

最近のグローバルデータセット事情



東京大学 新領域創成科学研究科
環境学専攻 自然環境学講座 自然環境学形成学分野
助手 安形康

われわれに必要なデータセット

- 自然側
 - 河川流量
 - 地下水賦存量(地域によっては重要)
- 経済社会側
 - 人口分布(都市域とそれ以外にわけて)
 - GDP
 - 土地利用と農事暦
- これらすべての、現状と将来予測

自然側Dataset

- 従来:GSWP
 - 1987,88年の気候DSから河川に入る水の量をシミュレーション計算.それをもとに河川流量算定
これを「現在」とみなす
- いま(これから):GSWP2
 - 1986~95年のDSが構築中
 - <http://grads.iges.org/gswp2/>

GSWP2でのデータ配布法

- DODSという方法を使う
 - <http://grads.iges.org/gswp2/dods.html>
- DODS: オンラインDB
 - 米国・フランス・日本(生産研)に3つのサーバ
- データ取得方法
 - シンプルなftp
 - DODSを利用したhttp/CGI
 - netCDF - DODSを利用してプログラムから
 - GrADSなど, DODS対応アプリケーションから

DODSのWWWインタフェース

- 生産研サーバから, 月別LAIデータを取得する

- http://ftp.tkl.iis.u-tokyo.ac.jp:9090/dods/gswp/grid/lai_uk.info

からデータの情報を得る

- lai_ukが変数名,
.infoがコマンド

- 結果はこんな感じ

GSWP2 Mirror Site - info for /gswp/grid/lai_uk : [dds](#) [das](#)

DODS URL: http://ftp.tkl.iis.u-tokyo.ac.jp:9090/dods/gswp/grid/lai_uk

Description: Monthly Vegetation parameters from ISLSCP-II at 1x1 degree
Documentation: none provided
Longitude: -179.5°E to 179.5°E (360 points, avg. res: 1.0°)
Latitude: -59.5°N to 89.5°N (150 points, avg. res: 1.0°)
Altitude: 1 to 1 (1 points)
Time: 00Z01JAN1982 to 00Z01DEC1998 (204 points, avg. res: 30.43 days)
Variables: (total of 1)
lai: monthly leaf area index (m²/m²)

[back to parent directory](#)

complete metadata listing:

Global attributes:

title: "Monthly Vegetation parameters from ISLSCP-II at 1x1 degree"
convention: "COARDS"

Variables

lai: Grid
_FillValue: 1.0E20
long_name: "monthly leaf area index (m²/m²)"
lai: Array of 32 bit Reals [time = 0.203][lat = 0.149][lon = 0.359]
_FillValue: 1.0E20

DODSのWWWインタフェース

- 必要な情報：
 - 行列の次元 / 各次元に対応する軸 (変数) 名
 - 各軸のサイズ
- LAIの場合，
 - 3次元 / 「time」「lat」「lon」 / 各々204,150,360サイズ
 - Infoの結果に書いてある

```
Variables

lat: Grid
  _FillValue: 1.0E20
  long_name: "monthly leaf area index (m^2/m^2)"
  lai: Array of 32 bit Reals [time = 0..203][lat = 0..149][lon = 0..359]
  _FillValue: 1.0E20
  long_name: "monthly leaf area index (m^2/m^2)"
time: Array of 64 bit Reals [time = 0..203]
  units: "days since 1-1-1 00:00:0.0"
  long_name: "Time"
  minimum: "00Z01JAN1982"
  maximum: "00Z01DEC1998"
lat: Array of 32 bit Reals [lat = 0..149]
  units: "degrees_north"
  long_name: "latitude"
  minimum: -59.5
  maximum: 89.5
lon: Array of 32 bit Reals [lon = 0..359]
  units: "degrees_east"
  long_name: "longitude"
  minimum: -179.5
  maximum: 179.5
```

軸名 軸サイズ

DODSのWWWインタフェース: 個々の値を得る

- Time=0 , lat=100 , lon=200の点でのLAIの値は？
- CGIとして,
`lai_uk.asc?lai[0][100][200]`
というhttpリクエストを送る
 - .asc = 命令 .
 - 「？」に続いてクエリを書く

DODSのWWWインタフェース:

