

地球水循環データアーカイブシステムにおける異種データ相互解析機能とその実装

Implementation of Mutual Analysis Functions for Different Kinds of Data in Global Water Cycle Data Archive System

根本 利弘[▼] 小池 俊雄[◆] 喜連川 優[▼]

Toshihiro NEMOTO Toshio Koike
Masaru Kitsuregawa

地球水循環データの相互比較、統合などの解析処理を行うためのアーカイブシステムとそのユーザインタフェースについて述べる。地球水循環データは、データの種類やデータ作成機関により異なる次元、座標系、データフォーマットを有するが、本システムは、利用者にデータの差異を意識させることなく、相互に利用することを可能としている。アーカイブシステム内における多様な地球水循環データの統一的表现法、相互比較するための座標軸、時間軸の変換法、複数データ比較のためのデータ表示を行うユーザインタフェースについて説明する。

In this paper, we explain the archive system and its analysis interface for mutual comparison and integration of different kinds of global water cycle data. The global water cycle data have various dimensions, coordination, data format depending on their type or their providers. This system hides these diversity and provides the data analysis environment without taking account of the variety. This paper introduces the unified data representation, transformation of coordinates, user interface for mutual data comparison in detail.

1. はじめに

地球水循環の変動メカニズムを理解し、予測精度を向上させるため、統合地球観測戦略 (IGOS: Integrated Global Observing Strategy) では、水循環テーマの第一ステップとして統合地球水循環強化観測期間 (CEOP: Coordinated Enhanced Observing Period) プロジェクトを開始し、2002年10月から地球水循環統合化データセットの作成を行っている。このデータセットは、地上観測データ、衛星観測データ、数値気象予報モデル出力データにより構成されるが、これらは、次元・時空間分解能・座標系・精度・フォーマットの異なるファイルであり、これらを統合的に利用し、水循環変動の理解・予測へ役立てるシステムが求められている。

[▼] 正会員 東京大学生産技術研究所
{nemoto,kitsure}@tkl.iis.u-tokyo.ac.jp

[◆] 非会員 東京大学工学系研究科
tkoike@hydra.t.u-tokyo.ac.jp

本稿では、CEOPプロジェクトにより作成される地球水循環データをデータベース化し、アーカイブするシステムにおいて、これらを相互に利用・解析することを可能とするための機能、およびその実装について述べる。まず、アーカイブの対象とする地球水循環データについて説明を行う。その後、水循環関連データを相互利用し解析を行うために必要とされる機能を示し、それらの機能を実装したユーザインタフェースを紹介する。

2. アーカイブ対象データ

CEOPプロジェクトにおいて作成され、アーカイブ対象となる地球水循環データについて説明を行う。

2.1 地上観測データ

地上観測データは、世界36ヶ所のリファレンスサイトにおいて各種センサにより観測された物理量の時系列データである。観測時間間隔、観測される物理量の種類、センサの地表からの高度等はリファレンスサイトごとに異なる。また、1つの物理量を複数の地点あるいは高度において測定を行うリファレンスサイトもある。

地上観測データは、各リファレンスサイトにおいてクオリティチェックがなされ、各値に対してクオリティフラグが付与される。地上観測データは、担当機関である UCAR (University Corporation for Atmospheric Research) により、1時間ごとのデータに変換され、CEOP統一フォーマットにより提供される。

2.2 数値気象予報モデル出力

数値気象予報モデル出力は、日本の気象庁を含む世界11の気象予報機関による、全球予報モデルあるいは同化システムからの出力データである。数値気象予報モデル出力は、グリッドデータおよびMOLTS (Model Output Location Time Series) の2つに分類される。

グリッドデータは、全球の2次元あるいは3次元のデータである。予報機関により座標系、時空間分解能、出力される物理量は異なり、また、同じ物理量を表す予報値であっても、瞬時値であるか平均値であるかなどの違いがある。データはGribと呼ばれるフォーマットで、数値気象予報モデル配布担当機関であるMPI (Max Plunk Institute) に転送され、提供される。

MOLTSは、36ヶ所のリファレンスサイトを含む世界41地点に対する数値気象予報モデルの出力である。MOLTSは、そのフォーマットは統一されておらず、予報機関ごとにデータフォーマットが異なる。

