

地球科学データの高度利用と流通のための基盤開発

堀之内武(北大・地球環境科学研究院) 西澤誠也(神戸大・惑星科学研究セ)、 渡辺知恵美(お茶大・理・情報)、 石渡正樹、小高正嗣(北大)、林祥介(神戸 大)



講演アウトライン

- 背景
 - 業界事情, 地球流体電脳倶楽部
- 地球流体のデータ解析・可視化基盤ライブラリ開発
- Webベースのデータアーカイブ・提供・解析・可 視化・発信システムの開発

歴史的背景

- 大気・海洋の研究と計算機
 - 数値予報, (のち)気候予測等 黎明期からの大口 ユーザ(大規模計算ほか, 研究のあらゆる局面で利用)
- データの嵐(情報爆発)
 - 多種多様, 多量なデータを使う必要性の増大:シミュレーション・諸観測(点, 線, 面,..; 地上/衛星), 学際協力ニーズ(環境問題)...やることは年々高度化
- 各研究者はなんとかやってきた(やれるひとが生き 残ってきた)
 - しかし膨大な車輪の再発明
 - DLして手元でファイルベースで…年々非効率に
 - ⇒ 要インフラ整備(誰かがやらねば/良品以外無意味だが)

地球流体電脳倶楽部



- 80年代末頃より. 創設メンバー: 林祥介(当課題メンバー), 酒井敏, 塩谷雅人, etc
- 気象学・海洋学etcの教育と研究のための計算情報資源の開発と蓄積を目指して協力する有志の集まり(ボランティア精神が基盤)
 - データ解析・可視化支援 本課 題はこの枠
 - 数値モデル群
 - 知識・教育資源群



データ解析と可視化 (Ruby以前)

- FORTRAN / DCL(電脳倶楽部描画ライブラリー 国内では高シェア)
- IDL, Matlab, コマンドラインシステムいろいろ...

...ライブラリの再利用性を高めるのは難しい

地球流体電脳Rubyプロジェクト

- 1999年より. 地球流体電脳倶楽部のプロジェクト. (参加自由. 宇宙物理や生物のひとも)
- <u>データ中心</u>にソフトの再利用性を考えたらオブジェクト指向になる.
- その他諸条件を考えるとRubyが適
 - 開発者=研究者(すばやく楽しく楽に開発, デバグ簡単), 完成ソフトよりライブラリ(研究は世界初を目指すもの), 対話利用(試行錯誤), FORTRAN / Cの資産継承, パフォーマンスも

電脳Rubyプロジェクト中心プロダクト

GPhys

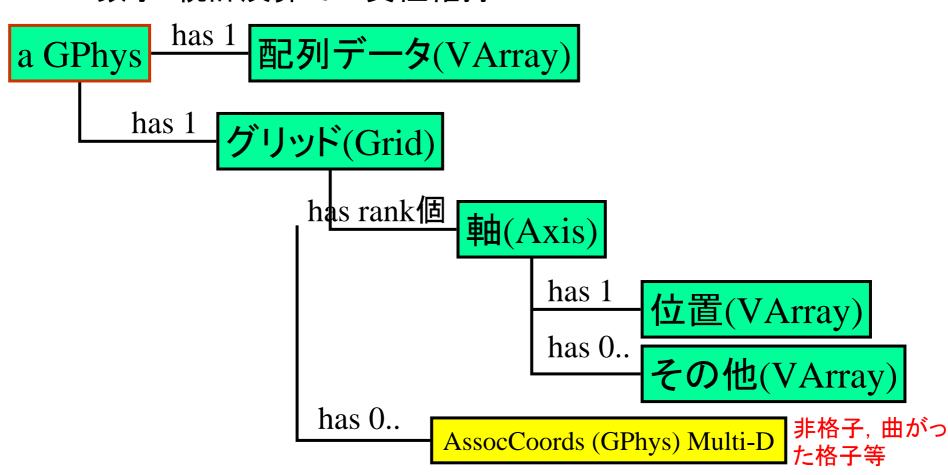
- 2003年より. 堀之内他.
- 我々があつかうデータを突き詰めて抽象化した、 汎用なデータ解析プログラミングの基盤(Ruby のクラスライブラリ)
 - 「多次元空間における離散的な物理量」 多様なデータ形式(実行時メモリ上も), 次元性, サンプリングをカバー. 遅延処理, 処理自動分割などで実用性高.
- 情報爆発での新開発(後述)の基盤
- 利用:多くの研究者が使ってるほか,最近では企業化されたシステムへの組込みも(気象庁,地球観測衛星関連)

GPhys: Gridded Physcal quantity

A01-14

inf@plosion

- 座標データと配列データからなる。(メタデータも)
- 多様な形態のデータを統一的に扱う(各種外部ファイル,実 行時メモリ上,他のGPhysのサブセットや複数統合)
- 数学・統計演算で一貫性維持



GPhysと類似のライブラリ

OPeNDAP (米): httpベースの遠隔データアクセスライブラリ. 提供データファイルを抽象化(地球流体データ中心)

〇:ネットワーク透過性

× : ユーザに(リクエストに応じてカットした)「食材」を届けるところまで. ユーザが手を加えた食材は管轄外 - 玄関でさよならされても...

