

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

根本 利弘¹ 小池 俊雄² 喜連川 優¹

あらまし CEOPプロジェクトにより作成される地球水循環に関連する現地観測・衛星観測・数値気象予報モデル出力データを統合的に管理するアーカイブシステム、およびこれらのデータを用いた解析を行うためのユーザインタフェースについて述べる。アーカイブ対象となるデータについて説明し、気象学・水文学の研究者がデータ解析時に必要とする機能をまとめる。さらに、現在構築中の、異なる次元・時空間分解能・座標系・精度・フォーマットを有する地球水循環関連データを統一的に扱い、利用者が異なるデータを意識することなしに扱うことができる環境を提供する地球水循環データアーカイブシステム、ユーザインタフェースについて述べる。

Construction of Global Water Cycle Data Archiving System and User Interface

Toshihiro NEMOTO¹ Toshio KOIKE² and Masaru KITSUREGAWA¹

Abstract On the CEOP (Coordinated Enhanced Observing Period) project, in order to improve our understanding of global water cycle system, large amount of data are being collected and archived. We describe characteristics of the global water cycle data and functions requested by users in order to analyze the data. We explain the architecture of archiving system now we are constructing and introduce user interface for browsing and analyzing.

1. はじめに

地球の表面には莫大な量の水が存在している。これらの水は海面・陸面から蒸発して大気中に放出され、大気中を移動し、上空で雲となり、雨・雪として地表へ降り、再び蒸発し、というように循環している。近年、水不足、豪雨災害、水質汚染、生態系の破壊など、水に関わる問題が世界各地で広がり、これらに起因する食糧難や伝染病が蔓延するなど、その影響は開発途上国においてますます拡大しているが、これらの問題の背景には、急激な人口増加による水需要の増大や都市開発、産業発展などの社会的要因に加えて、水循環の変動が大きな要因となっている。すなわち、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ことは、水危機の有力な解決策のひとつである。このような背景の下、国連関係機関、国際科学計画、地球観測衛星委員会らにより構成される統合地球観測戦

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、これらのデータを統合的に利用可能とし、水循環変動の理解・予測へ役立つシステムを構築することが求められている。

本稿では、CEOPプロジェクトにより作成される地上観測データ・衛星観測データ・数値気象予報モデル出力データをデータベース化し、アーカイブするシステムについて説明するとともに、利用者には、これら地球水循環データの検索、および検索されたデータを用

¹東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

Institute of Industrial Science, The University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

²東京大学大学院工学系研究科 〒113-8656 東京都文京区本郷 7-3-1

School of Engineering, The University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

いた解析を容易に行うことを可能とするユーザインタフェースについて述べる。まず、アーカイブシステムの要件として、アーカイブ対象となる水循環関連データについて説明をするとともに、利用者がデータ解析を行うために必要とする機能を述べる。さらに、現在構築をこなっているアーカイブシステムのアーキテクチャを述べ、その後、ユーザインタフェースについて紹介を行う。

2. 要求事項

2.1. 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。

2.1.1. 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観測された時系列データである。地上観測データは、リファレンスサイトに設置されたセンサにより観測された物理量の時系列データであり、地表面データ、土壌データ、上層大気データに分類される。地表データは、地表面付近において観測された気温、気圧、湿度、降水量、放射量など、土壌データは、地下数 cm～数 m における温度や湿度など、上層大気データは、ラジオゾンデなどにより観測される上空の気温、気圧、湿度などの時系列データである。CEOP プロジェクトにおいては観測対象となる物理量は定められているが、各リファレンスサイトにおいてすべての物理量が観測されるわけではなく、観測される物理量、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ている。また、観測時間間隔はリファレンスサイトごとに異なるが、地上観測データ配布担当機関である 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。データ量は年間数百 MB 程度である。

2.1.2. 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

データであり、各格子は、気温、気圧、湿度などの予報値を持つ。CEOP プロジェクトにおいては、出力されるべき予報値が定められているが、各機関におけるモデルの違いにより、必ずしもすべての予報値が出力されず、あるいは、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。また、CEOP プロジェクトによって定められた予報値に加え、それら以外の他の予報値を提供している機関もある。予報機関ごとに使用する数値モデルや予報システムの運用方法が異なるため、出力される 3 次元格子の座標系、時空間分解能は、予報機関ごとに異なる。例えば、水平方向では、緯度、経度を等間隔に分割した正方格子の他に、緯度、経度の間隔が不等なガウス格子と呼ばれる座標系が用いられ、鉛直方向では、等気圧面や、モデルでのレイヤそのものなどが用いられる。また、数値気象予報モデル出力においては、シミュレーションの初期値としてどの時刻の観測値を用いるか、すなわちシミュレーションの開始時刻をいつにするかにより、同じ時刻に対する数値気象予報モデル出力でも複数のデータが存在する。グリッドデータは、各予報機関において 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 Plank Institute) に転送され、提供される。データ量は年間数十 MB である。