2.3 衛星観測データ

衛星TRMM, DMSP, NOAA, GMS, Terra, Aquaなど、CEOPプロジェクトの観測期間に運用されている衛星による地表面観測データである。これらの衛星データはNASA, ESA, 宇宙航空研究開発機構などの衛星データ提供機関により幾何補正が行われ、世界36ヶ所のリファレンスサイトを中心とする250km四方の領域、5ヶ所のモンスーン領域、全球領域の3種類の領域の等緯度経度のデータに再配列されて提供される。空間分解能は、衛星に搭載されているセンサが有する分解能とほぼ等しくなるように定められているため、衛星、センサごとに異なる。また、提供される物理量データの種類も、衛星に搭載されているセンサに依存する。衛星データの大部分は地表面上の物理量の2次元データであるが、センサによっては鉛直方向の物理量の分布を測定、生成が可能なものがあり、この場合は3次元データとなる。データフォーマット

は衛星データ提供機関ごとに異なる。

3. 地球水循環データアーカイブシステム

本節では、現在構築を行っている地球水循環データアーカイブシステムの構成について述べる。

3.1 システム構成

図1に地球水循環データアーカイブシステムの構成を示す。システムはクライアントーサーバモデルにより実現しており、クライアントーサーバ間の通信は HTTP に基づく SOAP を採用し、ウェブサービスとして実装している。しかし、地球水循環関連データはサイズが極めて大きく、1つ1つの物理量を忠実に XML 化して通信を行うことは極めて非効率であるため、データの転送に関しては、データ列の各値を IEEE754 で表現して圧縮することによりバイナリ配列化して通信を行い、転送時間の短縮を図っている。

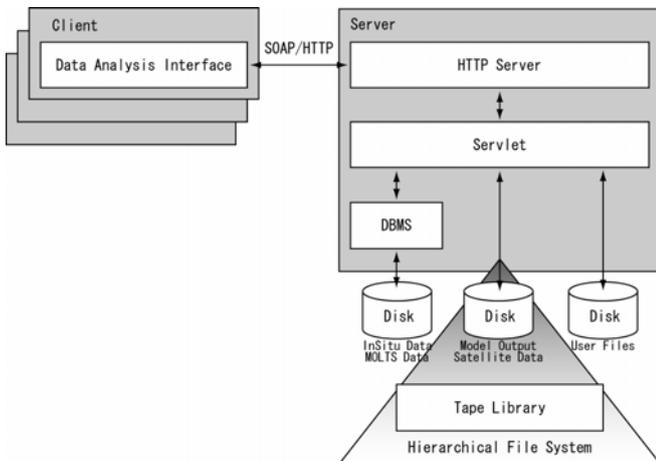


図1 システム構成

Fig.1 System architecture

3.2 データサーバ

サーバでは、システムへのログイン、指定した条件に基づくデータの検索、検索されたデータの時間軸、空間軸の変換、検索されたデータに対する演算や統計量の計算などのメソッドを提供する。

サーバ側においては、データ量が比較的小さい現地観測データ、MOLTS データは、メタデータとともにデータそのものを DBMS によって管理するが、数値気象予報モデル出力におけるグリッドデータ、および衛星観測データは、データ自体をディスクアレイ、あるいはテープライブラリシステムを利用した階層ファイルシステム上のファイルとして格納し、DBMS はそのインデックスとメタデータのみを管理する。

利用者により検索のメソッドが実行されると、データサーバは DBMS、あるいは外部ファイルから検索条件に合うデータを切り出し、内部表現形式に変換し、サーバ内の利用者データ領域に保存する。利用者は、この利用者データ領域内の切り出されたデータに対して、各軸の変換や演算、統計量の計算等のメソッドを実行する。メソッドの実行結果として生成されるデータもまた、新しいファイルとしてデータサーバ内の利用者データ領域に保存され、利用者は検索から直接作成されたデータと検索結果データに対して処理を加えて作成されたデータを区別することなく利用することができる。

3.3 データ検索・解析インタフェース (クライアント)

データ検索・解析インタフェースは、データサーバのウェブサービスに対するクライアントアプリケーションとして利用者側の計算機上で動作し、利用者の要求に応じて、データサーバのメソッドを実行する。データ検索・解析インタフェースは JAVA により記述されている。

データの解析処理は、データを作成する処理に関しては、データ検索・解析インタフェースがデータサーバのメソッドを実行し、処理自体はデータサーバで行われる。一方、データの表示は、データ検索・解析インタフェースがデータ転送メソッドを実行し、データサーバからデータ検索・解析インタフェースへ物理量データが転送され、データ検索・解析インタフェース上でグラフ、画像化が行われる。データサーバ上でグラフ化、画像化を行い、その画像をクライアントへ転送する方法も考えられるが、この場合にはグラフ化、画像化パラメータを変更するたびに画像を転送する必要が生じること、グラフ、画像上の指定された点の物理量の値の表示要求に対応するためには、画像データに加えて、結局物理量自体の値もクライアント側で必要となることによる。

