

C3-Grid - ein Werkzeug für die Klimaforschung

U. Ulbrich¹ für das C3-Grid Team
(uwe.ulbrich@met.fu-berlin.de)

mit Beiträgen von: H.Kupfer¹, I.Kirchner¹, T.Brücher²,
M.Stockhause³, B.Fritsch⁴, K.Fieg⁴, C.Kurz⁵

(1) Institut für Meteorologie, Freie Universität Berlin; (2) Institut für Geophysik und Meteorologie an der Universität zu Köln; (3) Max-Planck-Institut für Meteorologie / IFM-Geomar; (4) Alfred-Wegener-Institut Bremerhaven; (5) Deutsches Zentrum für Luft- und Raumfahrt

Problemstellung

Die Klimaforschung ist untrennbar mit der Nutzung verschiedener, teilweise extrem umfangreicher und in unterschiedlichen Standards abgespeicherter Datensätze verbunden. Die Daten aus Beobachtungen oder Modellexperimenten sind, bei kontinuierlich anwachsendem Datenvolumen, auf viele Datenarchive verteilt. Für den einzelnen Wissenschaftler wird es immer schwieriger, sich einen Überblick über die für seine Arbeit wichtigen und verfügbaren Datensätze zu verschaffen. Über die Recherche nach der bloßen Existenz der Daten hinaus müssen Informationen über Datenqualität und Formate mühsam zusammengesucht werden, Zugangsrechte organisiert und schließlich Datenextraktionen und –transfers in erheblichem Umfang bewerkstelligt werden. Alles in allem ist die Nutzung der für die Klimaforschung so wichtigen Daten nicht gut organisiert, obwohl es einen erheblichen Bedarf an diesen Daten gibt. Durch die anlaufenden Planungen, wie man sich an den Klimawandel anpassen kann, werden die Anforderungen an die Datennutzung in Zukunft zudem schnell weiter steigen. Dabei kann nicht davon ausgegangen werden, dass jeder, der die Daten nutzen will, über tiefgehende Erfahrungen mit den Rechnern verfügt, über die Datenzugriff und -verarbeitung erfolgen.

Zentraler Zugang zu verteilten Klimadatenarchiven

An dieser Stelle setzt das Projekt C3-Grid (Collaborative Climate Community Data and Processing Grid) an. Es will dem Wissenschaftler eine Arbeitsumgebung zur Verfügung stellen, mit dem er von zentraler Stelle aus - dem Portal - einen Überblick über den verfügbaren Datenbestand erhält. Er kann also über diesen Zugang Daten suchen, auswählen, bearbeiten und auf einen lokalen Rechner übertragen.

Standardisierte Metadaten

Diese unterschiedlichen Datensätze werden mit Hilfe eines einheitlichen ISO-konformen C3-Grid Metadatenschemas (ISO 19115/19139) beschrieben und mit Hilfe einer Harvesting-Schnittstelle in einen zentralen Metadatenkatalog indiziert. Zur Beschreibung von Variablen innerhalb des Metadatenschemas wird die Climate and Forecast (CF) – Konvention verwendet. So werden die Daten dem C3-Grid bekannt und suchbar gemacht.

Unter Nutzung von Grid-Technologien wird eine neue Infrastruktur aufgebaut, welche die vorhandenen Speicher- und Compute- Ressourcen der unterschiedlichen beteiligten Standorte

miteinander verknüpft. Über die Beschreibung der Daten in einem einheitlichen Metadatenformat wird ein einfacher Zugriff auch auf verschiedene Datenarchive realisiert. Durch die Verwendung von Zertifikaten wird ermöglicht, dass der Nutzer - an einer Stelle angemeldet - Zugang zu Datenarchiven erhält, ohne sich dort einer separaten Anmeldeprozedur zu unterziehen.

Datenaufbereitung und Transfer

Für die Arbeitsschritte nach dem Suchen und Auffinden der Daten bietet das C3-Grid standardisierte Arbeitsabläufe („Workflows“) von unterschiedlicher Komplexität an. Diese organisieren den Transfer der gewählten Daten einen lokalen Rechner, oder die Abarbeitung weiterer (standardisierte) Arbeitsschritte. Im letzteren Fall werden im C3-Grid verfügbare Speicher- und Rechenressourcen in einer optimierten Weise genutzt.