GPhysは調理を最後まで面倒見る(食材と料理は連続、言語特定は仕方なし、我々の料理は言語でするのだから、)

- ここまでは、理学の研究者のみで.(最近の内容は 当課題の成果でもあるが)
- 本特定領域より情報科学者と協力(協働レベル, 交流レベル…発表会等で多数のコメントを頂いたのも糧になってます.), データ科学WS

情報爆発での我々の(直接の)ミッション

 ネットワークを通じてデータ(&メタデータ)を提供し、 その上でデータ解析も可能な基盤ツールを作る。

(さらに得られる知見の動的な文書化と集積サポートも)

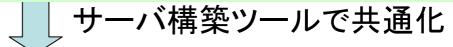
地球科学データに関する時代的要請

- データ公開(しかも検索・可視化等の機能付で)
 - 様々な機関・研究者がデータを所蔵
 - でも、地球科学者には敷居が高い。
 - → 一元的に集める / ITで敷居を下げる
- ネット上のデータも手元データも一緒に扱いたい
 - 研究の一線: 多種多量のデータを組合せが重要
 - いちいちDLし、ファイル形式を解釈…やってられない
 - 一方, 遠隔サーバだけで使えるサービスだと手元の データに使えず不便
- プロだけでなく「素人」も使えるものを
 - 環境問題には学際的取り組みが必要

Webベースの地球流体データサーバ inf plosion

研究機関それぞれが開発

(応用性に乏しい. それぞれ使い方が違う)



Webブラウザ用GUIサーバ (LASなど)

(クイックルックしかできない;手元のデータには使いづらい)



本研究

- ✓ データ公開から手元(デスクトップ)のデータの整理、解析、可視化まで
- ✓ クイックルックだけでなくデータ解析全般カバー (GPhys単独では台所に食材を用意してから ⇒ 食材を探し買ってくる/店を開くところまでカバー)
- ✓ 知見情報の集積も. 共同研究・学際研究サポート