4. 異種データ相互利用とその実装

4.1 統一的データ表現

地球水循環データは、データごとに次元が異なるが、地球水循環データアーカイブシステム内部においては、検索され切り出されたデータをすべて、時間、鉛直方向、南北方向、東西方向の4つの軸を有する4次元配列として表現し、この4次元配列にデータの情報、および4つの各軸の目盛値、および各軸の情報を付加して表現する。例えばリファレンスサイトにおける地上気温の時系列データは、時間軸方向のみ n のサイズを持ち、鉛直軸、南北軸、東西軸のサイズが1である、 $n \times 1 \times 1 \times 1$ の配列となる。配列の各要素には、物理量を表す数値に加え、そのデータのクオリティを表すフラグが付与される。データの情報には、データの種別、単位、データが検索された際に指定された条件、データ作成時刻等により構成され、また、軸情報は、軸の種類や目盛値の単位等により構成される。各軸の目盛値は、範囲の指定も可能としている。これは、その値が示す物理量が、ある範囲における集約値である場合にも対応するためである。この表現法はアーカイブ対象となる地球水循環データのフォーマットをほぼ包含しており、統一的に検索されたデータを扱うことが可能となる。このデータ表現は利用者には隠蔽されており、利用者にはデータ表現については意識をする必要はない。

```

DataInfo = {データ種別, 単位, ...}
Axis1[T] = {{目盛値11, 層厚11}, {目盛値12, 層厚12}, ...}
Axisx[X] = {{目盛値x1, 層厚x1}, {目盛値x2, 層厚x2}, ...}
Axisy[Y] = {{目盛値y1, 層厚y1}, {目盛値y2, 層厚y2}, ...}
Axisz[Z] = {{目盛値z1, 層厚z1}, {目盛値z2, 層厚z2}, ...}
Value[T][X][Y][Z] = {{{{
    {値1,1,1,1, クオリティフラグ1,1,1,1}, ...
}}}}
    
```

図2 統一的データ表現法

Fig.2 Unified data representation

4.2 データ検索

CEOP プロジェクトにおける地球水循環データは、データの種類、時刻、位置の3つのパラメータによって決定される。データ検索・解析インタフェースは、これらの3つのパラメータをメニュー形式で提示する。データ検索・解析インタフェースは、データサーバに対して選択されたパラメータを引数として検索メソッドを実行し、データサーバは検索結果を利用者領域に保存する。データの種類によってデータが存在する期間や位置が異なるが、データが存在するパラメータの組み合わせをデータサーバにおいてデータベース化しており、利用者があるパラメータを選択すると、データの存在に関する問い合わせメソッドを実行してデータの存在するパラメータの組み合わせを取得し、他のパラメータのメニューの文字の輝度を変更することで、データが存在しないパラメータの組み合わせが分かるようにしている。

4.3 フォーマット変換

地球水循環データアーカイブシステムを用いて解析を行う場合には、データサーバがそれぞれのフォーマットを自動的に判断して読み込むため、利用者はデータフォーマットの差異を考慮する必要はない。一方、利用者が自身のプログラムや地球水循環データアーカイブシステム以外の解析ツールで地球水循環データを利用するために、地球水循環データアーカイブシステムでは、ASCII テキストファイルもしくはは気象学、海洋学などで一般的に使用されている NetCDF フォーマットに変換してクライアント側のマシンに保存する機能を提供する。データ検索・解析インタフェースは、データの転送メソッドを実行し、データサーバより統一的数据表現形式の保存対象データを受け取り、データ検索・解析インタフェースが要求されたフォーマットに変換して、クライアント側マシンに保存する。

4.4 単位の変換

地球水循環データを直接比較するためには、統一された単位に変換する必要があるが、地球水循環データアーカイブシステムでは、地球水循環データを扱う上でよく用いられる単位の変換は、データ検索・解析インタフェースでメニューを選択することのみで実行できる。また、メニューに登録されていない単位の変換に関しても、利用者が変換式を入力することができる。データ検索・解析インタフェースが単位変換メソッドを発行することにより、データサーバは対象データの単位の変換を行い、利用者領域に新たなデータとしてファイルを作成する。