Die Datenaufbereitung (Preprocessing) wird als erster Schritt vom Datenanbieter übernommen. Sie basiert auf der Datenauswahl des Nutzers mit Hilfe des Portals. Hier wird auf Grundlage der Nutzeranfrage ein Teil-Datensatz „ausgeschnitten“ (Geographische Region, meteorologischer Parameter (CF-Variablen), vertikale oder zeitliche Ausschnitte). Des Weiteren werden Formatkonvertierungen sowie zusätzliche Datenanbieter-spezifische Optionen wie z.B. eine zeitliche Mittelung prozessiert. Für die so erzeugten abgeleiteten Datensätze werden automatisch Metadaten erstellt, und zwar auf Grundlage der Metadaten zu den Ausgangs-Datensätzen. Werden diese Metadaten dann dem C3-Grid bekannt gemacht, können sie als Zwischenergebnisse über das Portal gefunden und für weitere Anfragen genutzt werden.

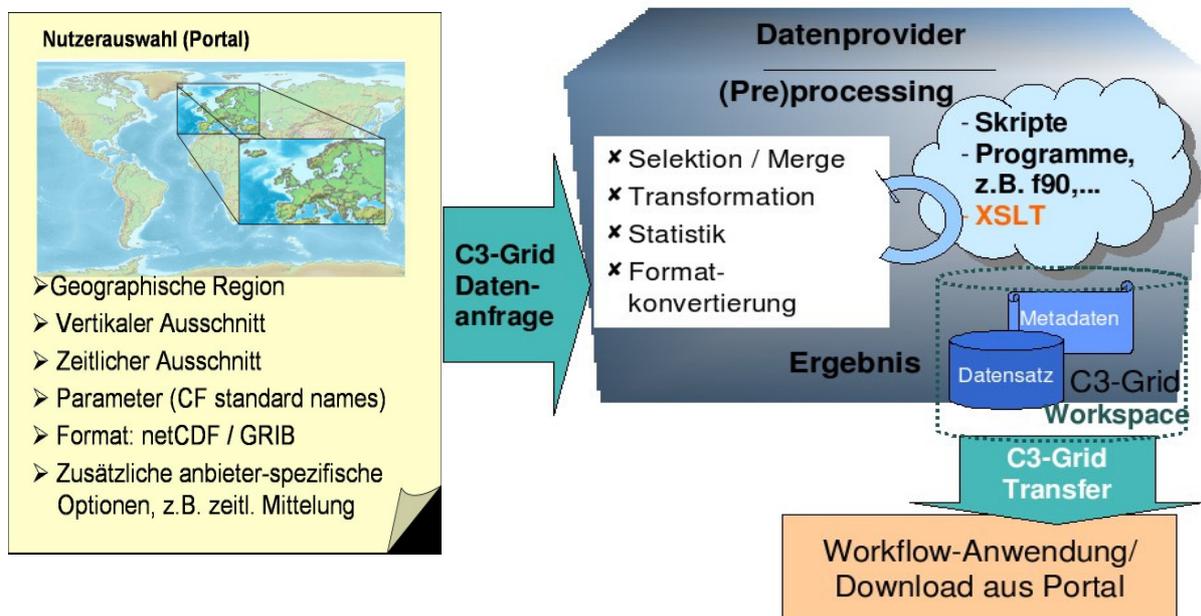


Abb.1: Schema zur Datenaufbereitung

Der Datentransfer nach dem ersten Arbeitsschritt kann sich somit auf die Zwischen- und Endergebnisse beschränken. Er wird durch den Daten-Management-Service (DMS) realisiert und durch den Scheduler gesteuert. Die Grid-Dienste übernehmen die Verarbeitung der vom Nutzer gewählten „Workflows“: Das Scheduling entscheidet nach den aktuell verfügbaren Ressourcen, bei welchem Ressourcenanbieter welcher Teil des Workflows ausgeführt wird.