研究成果はこの中に統合的に実装: A01-14



Gfdnavi: Geophysical fluid data navigator オープンソースで公開

http://www.gfd-dennou.org/library/davis/gfdnavi/

データ整理(メタデータ抽出・検索), データ分析, 可視化までの一連の作業を支援

Gfdnaviサーバをローカルホストで立ち上げる

スタンドアローンで利用できる統合的なデータ分析 ソフトウェア

GfdnaviサーバをWebサーバとして立ち上げる

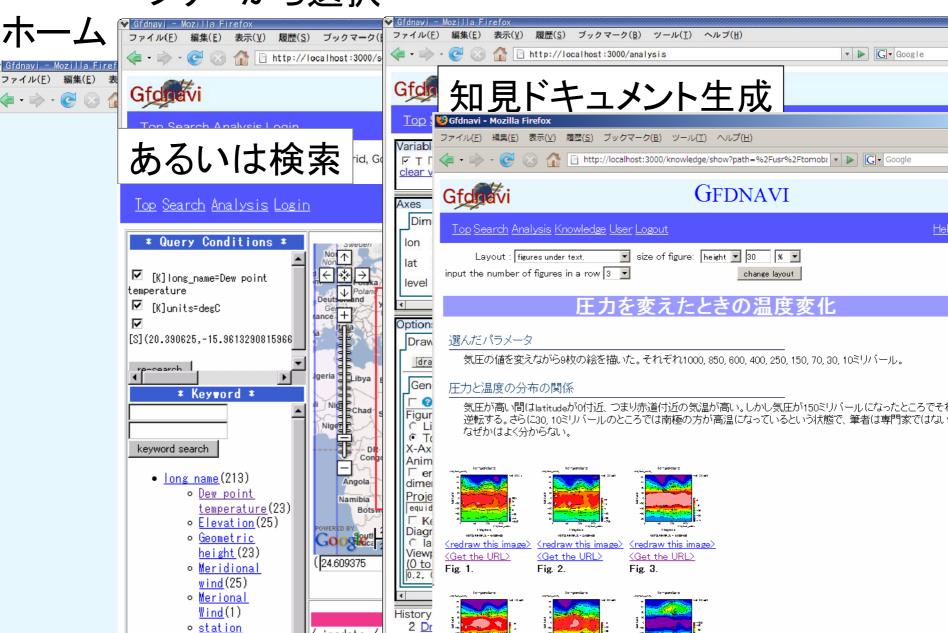
スタンドアローンで利用している環境と全く同じ環境を公開することができる



ツリーから選択

inf@plosion データ解析・可視化

A01-14



/ imadata /

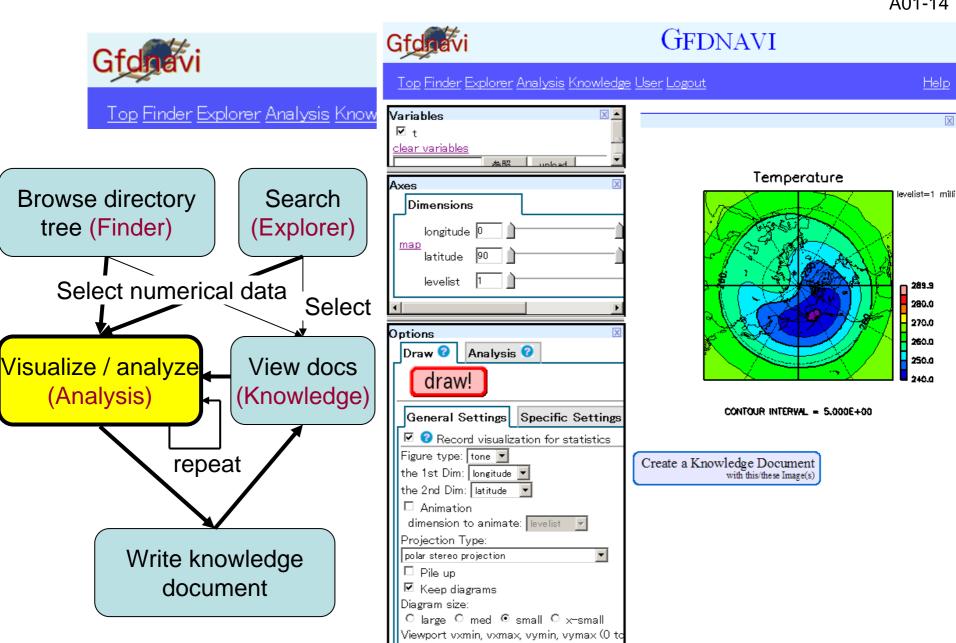
- GSMAP (マイクロ波による降雨観測を行う諸衛 星の統合データ)データサーバ
- SMILES (サブミリ波による大気環境(オゾン, ェアロ ゾル, etc.etc) 観測衛星) データサーバ
- 気象災害判断支援システム
- グローバル大気観測データベース

Gfdnavi利用で得られる知見の 文書化&DB化について

- Gfdnaviで行った可視化等をもとに文書を作成
- → 可視化再現スクリプト、元データへのリンクとと もにDB - 応用性大

利用例:共同研究プラットフォーム(共同作業や意思 疎通の補助,文書アーカイブ)、データ公開サイトにお ける情報発信(PR)、実演マニュアル、研究ノー トなど多様

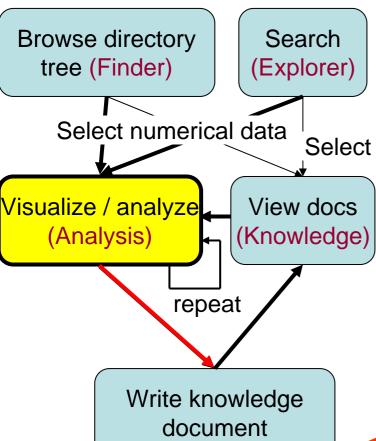


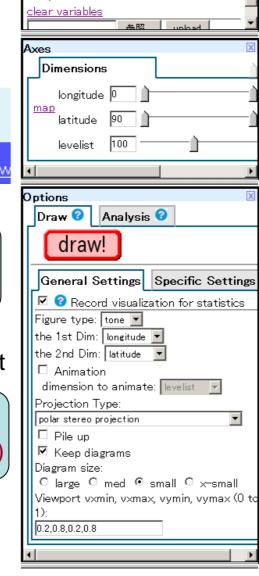


.

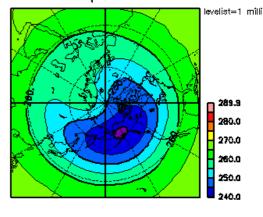
描画





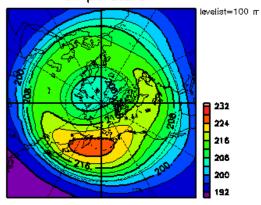


Temperature



CONTOUR INTERVAL = 5.000E+00

Temperature

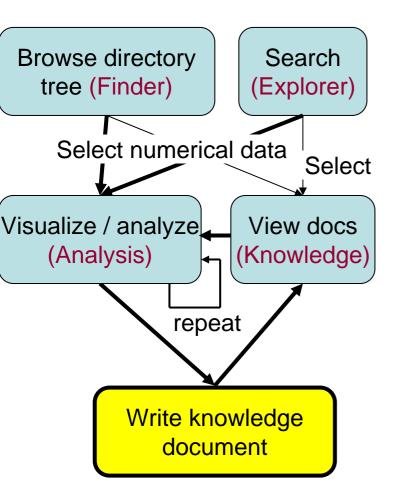


CONTOUR INTERVAL = 4.000E+00

Create a Knowledge Document with this/these Image(s)

