



Collaborative Climate Community Data and Processing Grid (C3-Grid)

Workflow-Dokumentation *qflux*

Arbeitspakete:	Workflows (übergreifend AP1,3,4,5,6,7)
Autoren:	Diagnose Cluster (FUB, ZAIK, IGMK)
Version:	Generation 1 (Stand: Jul.2008)
Veröffentlichungsdatum:	Juli 2008
Koordination:	Universität zu Köln (ZAIK, IGMK), FUB
Partner:	AWI, ZIB, DKRZ, M&D, MPI/IFM-Geomar, Uni-Dortmund
Ansprechpartner:	T. Brücher, V. Achter
Email:	bruecher@uni-koeln.de

Inhalt

Motivation

1 Einleitung

2 Einbindung des Diagnose-Tools „*qflux*“ in das C3-Grid

3 Modularisierung von „*qflux*“

4 Das Tool: *qflux*

4.1 Methode

4.2 Variablen

4.3 Optionale Steuerung

5 Entwicklungsausblick

Motivation

Durch die Erzeugung und Haltung von unterschiedlichen Datensätzen an ebenso verschiedenen Rechenzentren nehmen technische Aufgaben des Datenmanagements sehr viel Zeit des Wissenschaftlers in Anspruch. Dieses ist unter anderem durch ein unterschiedliches Datenhaltungskonzept, unterschiedliche Datenformate und ein uneinheitliches Rechtesystem bedingt. Damit dem Wissenschaftler mehr Zeit für die wesentlichen Aufgaben seines Spezialgebiets zur Verfügung stehen, ist eine einheitliche Schnittstelle, die die technischen Herausforderungen übernimmt, wünschenswert.

C3-Grid schafft, über diese einheitliche Schnittstelle zu den Datensätzen hinaus, die Kopplung von Diagnosetools an die verteilt liegenden Datensätze, so dass automatisiert Standardfragen / -untersuchungen schnell auf unterschiedlichste Beobachtungs- oder Modelldaten angewendet werden können.

1 Einleitung

Im C3-Grid werden diverse Tools zur Verfügung gestellt, die meteorologische Fragestellungen beantworten. Dieses sind unter anderem die Programme „*qflux*“ und „*cape*“, die sich dem Feuchtehaushalt der Atmosphäre widmen. „*qflux*“ berechnet den integrierten Feuchtefluß der Atmosphäre, während „*cape*“ (convective available potential energy) eine Maßzahl zur Stabilität der Atmosphäre berechnet.

Beide Tools benötigen neben der Temperatur und dem horizontalen Windfeld den Feuchtegehalt (egal, ob relative oder spezifische Feuchte) auf Druckniveaus der Atmosphäre für die zu untersuchende geographische Region.

Sollen mit diesen Diagnosetools klimatologische Fragestellungen beantwortet werden, so steigt die Anzahl der Einzelschritte enorm an, da auf Grund unterschiedlicher Datenanbieter, -formate und -quellen sowohl der Zugang, das Diagnosetool als auch die graphische Aufbereitung immer neu angepasst werden muss. Weiterhin stellen die zur Verfügung stehenden Ressourcen (z. B. Speicherplatz) oft Limitierungen dar.

2 Einbindung des Diagnose-Tools „*qflux*“ in das C3-Grid

Der Workflow besteht aus einer Kombination aus dem Workflow des Datenproviders (Preprocessing) und dem Workflow des Serviceproviders (Feuchtetransport-Analyse) und lässt sich auf verschiedene Datensätze anwenden.

Im trivialen Fall umfasst der Workflow nur die Selektion einer Datenquelle und der Raum-Zeit Parametern. In einer späteren Version wird zusätzlich die Auswahl diverser Optimierungen (Einbindung von Bodendatensätzen zur genaueren Diagnose) bzw. physikalischer Formulierungen (unterschiedliche Formulierungen in der Berechnung von *cape*) wählbar sein. Diese Selektionen werden im Portal durchgeführt, die anschließende Datenbereitstellung wird durch den Datenprovider vorgenommen und beim Serviceprovider werden die diagnostischen Parameter berechnet. Abschließend stehen die Ergebnisse im Portal zum Download bereit.