Datenanalyse und Diagnose

Bei der Analyse von Klimadaten werden häufig bestimmte, vordefinierte Werkzeuge (z.B. Diagnose-Routinen zur Berechnung bestimmter Kennzahlen für den Zustand der Atmosphäre) angewendet. Will man ohne das Grid diese Werkzeuge auf einen Datensatz mit spezifischer Abspeicherung anwenden, oder auf einer anderen Rechnerumgebung implementieren, müssen erhebliche Anpassungsarbeiten vorgenommen werden. Die Einbettung von Standard-Diagnose-Routinen ins C3-Grid (durch sogenannte „Service-Provider“) erspart dem Wissenschaftler den entsprechenden Aufwand. Informationen über Datenformate erhalten die entsprechenden Programme aus den Metadaten. Die Implementierung auf verschiedenen Rechnern im Grid erlaubt es, die vorhandenen Ressourcen optimal zu nutzen. Die Nutzung von Zwischen- oder Endergebnissen für weitere wissenschaftliche Fragestellungen verhindert eine unnötige Belastung der Ressourcen.

Als erste Routinen wurden im C3-Grid Diagnose-Tools zur Berechnung der „Stormtracks“ (synoptisch-skalige Variabilität von Geopotenzial bzw. Bodendruck) and Qflux (Feuchtflüsse in verschiedenen vertikalen Bereichen), sowie die 'chemische Wettervorhersage' zur Flugroutenbestimmung bei der Durchführung luftchemischer Messkampagnen implementiert. Routinen für die Baroklinität und für CAPE sind in Vorbereitung.

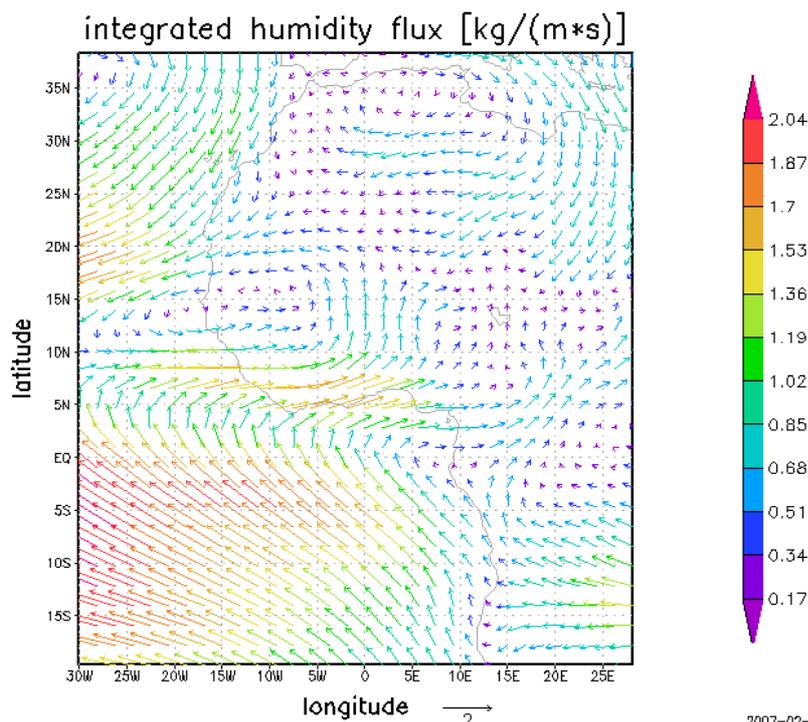


Abb.2.: Graphische Darstellung des mittleren, vertikal integrierten Feuchtetransports [kg / (m*s)] simuliert durch ECHAM5, 20C.1, Juni bis September, für ein Beispieljahr (1999) zwischen 1000 und 775 hPa.

Feuchteflüsse (Qflux)

Das Diagnose-Werkzeug „qflux“ berechnet den integrierten Feuchtefluß der Atmosphäre. Es benötigt neben dem horizontalen Windfeld den Feuchtegehalt (egal, ob relative oder spezifische Feuchte) auf Druckniveaus der Atmosphäre für die zu untersuchende geographische Region: Die Berechnung des integrierten Feuchtetransports beruht auf der vertikalen Integration der 2-dimensionalen Feuchteadvektion. So werden zunächst die Feuchtetransporte pro Windkomponente errechnet und schließlich über den durch den Nutzer bestimmten vertikalen Bereich integriert. Diesen Arbeitsschritten vorangeschaltet sind die Schritte der Datenaufbereitung (Preprocessing). Somit wird beim Aufruf von „qflux“ ein Workflow des Datenproviders (Preprocessing) und ein Workflow eines Serviceproviders (z.B. Feuchtetransport-Analyse) gestartet.

