP-Tabellen am PCMDI aus dem Datenrequest erstellt, so dass die MIP-Tabellen möglicherweise auf einer älteren Version des Datenrequests beruhen. Das muss noch geregelt und kommuniziert werden.

Das QA-Tool, dessen Kernprogramm in C++ geschrieben ist, ist in ein bash- und ein Python-Skript eingebettet (die doppelte Programmierung hat historische Ursachen). Es wurde u.a. erfolgreich am DKRZ für CMIP5- und HAPPI-Daten, und am DKRZ, KNMI, SMHI für CORDEX-benutzt.

Es wurde angeregt, die Anwendung des QA-Tools im Projekt-Redmine zu erläutern. Das wurde zugesagt.

CMIP6-Teilnehmer können einzelne Dateien auch auf einer Internetseite (<https://bovec.dkrz.de>) hochladen, und deren Konformität mit einem im Vorhaben entwickelten ‚spot-check‘-Tool feststellen lassen. Es wird grundsätzlich geraten spot-checks für einzelne Dateien durchzuführen, bevor große Datenvolumina formatiert werden.

Auch vom PCMDI wird ein Tool zur Qualitätskontrolle zur Verfügung gestellt (PREPARE). Dieses Tool wird im DKRZ-QA-Tool aufgerufen. Es prüft aber lediglich die Konsistenz und Richtigkeit von globalen Attributen mit den DRS-Komponenten und deren Anordnung im Dateinamen und in der Verzeichnisstruktur. PREPARE kann nicht auf andere Projektdaten angepasst werden (z.B. CORDEX, HAPPI).

Die minimale Anforderung an CMIP6-Daten für die ESGF-Publikation ist das unbeanstandete Passieren von PREPARE. Die Prüfungen des DKRZ-QA-Tools gehen darüber hinaus. Sie prüfen auch die Konsistenz zwischen verschiedenen Dateien, die zu einem ‚atomic data set‘ gehören, sowie die Werte der Target-Variablen in den Dateien auf unerwartete Werte und Muster.

Für DKRZ-Nutzer wird es ein ‚module‘ geben, das geladen werden kann. Für Nutzer anderer Plattformen wird ein Installationskript bereitgestellt.

Als problematisch wird die Frage angesehen, wann die dem QA-Tool zugrunde liegenden DatenRequests und MIP-Tabellen, sowie die CMOR3-Library von den Nutzern als ‚final‘ betrachtet werden können. Wenn die Finalität einmal zugesagt wurde, und die CMIP6-Teilnehmer mit der Datenproduktion und Formatierung angefangen haben, ist es nicht realistisch den Zugang zum ESGF zu verweigern, wenn die Weigerung auf späteren Änderungen von CMIP6-Standard-

Komponenten basiert. Demnach müssen ab der ‚Finalitäts‘-Ausrufung alle freigegebenen Versionen als Module am DKRZ zur Verfügung gestellt werden. Für externe Nutzer müssen entsprechende Versionen abrufbar sein.

AP5 Nationales CMIP6 Datenarchiv

Vortrag: Stephan Kindermann (DKRZ)
'DKRZ CMIP Datenpool und ESGF Knoten: Stand'

Für den CMIP-Datenpool ist ein Testservice aufgesetzt. Auf Anfrage wird der Zugang zu den Daten ermöglicht. Aktuell sind aber nur CMIP5-Daten im Pool, da noch keine CMIP6-Daten zur Publikation eingereicht wurden. Wenn das beginnt, und der Plattenplatz im Pool knapp wird, werden die CMIP5-Daten auf Bänder ausgelagert. Der Datenpool hält insgesamt 5 PB Speicherplatz vor, 1 PB davon für den deutschen Beitrag. Der mount point zu dem Poolbereich ist /work/kd0956.

Ein neuer ESGF-Knoten ist am DKRZ installiert und zur operationellen Nutzung bereit. Auf diesem Knoten sind Services für die Zuordnung von Persistent Identifiers (PI) eingerichtet.

Im Gegensatz zu CMIP5 werden in der CMIP6-Phase institutsbezogene Help-Adressen eingerichtet. Sie können also bei jedem Knoten unterschiedlich sein.

Der Datentransfer von anderen Instituten zum DKRZ ist langsam wg. geringer Bandbreite.