Der Datentransfer zwischen den einzelnen Partnern wird durch den Daten Management Service (DMS) realisiert und durch den Scheduler gesteuert¹. In der aktuellen Version (Generation 1) sind diese Schnittstellen nicht mehr fest verdrahtet, sondern der WF liegt in einer modularisierten, jedoch fest vorgeschriebenen Prozess-Kette vor. Das bedeutet, dass die wesentlichen, inhaltlich trennbaren Abschnitte separat voneinander durch den Scheduler gesteuert werden. Die prinzipielle Teilung dieser Prozesskette sieht wie folgt aus:

- a) Datenaufbereitung / Datenselektion
- b) Berechnung des / der Diagnoseparameter
- c) Statistische weiterverarbeitung der Ergebnisdaten
- d) Visualisierung der Ergebnisse

3 Modularisierung von Workflows:

Bei dem hier verwendeten Ansatz zur Modularisierung von Workflows sind die separaten Abschnitte so zu formulieren, dass kein „Wissen“ der Module untereinander vorausgesetzt werden darf. Das bedeutet, dass der Namensraum zum einen durch die global definierten Umgebungsvariablen bekannt ist, zum anderen wird jedem Module der komplette Satz an notwendigen Parametern übergeben. Via Ansteuerung eines Parsers zur Übergabe der Parameter an die Prozesskette werden alle Module nacheinander – durch den Scheduler auf deren Abhängigkeit geprüft und deren Ausführung kontrolliert – ausgeführt. Der Ablauf ist linear.

¹ Da die Vorgehensweise analog zum Stormtrack-Workflow läuft, wird hier für eine graphische Darstellung des Ablaufs verzichtet und auf Abbildung 1 dessen Workflow-Dokumentation verwiesen.

4 Das Tool: *qflux*

Das Diagnosetool „*qflux*“ ist in einer PERL bzw. SHELL Umgebung eingebettet, die sich um die Analyse der Daten und Metadaten kümmert, den eigentlichen Programmaufruf samt Kommandozeilenparametern startet und ein Postprocessing, bestehend aus einer einfachen statistischen Analyse. Für die abschließende graphische Aufbereitung der Ergebnisse wird dynamisch ein Skript erzeugt, das mit Hilfe von GrADS ein Bild im gif-Format erzeugt.

Die visuellen und numerischen Ergebnisdaten werden vom DMS an einen Webserver gesendet, auf den das Portal zugreift und die Ergebnisse dem Nutzer sichtbar macht. Die Ergebnisse können von dort übers Web auf den lokalen Klienten des Nutzers transferiert werden.

4.1 Methode

Die Berechnung des integrierten Feuchtetransports beruht auf der vertikalen Integration der 2 dimensional Feuchteadvektion. So werden zunächst die Feuchtetransporte pro Windkomponente errechnet und schließlich über den durch den Nutzer bestimmten vertikalen Bereich integriert. Insbesondere die Kombination von Windfeld und transportierter Feuchte bilden einen wesentlichen Bestandteil der Untersuchungen des Feuchtegehalts der Atmosphäre, da hierdurch Senken und Quellen ermittelt werden können.

4.2 Variablen

Parameter:	Temperatur, relative oder spezifische Feuchte, 2d-Windfeld auf Druck-Niveaus, optional die Bodenparameter Temperatur, Taupunkt, Bodendruck, Wind in 10m Höhe,
Geographische Region:	vom Nutzer frei wählbar (vertikal und horizontal)
Zeitraum:	vom Nutzer frei wählbar
Zeitliche Auflösung:	vom Nutzer frei wählbar

4.3 Optionale Steuerung

- Bestimmung des In- und Output Datenformats (grb oder NetCDF)
- Zuschaltung von Bodenparametern, damit insbesondere die Analysen im unteren Bereich der Atmosphäre optimiert / korrigiert werden können