4.5 時空間軸変換・整合

地球水循環データの相互比較のため、地球水循環データアーカイブシステムにおけるデータサーバは、最近隣値、線形補間値、最大値、最小値、平均値によって再サンプリングする方法により、利用者が定義する新たな時間軸に既存データを整合させるメソッドを提供する。また、既存のあるデータの時空間軸に他のデータを整合させるメソッドも提供し、この場合には、最近隣法、線形補間法が適用可能である。これらの時空間軸変換・整合処理時には、データが瞬時値のようにある1点の物理値ではなく、ある範囲における集約値である場合も自動的に考慮され、変換元データのメタデータに従い、自動的に適切な値を選択する。データ検索・解析インタフェースより、時空間軸変換・整合メソッドが発行されると、データサーバは利用者領域に新たなデータとして、時空間軸変換・整合を行ったデータを作成する。

4.6 演算・解析機能

データ解析を進めるためには様々な演算が必要とされる。

地球水循環データアーカイブシステムでは、複数のデータの相関係数、回帰係数を、散布図の表示ウィンドウのメニューの係数表示コマンドを選択することにより表示することができる。また、利用者が四則演算や算術関数によって構成される任意の数式を入力することにより、様々な演算を行うことも可能である。演算メソッドが実行されると、データサーバは指定された演算を実行し、演算結果は利用者領域に新たなデータとして作成される。

4.7 データ表示機能

地球観測データの解析を行うために、地球水循環データアーカイブシステムは、様々なデータ可視化機能を提供する。データ検索・解析インタフェースがデータ転送メソッドを実行しデータサーバより物理量データを取得し、データ検索・解析インタフェースがグラフ画像、ビットマップ画像を作成して、表示を行う。

4.7.1 複数データのグラフ表示

データ検索・解析インタフェースは、データを比較するために、複数のデータを1つのグラフ上に表示する機能を提供している。グラフのx軸、y軸には、時間軸、鉛直軸、南北軸、東西軸、物理量の値を任意に指定可能である。注目領域を拡大する機能が提供しており、また、表示対象としたデータが異なる物理量、時空間次元を有する場合でも、表示可能である。x軸、y軸以外の軸は、画面内にスライドバーが表示され、このスライドバーを操作することにより表示対象値を変更できる。

4.7.2 散布図表示

複数データの関連を表示するために、データ検索・解析インタフェースには、散布図を作成し、表示する機能が実装されている。散布図は複数のデータの対応をとり、それぞれの物理値を軸として作成されるが、異なる時間、高度、位置のデータを比較する場合もあり、4つの軸に関して、いずれをデータ対応の対象とするかの選択ができる。また、散布図には回帰直線も同時に表示され、さらにこのウィンドウのメニューから相関係数・回帰係数の表示を選択することで、その数値が表示される。

4.7.3 画像の重ね合わせ

データ検索・解析インタフェースには、データを2次元ビットマップ画像として表示する機能が実装されている。複数のデータを選択して、それらを重ね合わせて表示することが可能である。図3は数値気象予報モデル出力のグリッドデータと衛星観測データを重ね合わせて表示している。グラフの場合と同様、x軸、y軸にはデータの任意の軸を指定することが可能であり、図3の例では、x軸に東西軸（経度）、y軸に南北軸（緯度）を指定している。また、図3で画像化の対象としたデータは時系列データであり、時間軸のサイズが2以上であるが、このようにグラフ化、画像化の際にx軸、y軸として指定しない軸でそのサイズが2以上である場合には、ウィンドウ上にスライドバーが表示され、任意にそれらの軸の値を変えることができる。

画像化パラメータとして、カラーリング（カラーチャート、グレイスケール）の選択、等位線、海岸線、地点名の表示の有無等を指定することができる。複数のデータを重ね合わせて表示した場合、個々のデータに対して画像化パラメータを指定することができ、また、図4のウィンドウで指定することにより、重ね合わせる画像の順序の変更、およびそれぞれの画像に対する透明度のスライドバーで操作することが可能である。



図3 重ね合わせ表示
Fig.3 Example of overlaid images

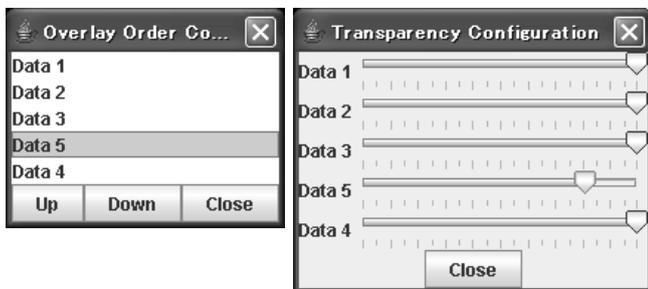


図4 重ね合わせ順序 (左) と透明度 (右) の指定
Fig.4 Specification of overlaying order (left) and transparency (right)