In jedem Fall wird der Nutzer des „qflux“ Workflows die zu verwendenden Daten sowie die räumliche und zeitliche Begrenzung des Untersuchungsgebiets vornehmen. In einer späteren Version sollen zusätzlich diverse Optimierungen (Einbindung von Wind- und Feuchte am Boden in die Berechnung) bzw. physikalische Formulierungen über das Portal wählbar sein.

Stormtracks

Die Berechnung der Stormtracks beruht auf der Berechnung der Standardabweichung der 500hPa Geopotentiellen Höhe, wobei der Varianzberechnung eine Bandpass-Filterung mit Durchlass auf der synoptischen Zeitskala (2.5-6 Tage) vorangestellt ist. Die Gebiete mit einem hohen Wert der Standardabweichung stimmen mit Regionen starker synoptischer Aktivität überein, charakterisiert durch die Abfolge von tiefem und hohem Druck/Geopotenzial. Die zugrunde liegenden Daten müssen eine zumindest 24-stündige zeitliche Auflösung haben. Die Auswahl der Filtercharakteristika passend zur zeitlichen Auflösung der Daten ist vorgesehen.

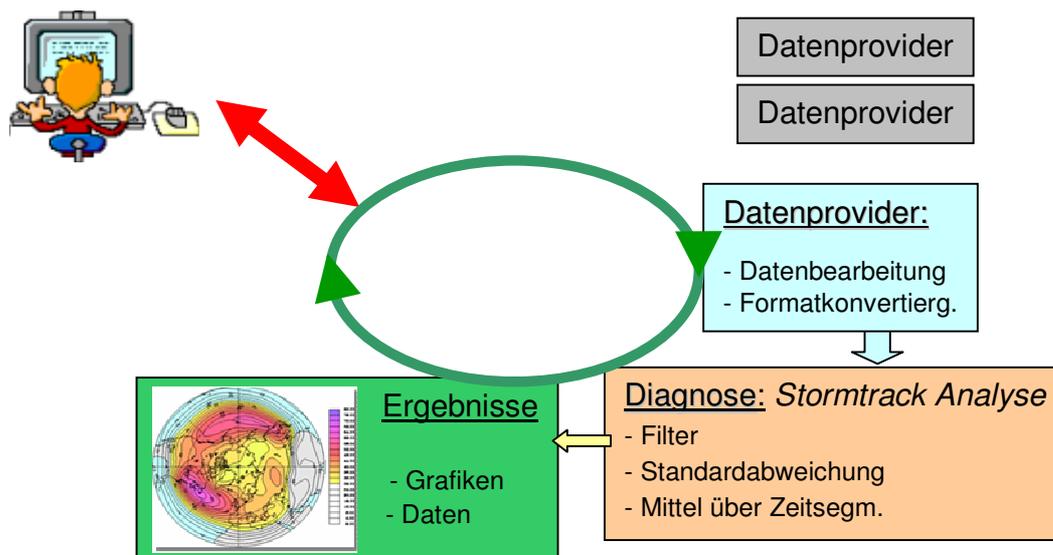


Abb. 3: Ablauf des Stormtrack-Workflows in der Generation 0 des C3-Grids im Kontext mit dem Präprozessing und dem Portal

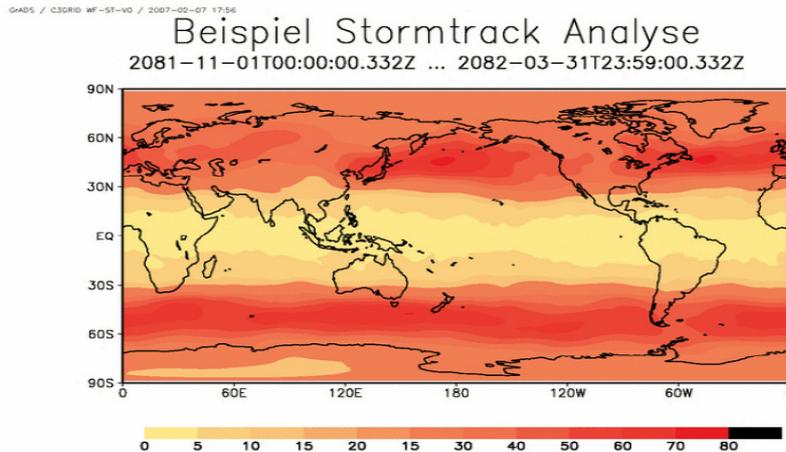


Abb.4: Graphische Darstellung der Stormtrackanalyse (geopotential_height 500) für ECHAM5, Lauf 1, A1B Szenario für den Winter 2081/2082