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。リファレンスサイトにおける現地観測データとのより詳細な比較目的として作成され、グリッドデータよりも時間分解能は高い。グリッドデータ同様、予報機関ごとに提供される予報値の数、種類、鉛直方向の座標系、分解能が異なる。また、同時刻の予報値も複数存在する。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。

2.1.3. 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、4～5 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、センサごとに提供されるデータの分解能

は異なる。また、提供される物理量データの種類の、衛星に搭載されているセンサにより異なるが、輝度温度、アルベドなどの低次物理量のみならず、海面温度、土壌水分量等の高次物理量データも提供される。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、センサによっては高度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。データ量は年間数百TBと想定されている。

2.2. 要求機能

地球水循環データは、気象現象の解明、数値気象予報モデルの予報精度向上、衛星データ処理アルゴリズムの改良などのため、統合的に利用される。このために必要とされる機能は、利用者の行うデータ解析手法により様々であるが、大まかに以下のようにまとめられる。

2.2.1. データ検索

地球水循環データは、データの種類（現地観測データ、衛星観測データ、数値気象予報モデル出力、さらにはその中の気温、気圧、土壌水分量など）、時刻、位置（全球、4～5カ所のモンスーン領域、36ヶ所のリファレンスサイト）によって決定され、これらの条件に基づくデータの検索、取得は不可欠である。さらに、現地観測データや数値気象予報モデル出力においては、データ提供機関やリファレンスサイトの位置によって提供されるデータが異なるため、例えば、ある地点において利用可能な海面水温のデータにはどのようなものがあるか、というような検索も必要とされている。

2.2.2. フォーマット変換

地球水循環データは、データごとにフォーマットが異なる。現地観測データでは、すべてのリファレンスサイトのデータで統一のフォーマットが採用されているものの、CEOP独自形式であり、また、MOLTSデータではデータ提供数値予報機関ごとに異なる形式のフォーマットが採用されている。利用者が既存のアプリケーションで使用する場合には、そのアプリケーションが対応しているフォーマットに変換する必要がある。また、利用者が自ら作成したプログラムで利用する場合にも、利用しやすいフォーマットに統一されている方が好ましい。このため、気象学、海洋学などで一般的に使用されているフォーマットに変換する機能が必要とされている。

2.2.3. 時空間軸変換・整合

地球水循環データは、データごとに時空間分解能や

座標系が異なり、そのままでは相互に比較することはできない。このために、時空間の分解能や座標系を変換し、データ間の時空間軸を整合させる機能が必要とされる。

一般的には、最近隣法、線形補間、スプライン補間などにより再サンプリングする方法が用いられ、地球水循環データにおいても、これらの手法は適用可能であるが、さらに、これらの手法に加えて実際の物理量に基づく方法も必要とされる。例えば、1時間単位の降水量を1日単位に変換する場合には、24時間分のデータをすべて加算しなければならない。最高気温、最低気温などを変換する場合には、変換前の対応する時間、空間の最大値、最小値をとる必要がある。また、平均気温などの場合には、必要な重み付けをして平均値を再度計算する必要がある。

2.2.4. 統計量計算

データの解析のために、平均値、最大値、最小値、分散、総和などの各種統計量の計算機能が要求されている。単一の領域の統計量（例えば、あるリファレンスサイトを中心とした250km四方の領域の平均値）や時空間のある単位ごとの統計量（例えば、12時間ごとの平均値）などに加え、日中変動、季節変動、あるいは緯度帯、高度帯などにおける統計値、いわゆる串刺し統計量を得るための計算機能、複数の任意の地点、領域の統計量、さらには、陸面領域のみ、海面領域のみ、ある標高以上の地点のみの統計量などの計算機能が要求されている。

2.2.5. 演算・解析機能

データ解析を進めるための様々な演算が要求されている。例えば、提供されている物理量から、直接提供されていない物理量の導出、複数のデータの比較のための差分、相関係数、回帰係数の計算、時系列データの時間微分、2次元データ、3次元データでの渦度、発散の計算などが必要とされている。また、これらの演算が繰り返し適用可能でなければならない。

また、多くの利用者が共通して使用するような演算が容易に適用可能であることに加え、利用者ごとに頻繁に使用される演算は異なるため、利用者ごとに演算を登録し、簡単に呼び出せるようになっていることが望まれている。

2.2.6. データ表示

地球水循環データを表示し、相互に比較できるようにする機能が要求されている。提供されているデータは多様であるため、データ、解析手法に即した様々な表示方法が要求されている。2次元の折れ線グラフ、