4.7.4 複数グラフ・画像の同期

グラフ表示、画像表示においては、時間軸、鉛直軸、東西軸、南北軸の4つの軸のうち、x軸、y軸に指定された以外の軸のサイズが2以上の場合、スライドバーが表示される。このとき、複数のグラフ表示ウィンドウ、画像表示ウィンドウのスライドバーの同期ボタンをセットすることにより、それらの軸を同期させることが可能である。同期対象となっている軸の値は、最近隣法により選択される。同期ボタンが押されると、データ検索・解析インタフェースは同期ボタンが押されているウィンドウの同期対象軸間の対応テーブルを作成する。スライドバーによりあるウィンドウの値が変更されると、データ検索・解析インタフェースはこの対応テーブルを参照し、他のウィンドウの軸の値を変更する。

5. おわりに

本稿では、現在構築を行っている地球水循環データアーカイブシステムにおいて、異種データを相互に利用し、解析を行うための機能について述べた。利用者が異種データを相互に利用し、解析を行う際に利用される基本的な機能の実装はほぼ完了したが、より高度な機能も要求されており、今後、実装をさらに進め、機能の充実を図る予定である。

[文献]

- [1] Barclay, T., Slutz, D. R., and Gray, J.: "TerraServer: A Spatial Data Warehouse", Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp.307-318 (2000).
- [2] Szalay, A. S., Gray, J., Thakar, A. R., Kunszt, P. Z., Malik, T., Raddick, J., Stoughton, C. and vandenBerg, J.: "The SDSS SkyServer - Public Access to the Sloan Digital Sky Server Data", Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, pp.570-581 (2002).
- [3] Stolte, E., Praun, C., Alonso, G. and Gross, T.: "Scientific Data Repositories - Designing for a Moving Target", Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp.349-360 (2003).
- [4] Koike, T.: "The Coordinated Enhanced Observing Period - an initial step for integrated global water cycle observation", WMO Bulletin, vol.53, no.2, pp.115-121 (2004).
- [5] 根本利弘, 小池俊雄, 喜連川優: "地球水循環データアーカイブシステムとユーザインタフェースの構築", 電子情報通信学会技術報告, vol.105, no.173, pp.49-54 (2005).
- [6] 根本利弘, 小池俊雄, 喜連川優: "地球水循環データアーカイブシステムにおける異種データ相互解析機能の実装", 電子情報通信学会第18回データ工学ワークショップ(2007).
- [7] Nemoto, T., Koike, T. and Kitsuregawa, M.: "Data Analysis System Attached to the CEOP Centralized Data Archive System", Journal of the Meteorological Society of Japan, Vol.85A, pp.529-543 (2007).

根本 利弘 Toshihiro NEMOTO

東京大学生産技術研究所戦略情報融合国際研究センター助教。1994年東京大学大学院工学系研究科博士課程退学。博士(工学)。地球環境データアーカイブ、大規模ストレージシステムの研究に従事。情報処理学会、日本データベース学会会員。

小池 俊雄 Toshio KOIKE

東京大学大学院工学系研究科教授。同大学地球観測データ統合連携研究機構機構長。1985年東京大学大学院工学系研究科土木工学専門課程博士課程修了。工学博士。水文学、河川工学、リモートセンシング工学に関する研究に従事。IPCC第4次評価報告WG1評価編集者、世界気候研究計画(WCRP)統合地球水循環強化観測期間プロジェクト(CEOP) Lead Scientist, 地球観測に関する政府間グループ(GEO)10年実施計画タスクチームメンバー(H16,H17), 同構造とデータ委員会共同議長(H17,H18)。

喜連川 優 Masaru KITSUREGAWA

東京大学生産技術研究所教授。同所戦略情報融合国際研究センター長。1983年東京大学大学院工学系研究科情報工学専攻博士課程修了。工学博士。データベース工学、並列処理、Webマイニングに関する研究に従事。本会理事、情報処理学会フェロー、SNIA-Japan顧問、ACM SIGMOD Japan Chapter Chair (H11-H14), 電子情報通信学会データ工学研究専門委員会委員長(H9,H10)。VLDB Trustee, IEEE TKDE Assoc. Editor, IEEE ICDE, PAKDD, WAIM Steering Comm. Member.