結果1

- こんな結果がテキストで返ってくる

```
http://ftp.tkl.iis.u-tokyo.ac.jp:9090/dods/gswp/grid/lai_uk.asc?lai[0][100][200]
ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
← 戻る → × ↻ 🏠 | 📁 お気に入り 🕒 🗑️ 🖨️

lai, [1][1][1]
[0][0], 0.828
結果

time, [1]
723547.0
lat, [1]
40.5
lon, [1]
20.5
各軸変数の,
その点での実際の値
```

DODSのWWWインタフェース: 個々の値を得る2

- Lat,lonをもう少し広げて,複数の点でのLAIの値を得たい
- CGIとして,
`lai_uk.asc?lai[0][100:105][200:202]`
といったhttpリクエストを送る
 - 座標値指定の[]の中では,コロンを用いて範囲指定が可能

DODSのWWWインタフェース:

結果2

- **こんな結果**

```
http://ftp.tkl.iis.u-tokyo.ac.jp:9090/dods/gswp/grid/lai_uk.asc?lai[0][100:105][200:202] - Microsoft Inter
ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)
← 戻る ▾ → ▾ × 🔄 🏠 📁 お気に入り 🌐 📄 📄
アドレス(D) http://

lai, [1][6][3]
[0][0], 0.828, 0.673, 0.716
[0][1], 0.599, 0.659, 0.741
[0][2], 0.611, 0.549, 0.695
[0][3], 0.544, 0.541, 0.603
[0][4], 0.411, 0.415, 0.527
[0][5], 0.333, 0.338, 0.435

time, [1]
723547.0
lat, [6]
40.5, 41.5, 42.5, 43.5, 44.5, 45.5
lon, [3]
20.5, 21.5, 22.5
```

結果

**各軸変数の,
その点での実際の値**

DODSのnetCDFインタフェース

- netCDF: グリッドデータのためのファイルフォーマット
 - GSWPで「静的に」データ送受を行うときの「共通言語」
- データ読み書きにはC/Fortran用インタフェースがよく使われる
 - PerlもJavaも, pythonすらもあるが...

netCDFの読み書き (C原語)

- プログラム中では...
 - `nc_open()`でローカルにある(見える)netCDFファイルを開く
 - `nc_inq_ndims()`, `nc_inq_nvars()`で次元数, 配列数を得る
 - `nc_inq_varid()`で変数IDを得る
 - `nc_get_var_<type>()`で変数を得る
 - ...
- コンパイル・リンク時には...
 - `libnetcdf.a`をリンクする
(リンク時に`-lnetcdf`をつける)
- ではDODSだと？

netCDFの読み書き (DODS)

- プログラム中では...
 - いったい変える必要はない
 - ただし, `nc_open()` でURLを指定することができる
 - つまりネットのむこうのnetCDFファイルに, まるでそれが自分のHDDに入っているかのようなつもりでアクセスできる.
- コンパイル・リンク時には...
 - `libnc-dods.a`をリンクする
(リンク時に`-lnc-dods`をつける)
 - つまりMakefileの変更は必要

netCDF-DODS:試してみました

- 体験記をWWWに書いています

- <http://hydro.iis.u-tokyo.ac.jp/~agata/archive/GSWP2/startdods.html>

- netCDFファイルを簡単にブラウズするtool:

- 元祖netCDFではncdump

- netCDF-DODSではdncdump

- dncdumpを試してみると

netCDF-DODS:試してみました

- 結果

```
# /usr/local/DODS/bin/dncdump -h
```

```
http://www.monsoondata.org:9090/dods/gswp/vector/tair_cru
```

```
netcdf tair_cru {
```

```
  dimensions:
```

```
    time = 39456 ;
```

```
    lat = 1 ;
```

```
    lon = 15238 ;
```

```
  variables:
```

```
    float tair(time, lat, lon) ;
```

```
    tair:_FillValue = 1.e+20f ;
```

```
    tair:long_name = "air temperature at 2m (kg/kg) " ;
```

```
    ***** (中略)
```

```
// global attributes:
```

```
:title = "NCEP reanalysis-2 data hybridized with CRU data" ;
```

```
:convention = "COARDS" ; }
```