Datenpool nahes Prozessieren ist möglich auf der mistral und VMS. Demnächst soll ein sogenanntes Jupyterhub eingerichtet werden.

Diskussion

Der PI ersetzt die tracking_id von CMIP5 (NetCDF Attribut dessen Wert von der UUID gestellt wird). Wie die tracking ID bei der Formatierung mit CMOR2 erzeugt und in die Datei geschrieben wird, wird der PI durch CMOR3 gesetzt. Es handelt sich dabei ebenso um eine UUID, allerdings mit einem Prefix, der den Publikationsknoten angibt. Im Gegensatz zur tracking ID wird der PI in einem handle-Server registriert, aber erst bei der ESGF-Publikation. Da der PI bei der Formatierung mit CMOR3 erzeugt wird, bekommt jede Datei einen eigenen PI, auch wenn sie Teil eines ‚atomic data sets‘ ist. Die PI-Services erlauben jetzt, einer beliebigen Dateigruppe einen einzelnen PI zuzuordnen, vorausgesetzt jede Datei der Gruppe besitzt einen PI (z. B. alle zu einem ‚atomic data set‘ gehörenden Dateien, alle auf einem Modell basierenden publizierten Dateien von allen Realisationen eines Experiment, etc.).

Die CMIP6-Verzeichnisstruktur endet mit einem Verzeichnis, das entsprechend dem Tag der Dateierzeugung benannt wird (vYYYYMMDD). Da sich bei einem Workflow wie er am DKRZ aufgesetzt wird, die Erzeugung eines ‚atomic data sets‘ über eine längere Zeitspanne hinziehen kann (eventuell Monate), würden Dateien in verschiedenen Verzeichnissen landen, obwohl sie zu demselben ‚atomic data set‘ gehören. In der letzten Version von "CMIP6 Global Attributes, DRS, File-names, Directory Structure, and CV's" (Taylor et al., 2017; <http://goo.gl/v1drZl>) heißt es, "Note that files contained in a single <version> subdirectory at the end of the directory path should represent all the available time-samples reported from the simulation; a time-series can be split across several files, but all the files must be found in the same subdirectory. This implies that <version> will not generally be the actual date that all files in the subdirectory were written or published."

Es ist nicht klar, wie dies gewährleistet werden soll, wenn die Versions-Verzeichnisse automatisch von CMOR3 erzeugt werden. Weder haben die ESGF-Leute am DKRZ noch die Datenproduzenten Lust, die Verzeichnisse glatt zuziehen. Fabian wird sich darum kümmern.

AP6 – Standardisierte Diagnostiken und Modellevaluation

Vortrag: Björn Brötz (DLR)
'ESMValTool als Monitoring- und QA-Tool'

Die Überwachung von laufenden Simulationen mittels des ESMValTool-Quicklooksystems wurde vorgestellt. Das ESMValTool-Quicklooksystem ist als Modul auf der mistral ladbar. Die webbasierte Darstellung der Monitoring-Ergebnisse wurde vorgeführt. Aktuelle Arbeitsergebnisse wurden mit Milestones abgeglichen.

Vortrag: Lisa Bock (DLR)
'Nutzung des ESMValTools zur Unterstützung der CMIP6-Wissenschaftler'

Der Stand der Entwicklung und Nutzung des Tools zur routinemäßigen Evaluation der CMIP6-Simulationen wurde vorgestellt. Aktuell wird an einer Integration von Tags gearbeitet, die eine Sortierung der Plots möglich macht. Zusätzlich soll damit eine Reproduzierbarkeit und die Provenance der Ergebnisse sichergestellt werden.

Diskussion

Sollten die Modellierungsgruppen ihren Workflow derart gestalten, dass der Modelloutput während der Laufzeit der Experimente bereits "stückweise" cmorisiert wird (https://redmine.dkrz.de/attachments/download/2828/TP1_workflow.png), ist die Verwendung des ESMValTools im Rahmen des Quicklooksystems ohne weitere Anpassungen möglich. Im Falle der Verwendung der Modellrohdaten für das Monitoring mit dem ESMValTool wurde der Konsens erzielt, dass in Zusammenarbeit mit den Modellierern eine Schnittstelle für das jeweilige Modell geschaffen werden muss. Als erster Schritt wären dem DLR zunächst Testdaten bereitzustellen. Dazu wurde bereits am 16.8.2017 eine Mail an die Modellierer aller an CMIP6 beteiligten Institute geschickt.