知見文書作成





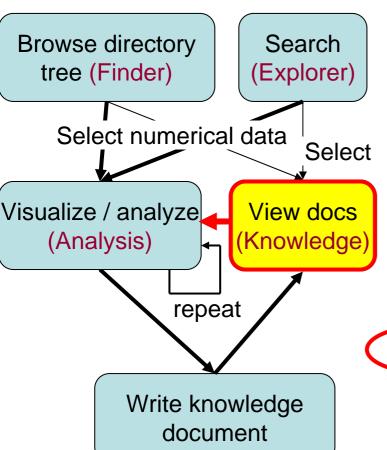
Save Images and Create a New Knowledge

Title: ECMWF Reanalysis January Climatology Author: T Horinou	chi
Textbody:	
((<figure 1="">)) shows the climatological temperature at 1 hPa using the ECMWF Reanalysis (ERA40) in the northern hemisphere. It shows that the climatological polar vortex is shifted to the Pacific side.</figure>	
((<figure 2="">)) shows is the same as Fig.1 but for 100 hPa. It shows that the westerly jet is strong in the Pacific strom track.</figure>	
Path: /usr/root/knowledge/tmp/eraT knlge ex) /usr/root/knowledge/folder1/folder2/writing.knlge	
visible to: visible to: visib	
Choose a default layout : figures under text.	100 % 🔽
input the number of figures in a row 1	
Figure 1	
Caption:	delete
ERA Jan T at 1 hPa	Temperature
	Mark of the second of the seco
File Name:	3104
image001	C)-10(F PTSML = 5,003+C)
	view this image in the original size
Figure 2	
Caption:	delete
ERA Jan T at 100 hPa	Tempercture
File Name:	127
image002	C)-POP HTSML = MODINGS
	view this image in the original size

More Figure
Create







GFDNAVI

Top Finder Explorer Analysis Knowledge User Logout

Layout : figures under text. size of figure: height 100 %
input the number of figures in a row 2
change layout

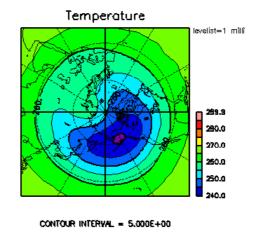
ECMWF Reanalysis January Climatology

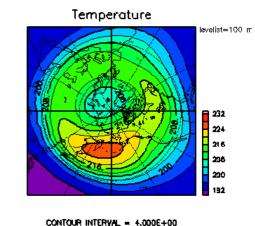
Author: T Horinouchi

Help

<u>Figure 1</u> shows the climatological temperature at 1 hPa using the ECMWF Reanalysis (ERA40) in the northern hemisphere. It shows that the climatological polar vortex is shifted to the Pacific side.

Figure 2 shows is the same as Fig1 but for 100 hPa. It shows that the westerly jet is strong in the Pacific strom track





<u>⟨redraw this image⟩</u> <u>Get the URL⟩</u>

⟨redraw this image⟩ ⟨Get the URL⟩
Fig. 2. ERA Jan T at 100 hPa

Path: /usr/root/knowledge/tmp/eraT.knlge

Edit Back to List

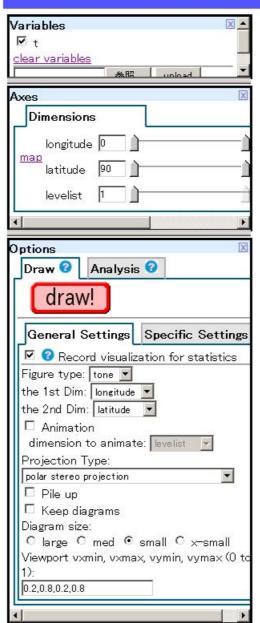
There are no comment on this document.

Write a Coment on this document.

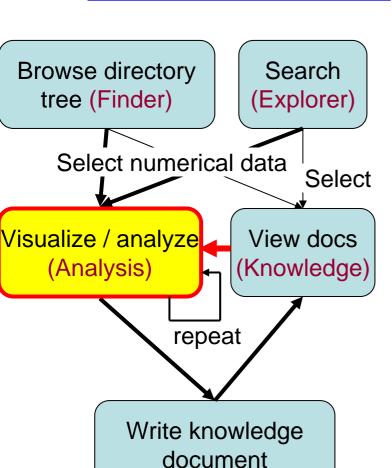
Back to List



Top Finder Explorer Analysis Knowledge User Logout







- 文章に別の文章(コメント)をつけられる.
- データと文書にまたがった検索
- etc

GUI vs API

- GUI: quick-look・試行錯誤に適 (研究初期段階,「素人」に適—学際研究プラットフォーム) (余談:地球流体電脳倶楽部家訓「隣の専門家がわかる情報を提供すべし」⇒いまの場合は「隣の専門化が使えるようデータを提供すべし」)
- しかし、研究中・後期段階では必ずしも便利でない.
- ⇒ GUIとAPIの両方が必要. 両者にギャップのないよう, 多様なプログラム性が必要. (既存の地球流体データサーバはどちらか.)

Gfdnavi は複数の手段によるプログラム性を提供することで、GUIとAPIをつなぐ

- 必要最小限データ+ 再現スクリプトのダウンロード
- GUIで実行するスクリプトの登録
- Webサービス
 - ローカルでのプログラミングにより、サーバー上で解析・可視化することが可能. ただしWebサービスのみでは不十分. ... 他の手段との親和性薄
 - RESTful Webサービスとオブジェクト指向プログラミングの親和性の追及.(論文発表: Nishizawa et al, 2010)
 - 遅延メカニズム&キャッシュによる効率化
 - クライアントライブラリ in Ruby: これで「プロ」にも 市場から厨房まで繋がる. (知見文書で配膳も)

典型的ワークフロー

Gfdnavi GUI上で

データ検索(複数Gfdnavi サーバより)