ファイル名でなく、URLを指定
していることに注意

表示内容は、ローカルファイルに
ncdumpをしかけたときと同じ

DODS:現バージョンの問題点

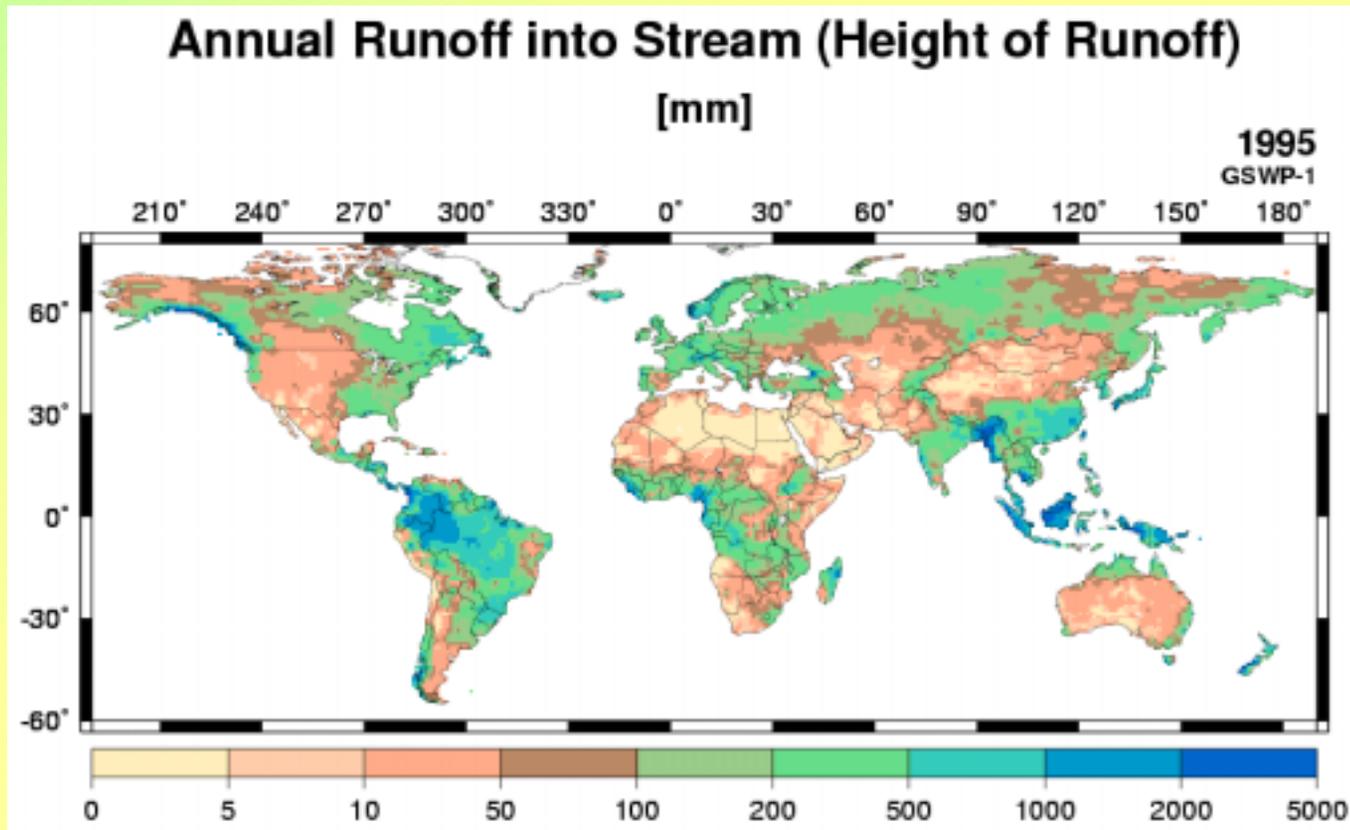
- gcc3ではインストールできない
 - ユーザからのコメントにより判明
 - C++部分でのバージョン互換性の問題か？
- Fortranコンパイラとして,
Intel Fortran(`ifc`)が考慮されていない
 - リンカオプションの追加変更など,
Makefileを自分で書き換える必要がある.
- スループットは未知数
- Webserviceとはあまり関連性がない