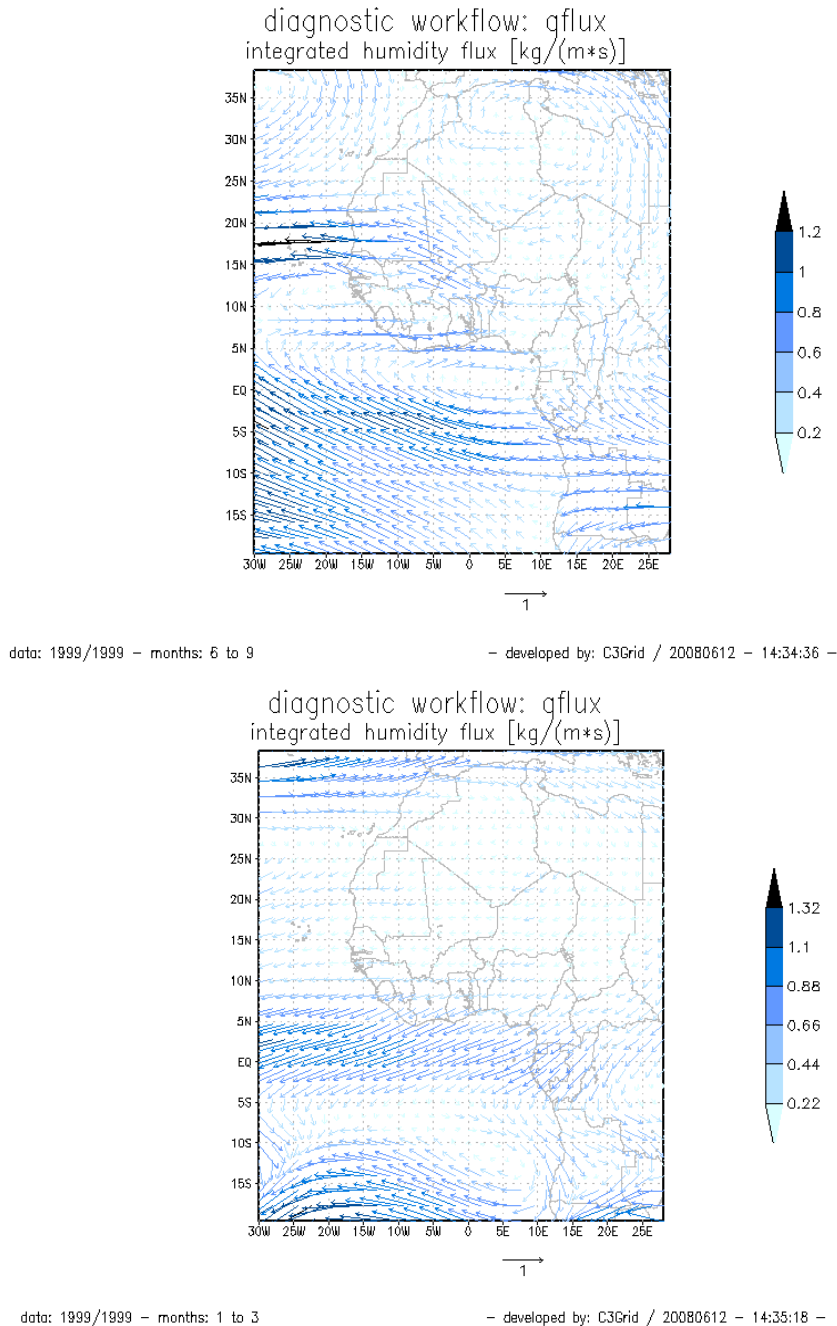


Abb.1: Graphische Darstellung des mittleren, integrierten (1000 hPa bis 750 hPa) Feuchtetransports [kg / (m*s)] simuliert durch ECHAM5, 20C.1 für das Modelljahr 1999. Die Feuchtperiode Westafrikas (Sommermonsun, Juni bis September; oben) und die Trockenzeit mit den heißen Harmatan-Winden aus dem Landesinneren (Januar bis März; unten) sind an den Richtungen der Strömungsrichtung gut zu erkennen.

5 Entwicklungsausblick

Folgende Erweiterungen / Ergänzungen bieten sich an, den Funktionsumfang (aus Meteorologischer Sicht) zu erweitern und mit dieser erhöhten Funktionalität den Benutzerkreis des Diagnosetools zu vergrößern.

- Zuschaltung von Bodenparametern, damit insbesondere die Analysen im unteren Bereich der Atmosphäre optimiert / korrigiert werden können
- Erweiterung für Daten auf Modell-Niveau

Insbesondere die Ansteuerung optionaler Parameter bietet für C3Grid inhaltlich eine große Erweiterung, WFs zukünftig komplexer zu gestalten. Diese Funktionalität erwartet, dass Abhängigkeiten definiert werden können, die im Portal aufgelöst werden können. So sollte zum Beispiel dem Nutzer angezeigt werden, daß eine Erweiterung der Standardberechnung in Abhängigkeit des selektierten Datensatzes möglich ist.

Mit der Erweiterung in C3Grid um eine frei kombinierbare Prozesskette auf Modulbasis ist sichergestellt, daß „qflux“ auf Grund seiner Modularisierung mit weiteren Anwendungen kombinierbar ist. Jedoch ist zu überprüfen, ob die aktuelle Version der vollen Flexibilität von zukünftigen Generationen gerecht wird, oder weitere technische Arbeiten an der Ansteuerung des Diagnosetools notwendig sind.