いろんなデータ解析・可視化をしてみる(試行錯誤)

画像保存/

再現スクリプトを取得し保 存(⇒プログラミングへ)/

知見ドキュメントを書き サーバ上に保存

1つのRubyスクリプトで

Gfdnavi Webサービスを 通じて遠隔データを解 析・可視化

キャプションや本文をつけて知見ドキュメントとして Gfdnaviサーバに保存

(おなじ処理を複数のデータ セットでand/or時間等を変え て多数行い可視化し整理す るなど、プログラムでやりや いことは多い)

Gfdnaviメタサーチエンジン

● 背景

- 科学データは大規模な研究組織だけでなく、各研究者が個々に研究経過や成果として所有・公開
 - ●観測データ,分析結果など
 - ●論文として掲載した結果に対して、その結論 に至るまでの詳細な分析や検討経過など
- 世界中に散在するこれらの情報を集めて検索できるようにしたい!

Gfdnaviによって 蓄積、公開され るデータ

Gfdnaviによるメタサーチェンジンの提供

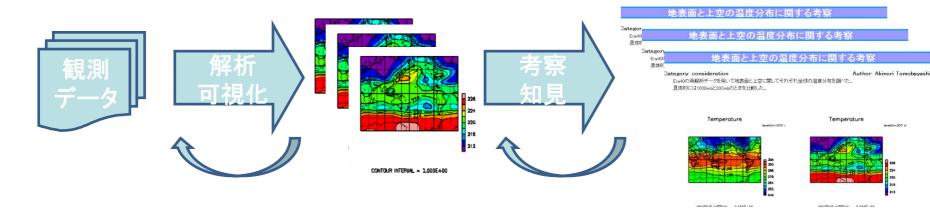
Gfdnaviの統合分析環境との連動

A01-14

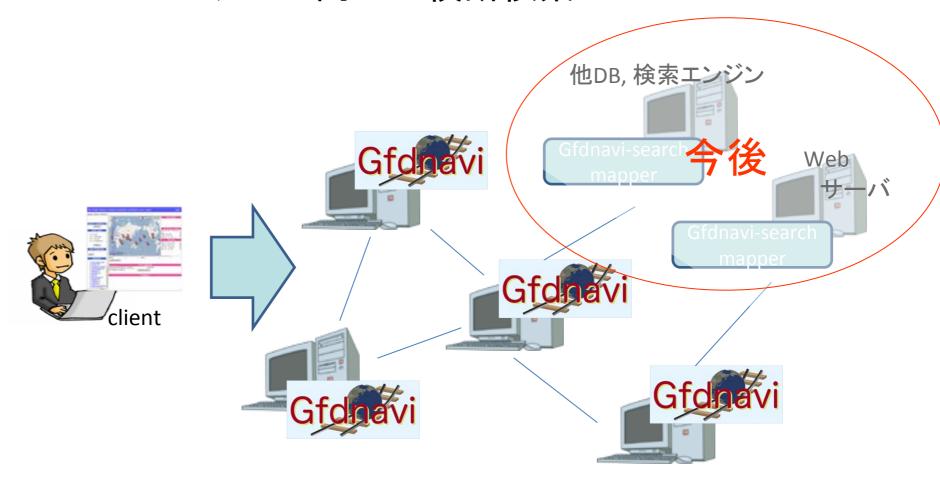
- 関連事例: Sciencenet
 - P2Pによるメタサーチェンジン YaCyを使った科学関係サイト専門 のメタサーチェンジンを提供
 - 2億4000万ページを収集 Webページが検索対象