Eine Beschreibung des Quicklook-Workflows wird demnächst in das Projekt-Redmine gestellt.

Vortrag: Thomas Schartner (FUB)
'Vorstellung des CMIP6-DICAD Portals Freva'

Die aktuelle Internetpräsenz von Freva wird vorgestellt. Miklip-Erfahrungen sind in die Entwicklung eingeflossen. Jeder auf Mistral registrierte User kann sich einloggen. Zukünftig sollen sich Externe auch einloggen können. FreVa ist auch als Modul ladbar und verfügt über Autovervollständigung.

Der Result-Browser und verfügbare Search-Facets wurden vorgestellt.

Aktuell sind die Graphiken mit Wasserzeichen versehen, um Missbrauch vorzubeugen. Der Zugang zu Graphiken ist offen. Um die Kommentarfunktion zu nutzen, ist es aber erforderlich sich einzuloggen.

Diskussion

Es wurde angemerkt, dass der Zugang zum Portal nicht über das https: Protokoll realisiert ist, also eventuell nicht sicher ist. Sicherheitsmaßnahmen sind aber aktuell nicht geplant. Es ist auch nicht geplant, Diagnostiken bzw. Plots von Diagnostiken in Freva bereitzustellen, die nicht auf Daten basieren, die im ESGF publiziert sind. Dies bezieht sich zum Beispiel auf Daten der 6hrLev-MIP-Tabelle. Daten dieser Tabelle werden von CORDEX für den Antrieb von RCM-Simulationen verlangt. Die ECHAM-Daten dieser Tabelle sollen wg. des mit dem CMIP6-Standard verbundenen Datenvolumens nicht ins ESGF gestellt werden. Damit können die Daten auch nicht in Diagnostiken berücksichtigt werden, die mit Hilfe des ESMValTools für CORDEX als Auswahlhilfe erstellt werden. Das Problem ist nicht die Datenformatierung, sondern das für Spektralmodelle besonders ungünstige Verhältnis der Datenvolumina im CMIP6-Standard oder nativen Format. Allerdings sind die Daten im CMIP5-Archiv nicht NetCDF-komprimiert publiziert worden. Damit könnte das Verhältnis für CMIP6-Daten etwas besser sein. Die Diskussion der Frage wurde wg. der fortgeschrittenen Zeit verschoben. Sie kann in der zu schaffenden CORDEX-Task-Force (s.o.) thematisiert werden.

AP7 Zitierfähigkeit und Dokumentation im CMIP6+-Archiv

Vortrag: Martina Stockhause (DKRZ)

'Vorstellung/Demonstration der Bereitstellung von Datenzitatinformationen'

Diskussion

Der aktuelle Entwicklungsstand des Web-GUI zum Einfüllen von Datenzitatinformationen wurde demonstriert. Nach dem Einfüllen kann ein JSON-File heruntergeladen und bearbeitet werden. Danach kann ein Python-Programm die modifizierten Informationen wieder hochladen. Durch Krankheitsausfall kann aktuell nur mit halben Ressourcen an dem AP gearbeitet werden.

Es wird auf die Wichtigkeit hingewiesen, eine Verknüpfung von Daten mit Autoren und Institut zu erstellen. Die Verknüpfung wird benötigt, um den Daten einen DOI zuzuordnen. Auch die Möglichkeit zur Frühzitation (vor einer endgültigen DOI-Zuweisung) ist ohne die Verknüpfung nicht gegeben. Es konnte nicht geklärt werden, ob ein ausgefülltes CIM-template mit einer PID-ähnlichen Kennung identifiziert werden kann, ohne dass die Datei kopiert werden muss.

Im Gegensatz zu CMIP5 gibt es in CMIP6 keinen Ansprechpartner für das Ausfüllen des CIMs. Es heißt aber, dass die Beschreibung von ESMs, die sich als Nachfolgemodelle von CMIP5-Modellen geoutet haben, mit dem für CMIP5 ausgefüllten CIM-Formblatt fortgesetzt werden kann.