Chemische Wettervorhersage

Der "Chemical Weather Forecast" Workflow dient zur Begleitung und Flugroutenplanung während Messkampagnen. Es wird täglich die Verteilung chemischer Spurengase in der Atmosphäre mit dem ECHAM5/MESSEY CCM prognostiziert. Der Workflow kann dabei auf verschiedene HPC Ressourcen zugreifen. Der Nutzer kann über ein Webinterface aus den Modellergebnissen Variablen extrahieren oder Regionen ausschneiden, die dann graphisch dargestellt werden. Er muss dazu selber nicht auf die Modellrohdaten zugreifen.

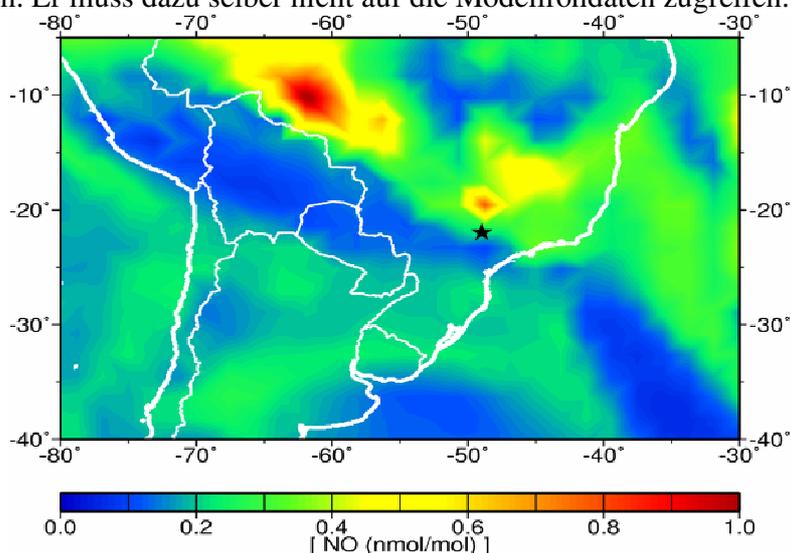


Abb. 5: 3-Tage-Vorhersage von Stickstoffoxiden. Mischungsverhältnis bei 200 hPa über Südamerika. Der Stern zeigt den Ort der Messstelle an.

Visualisierung

Die Darstellung der Ergebnisse ist im Moment noch fest mit den Tools gekoppelt. Ein separates Visualisierungs-Modul für eigene Farb- oder Projektionseinstellungen ist in Planung.

Ausblick

In Zukunft soll es möglich sein, Bearbeitungsschritte nutzerspezifisch zu formulieren und flexibel zu arrangieren. Als Teilziel des Projektes wird im C3-Grid eine Umgebung eingerichtet, die es dem Anwender ermöglicht, einen eigenen Workflow zu formen. Das C3-Grid unterstützt den Anwender dabei, standardisiert Daten zu bearbeiten, erleichtert dabei Routinearbeiten und hilft so, die Arbeit effektiver zu gestalten.

Einbindung zukünftiger Nutzer in das Projekt

Im Juni 2007 fand ein Nutzerworkshop statt, bei dem zukünftige Anwender des C3-Grid zur Teilnahme und Diskussion eingeladen waren. Dabei konnten zahlreiche Anregungen mit in die Projektplanung aufgenommen werden. In einem weiteren Schritt werden nun Wissenschaftler, die Daten mit Bezug zur Klimaforschung nutzen oder produzieren eingeladen, die erste Generation des C3-Grid zu testen und zu kommentieren (<http://www.c3-grid.de/portal>). Zusätzlich stehen wir natürlich für Anregungen weiterhin offen.

Referenzen:

Siehe www.c3-grid.de , Abschnitt: Ergebnisse