Gfdnaviと連携したメタサーチェンジンではGfdnavi 利用で逐次生成される解析結果や知見が対象



• Gfdnaviサーバ同士の横断検索

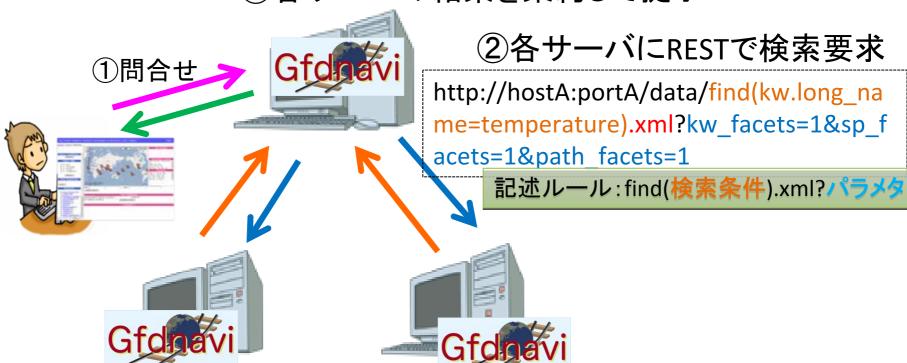


Gfdnavi横断検索画面

から抽出

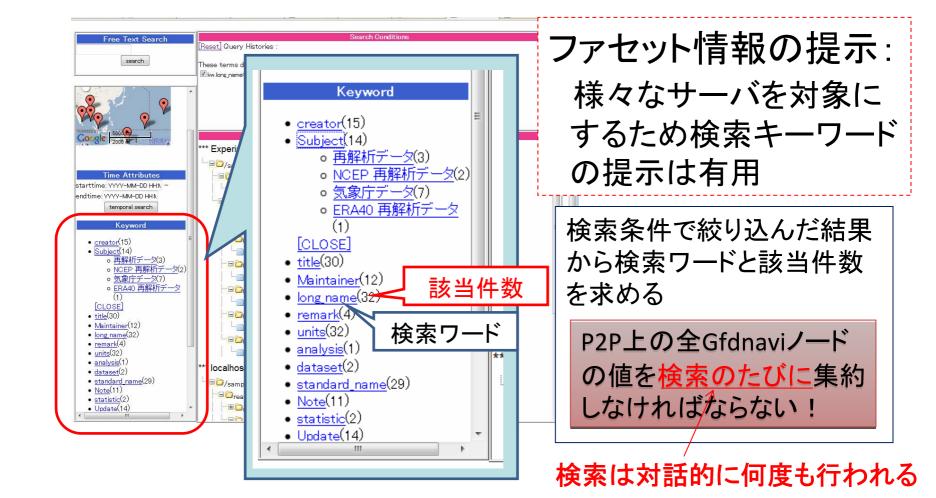


③各サーバの結果を集約して提示



- 成果: RESTを用いたGfdnavi Webサービス提供
 - 諫本, 渡辺, 堀之内, 西澤: DEIM, B2-1 (2009)
 - Nishizawa et al., DIEW 2010 → LNCS論文

• P2P環境でも同じ検索インタフェースを使用



P2P環境におけるファセット検索

A01-14

- 効率的なファセット検索のためのデータ配置方法
 - 検索ワードごとにファセット情報を (できるだけ)一つの検索サーバに配置
 - 問合せ要求はサーバの負担が均等になるよう考慮

Gfdnavi対応 から昇華させ た研究



- C. Watanabe, M. Saito: A Data Assignment for Faceted Search on P2P Environment, PDPTA2010, pp.646-651.
- 東本有加, 渡辺知恵美: P2P環境でのFaceted Searchにおける効率的な問合せの実現に向けて,第151回 データベースシステム研究発表会, Vol.2010-DBS-151 No.3

応用事例:



東南アジアでの天気予報改善のための国際 共同プロジェクト – by日本(京大etc, 気象庁・気 象研)&東南アジア各国の予報機関+大学



International Research for Prevention and Mitigation of Meteorological Disasters in Southeast Asia



MEXT Special Coordination Funds for Promoting Science and Technology for FY 2007 - 2009

Asia S&T Strategic Cooperation Program

Home

This is the Home Page of "International Research for Prevention and Mitigation of Meteorological Disasters in Southeast Asia" under the MEXT Special Coordination Funds for Promoting Science and Technology, supported for FY 2007 - 2009 under Asia S&T Strategic Cooperation Program.

Risk of high-impact weather associated with global warming and/or economical development is potentially increasing, and utilization of probability information obtained by ensemble NWPs (Numerical Weather Predictions) is a challenge for the development of decision support tools.

Thus, we establish "International Scientist-Network for Prevention and Mitigation of Meteorological Disasters in Southeast Asia" through research and development of downscaling NWP systems.

Name of affiliation: Kyoto University

Project leader: Shigeo YODEN (Professor of Meteorology)

Partnered institution in Meteorological Research Institute, Japan Meteorological Agency

Japan:

Group leader: Kazuo SAITO (Head of the 2nd Research Laboratory, Forecast

Home

<u>Motivations</u>

Major Research

Subjects

(a) Downscale

<u>NWPs</u>

(b) New Data

(c) Decision

Support System

<u>Topics</u>

<u>Meetings</u>

Newsletters
Publications

Presentations

Reports -

東南アジアにおける天気予報(+気象災害警

報作成)の現状

 気象観測 → 全地球気象データ網へ → 先進国が数値 予報 → 結果をもとに予報や警報注意報(∴気象機関は 数値予報結果のユーザ)

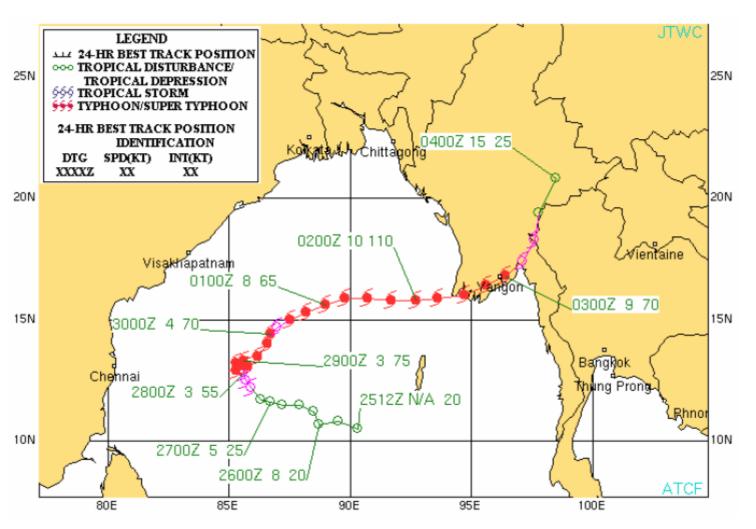
このプロジェクトの柱

- 1. 各国自前で数値予報できるように(気象庁の予報モデル 利用)
- 2. 計算コストもデータ量も膨大な最新予報技術である「アン サンブル予報」を東南アジアでも活用できるようにする(計 算コスト大⇒当面はデータ利用)

⇒ 2.にGfdnaviが活用された

: GUIで使いやすい. データ量膨大なので準リアルタイム転送不可.