AP8 LZA im WDCC/IPCC-DDC & DataCite Datenpublikation

Krankheitsbedingt ist der Vortrag zu diesem AP ausgefallen.

Informationen: Martina Stockhause (DKRZ)

Durch Krankheitsausfall kann aktuell nur mit halben Ressourcen an dem AP gearbeitet werden. Ende September fand ein Daten-Workshop zwischen IPCC-DDC und IPCC TSUs WG I und II am DKRZ statt. Der Data Pool und unsere Compute Ressourcen sowie das Freva Portal als am DDC des WDCC/DKRZs im Rahmen von DICAD entwickelte Services wurden in Hinblick auf deren Nutzbarkeit für die IPCC Autoren diskutiert. Als Ergebnis des Workshops wurde eine engere direkte Zusammenarbeit zwischen IPCC-DDC und WGs in Datenfragen über die Gruppe 'DDC Support' vereinbart.

Sonstiges

- Die Projektpartner wurden aufgefordert, den vom BMBF ‚angeregten‘ repräsentativen Internetauftritt mit Inhalten zu befüllen.
- Die Projektpartner wurden auch aufgefordert das Projekt-Redmine zu nutzen und zu befüllen.
- Der RZ-Antrag für die mit dem Vorhaben in Zusammenhang stehenden Modellexperimente am DKRZ-Mitarbeiter muss bis zum 31.10. eingereicht werden. Der Antrag soll wieder von den DKRZ-Mitarbeitern gestellt werden. Urlaubsbedingt müssen alle Beiträge bis 24.10. bei Martin und Stephanie eintreffen, sonst muss die Einreichung delegiert werden.
- Mittlerweile hat sich das AWI bereit erklärt, das nächste Projekttreffen (Monat 18) im April oder Mai in Bremerhaven zu organisieren. Ein Doodle wird aufgesetzt.

1. Jahrestreffen CMIP6 BMBF-Projekt, Mi. 11. Oktober 2017, 10:00-16:00
 Raum 23; DKRZ, Bundesstrasse 45a, Hamburg; (+49-40-460094-205)

Name	E-mail	Institut	AP	Unterschrift
Lisa Bock	Lisa.Bock@dlr.de	Deutsches Zentrum für Luft- und Raumfahrt (DLR)	AP6-DICAD	
Björn Brötz	Bjoern.Brötz@dlr.de	"	AP6-DICAD	
Markus Kunze	Markus.Kunze@dlr.de <i>M.Kunze@dlr.de</i>	"	Chemie	
Christopher Kadow	christopher.kadow@met.fu-berlin.de	Freie Universität Berlin (FUB)	AP6-DICAD	
Ingo Kirchner	ingo.kirchner@met.fu-berlin.de	"	"	
Thomas Schatner	Thomas.Schatner@met.fu-berlin.de	"	"	
Heinz-Dieter Hollweg	hollweg@dkrz.de	Deutsches Klimarechenzentrum (DKRZ)	AP4-DICAD	
Stephan Kindermann	kindermann@dkrz.de	"	AP5-DICAD	
Stephanie Legutke	legutke@dkrz.de	"	AP1,2,3-DICAD	
Martin Schupfner	schupfner@dkrz.de	"	AP3-DICAD	
Martina Stockhause	stockhause@dkrz.de	"	AP7-DICAD	
Fabian Wachsmann	wachsmann@dkrz.de	"	AP2-DICAD	
Dmitri Sein	dmitry.sein@awi.de	Deutscher Wetterdienst (DWD)	AP1,3-DICAD	
Bernadette Fritzschi	Bernadette.Fritzschi@awi.de	Alfred-Wegener-Institut (AWI)	AP1,3-DICAD	
Matthias Bitner	matthias.bitner@mpimet.mpg.de	Max-Planck-Institut für Meteorologie (MPI-M)	AP1-DICAD	
Johann Jungclauss	Johann.Jungclauss@mpimet.mpg.de	"	AP1-DICAD	
Karl-Hermann Wieners	Karl-Hermann.Wieners@mpimet.mpg.de	"	AP1-DICAD	
Nikolay Koldunov	Nikolay.Koldunov@awi.de	AWI	AP1,3-DICAD	
Katharina Berger	berger@dkrz.de	DKRZ		