人文系データ

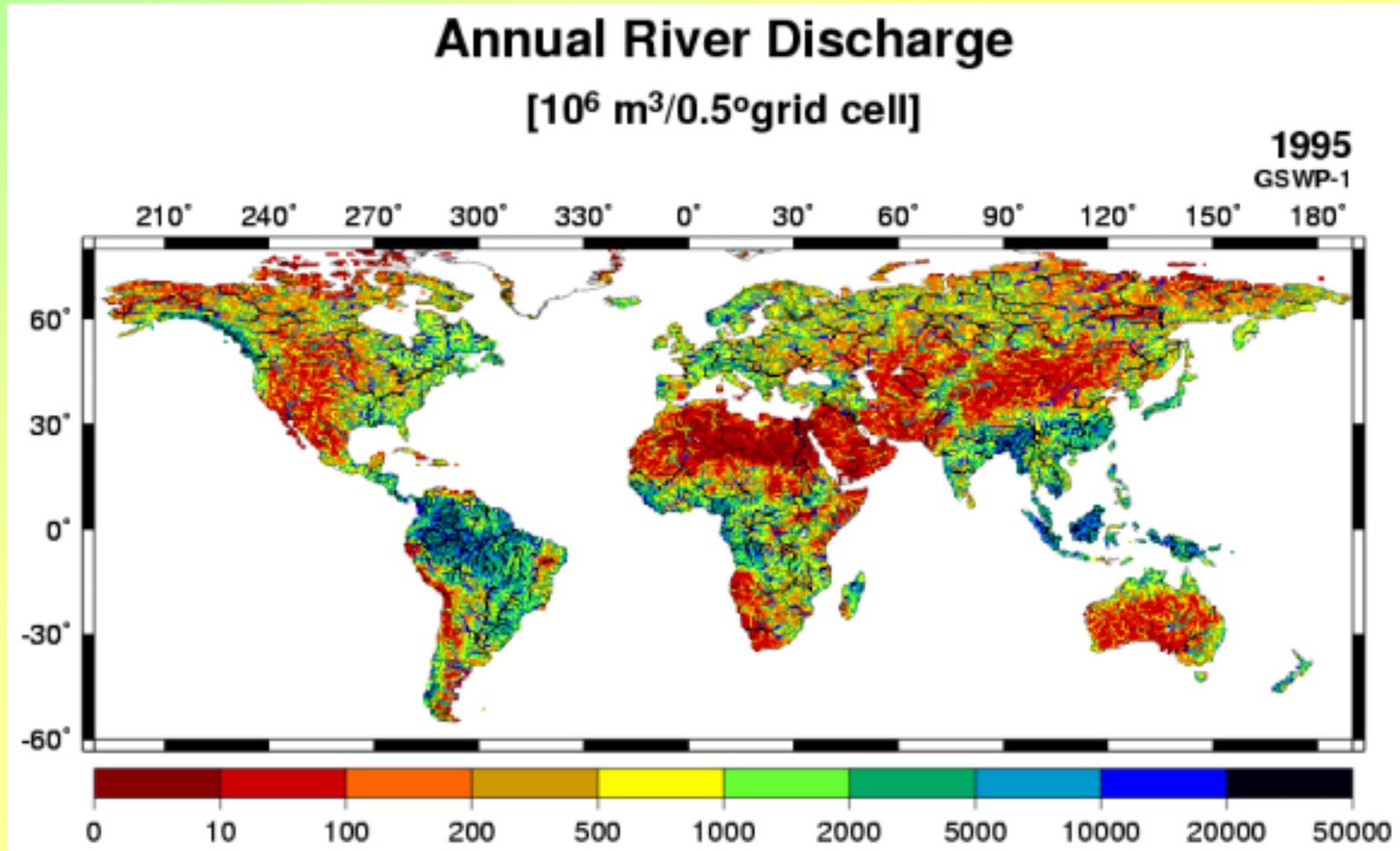
- これまで：
「現在」としては1990年ないし1995年がターゲット
– 2001年ころに仕事を始めたため
- これからは「現在」とは2000年がターゲット？
- データの公開・整備状況は？

GSWPのプロダクト: LSMからの 年河川流出高

- 1987・1988年平均 .



結果：年河川流量 0.5度グリッド版



任意地点での水資源量がわかるのがメリット

人口

- **CIESINの2.5分人口密度データ**
 - <http://sedac.ciesin.org/plue/gpw/index.html?main.html&2>
 - 現在Version2, 残念ながら1995年のもの
- **国連人口予測 (POPINおよびUNPOP)**
 - <http://www.un.org/popin/>
 - <http://www.un.org/esa/population/unpop.htm>
 - 後者は2000,2050,2100,2200,2300年, 国毎

水の利用

- **WRI**

- <http://earthtrends.wri.org/datatables/index.cfm?theme=2&CFID=105193&CFTOKEN=13304985>
- **Freshwater Withdrawals2003が最新**
- **国別の水使用量について, データ年度が (アフリカ以外は) おおよそ1990年代に揃ってきた**

- **FAOSTAT**

- <http://apps.fao.org/>
- **国毎の灌漑面積: 2001年度のものが最新**