Best track of Nargis



http://metocph.nmci.navy.mil/jtwc/atcr/2008atcr/2008atcr.pdf

アンサンブル各メンバの進路予報

科振費発表会のス ^A ライド(by大塚)より

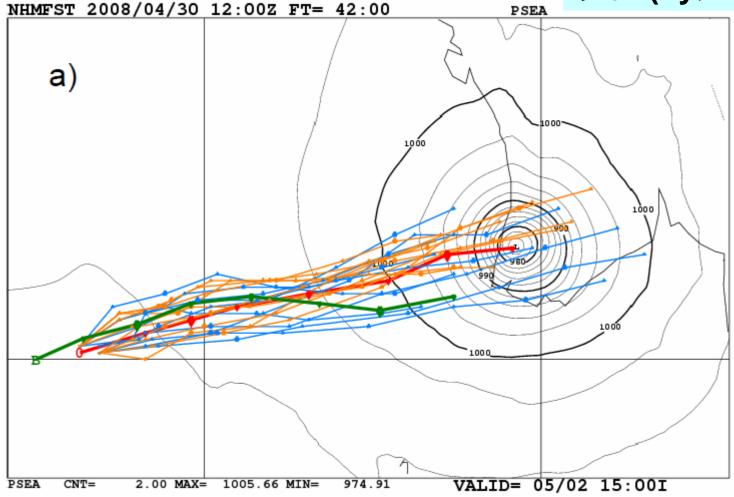


Fig. 11. Same as in Fig. 8 but NHM ensemble prediction with initial and lateral boundary perturbations.

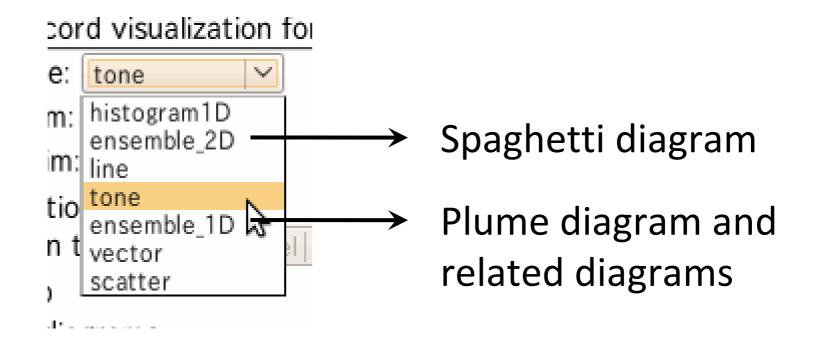
Fig. 8. a) Predicted tracks of Nargis until valid time 06 UTC 2 May 2008) by the 10 km NHM ensemble prediction (FT=42). Control run is shown by red line, and predicted sea level pressure at FT=42 is superimposed. Positive members (p01 – p10) are shown by orange lines while negative members (m01-m10) blue. Corresponding best track is also indicated (green). b) Time evolution of central pressures of Nargis predicted by NHM ensemble forecast.

5. A prototype decision support system

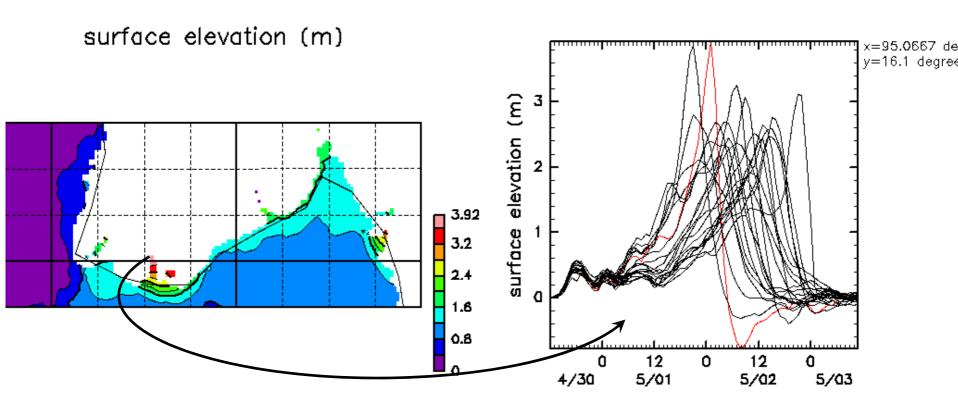
- Based on Gfdnavi
- Visualization functions and mathematical functions for ensemble NWP data
- Documentation to utilize Gfdnavi and above functions to analyze ensemble NWP data

5.2. New functions

Dropdown menu



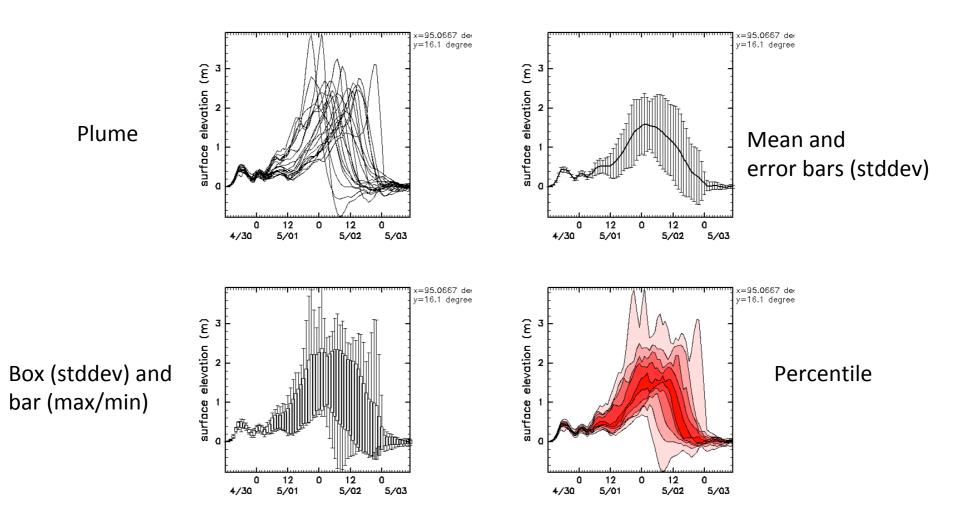
Plume diagram

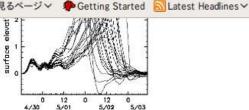


- Output from POM: Surface elevation (m)
- Left: Horizontal distribution of the maximum surface elevation $h_{\max(t,\ i)}(lon,\ lat)$
- Right: Plume diagram of $h_{i=0, ..., 20}(t)$ at Irrawaddy Point (95.07E, 16.10N)

科振費発表会のス Plume diagam and related diagra ライド(by大塚)より

Different expressions for the same quantity





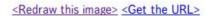
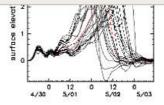
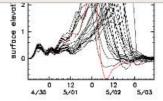


Fig.1. Time series of surface elevation at Irrawaddy point (95.07 degE, 16.10 degN) for 21 members.



<Redraw this image> <Get the URL>

Fig.2. Same as Fig. 1, but the control run is highlighted.



<Redraw this image> <Get the URL>

Fig. 3. Same as Fig. 1, but the member 1 (which shows the heighest surface elevation) is highlighted.

Previous Next

6. Decision support tools for ensemble numerical weather prediction: I. Basic diagrams

6.1 1D line plot

Data

/Nargis/NHM/POM/h.nc (lon, lat, t, member)

Settings

- Axes
 - o h member(t)
 - Ion = 95.07 degE
 - lat = 16.10 degN
 - = (X) t = [0 h, ..., 71 h]
 - (Ens) member = 0, ..., 20
- General Settings
 - o Draw method
 - ensemble 1D
- Specific settings
 - o style: lines

This diagram is called "Plume diagram".

Result

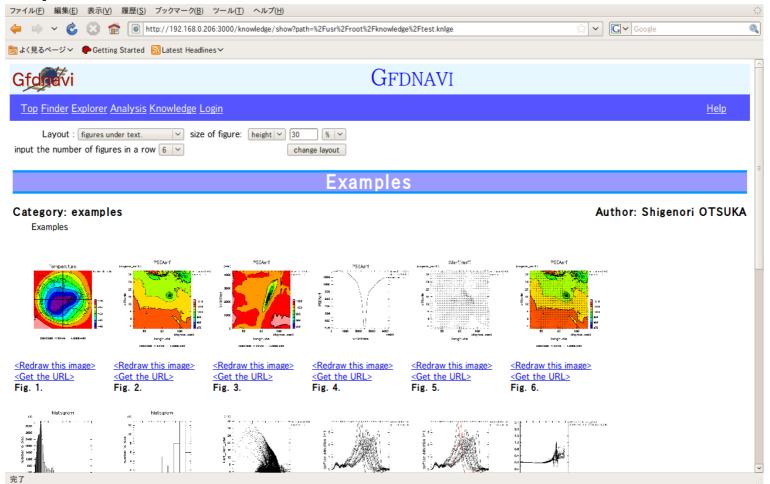
Time series of surface elevation at Irrawaddy point (95.07 degE, 16.10 degN) for 21 members Fig. 1. Some members show storm surge of more than 3 m in height.

Advanced usage

アンサンブル予報可 視化の手引きを用 意:それぞれがデー タにリンクされ再現可

Knowledge database

- Interactive documentation
 - Figures can be reproduced with the same parameters



- 地球流体科学における情報爆発=データの嵐
- データ解析のための優れた基盤ライブラリ開発 (GFD電脳Rubyプロジェクト~本特定領域)
- 地球流体分野のデータ解析可視化に役立つ WebベースのデータサーバGfdnaviを開発
 - 市場の店づくりから(プロや素人の)料理, 配膳まで
 - 地球科学者と情報科学者の(双方に益ある)協力... 今後へ