各物理量を輝度などで表現した 2 次元の画像、等位線により表現された画像、さらには 3 次元表示など様々なデータ表現法が要求されている。これらの表示においては、時間軸、水平方向空間軸（緯度、経度）、鉛直方向空間軸（高度）、およびデータの値をそれぞれ 2 次元表示では x 軸、y 軸に対して、3 次元表示ではさらに z 軸に対して、任意に割り当て可能であること、すなわち時間軸を含めた 4 次元データの各軸に垂直な平面による切断面の表示が要求されている。さらには、これらの軸に垂直な平面とは異なる、いずれの軸にも直交しない平面、さらには利用者が指定した曲面での切断面の表示も求められている。また、切断面を移動させることによる動画表示も要求されている。

また、データの比較を行うために、複数のデータをひとつのグラフ上に表示したり、2 次元画像として表現したデータを重ね合わせたりする機能、複数の画像を連携させ、同じ時刻、地点の対応付けが容易にできることも望まれている。

3. 地球水循環データシステム

本節では、現在構築を行っている水循環関連データシステムの構成、および現地観測データ、衛星観測データ、数値気象予報モデル出力データを統一的に扱う内部データ表現法について述べる。

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。地球水循環データの利用者は世界各地に散在し、その計算機・ネットワーク環境は様々である。ファイアウォールを越えてアクセスする必要がある場合も多いが、HTTP を採用することで、より多くの利用者への対応が容易となる。また、ウェブサービスとして提供することにより、利用者が自ら作成したアプリケーションで直接水循環関連データへアクセスすることも可能となる。一方、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的である。このため、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。

3.2. 内部データ表現

地球水循環データは、データごとに次元が異なるが、

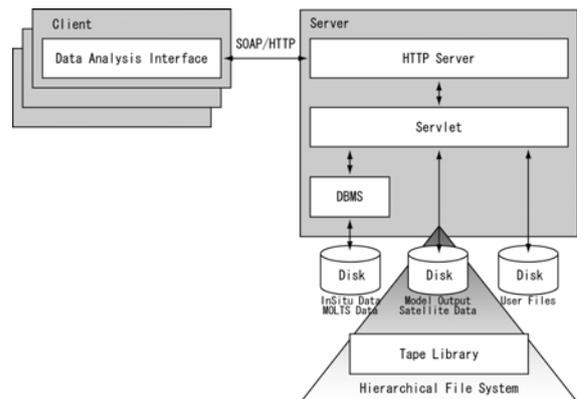


図 1 システム構成

本システムにおいては、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、データの種類、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、ある範囲の総和や平均値といった明確な 1 点ではなく、ある範囲における値である場合にも対応するためである。

3.3. 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、1 地点における物理量であるためデータ量が比較的小さい現地観測データ、数値気象予報モデル出力における 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、2 次元、3 次元的な広がりを持つ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 においてはそのインデックスとメタデータのみを管理する。数値気象予報モデル出力のグリッドデータ、衛星観測データはそのサイズが大きく、すべてのデータをディスク

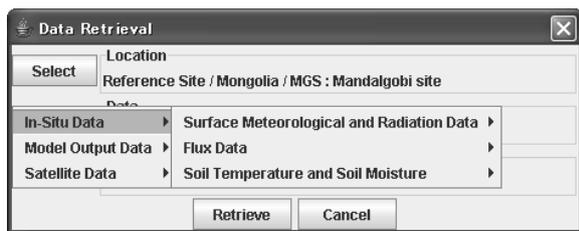


図2 データ検索条件入力

No.	Label	Dimension	Data	Location	Period	Creation Time
1			0 In-Situ / Meteorolo	Reference Site / N.	2001/09/05 00:00	Tue Mar 01 16:13
2			0 In-Situ / Surface /	Reference Site / N.	2001/09/05 00:00	Tue Mar 01 16:14
3			2 Satellite / OMS S-V	Reference Site / N.	2001/09/05 00:00	Tue Mar 01 16:14
4			3 Satellite / TRIM P	Reference Site / N.	2001/09/05 00:00	Tue Mar 01 16:15
5			2 Model Output / EC	Global	2001/09/05 00:00	Tue Mar 01 16:17
6			2 Model Output / EC	Global	2001/09/05 00:00	Tue Mar 01 16:19
7			2 Model Output / EC	Global	2001/09/05 00:00	Tue Mar 01 16:21
8			2 Model Output / EC	Global	2001/09/05 00:00	Tue Mar 01 16:22
9			0 MOLTS / JMA / G4	Reference Site / E.	2002/01/01 00:00	Wed Mar 30 14:53
10			0 MOLTS / UKMO / J3	Reference Site / E.	2002/01/01 00:00	Fri Apr 01 14:35.2
11			3 Model Output / NC	Global	2002/01/01 00:00	Wed Apr 08 10:47
12			2 Model Output / NC	Global	2002/01/01 00:00	Fri Apr 08 17:30.3
13			2 Model Output / CP	Global	2001/07/01 00:00	Fri Apr 08 17:40.2
14			2 Masked / Satellite	Reference Site / N.	2001/09/05 00:00	Wed Apr 20 16:02
15			0 Processed / Zonal	Reference Site / N.	2001/09/05 00:00	Wed Apr 20 16:04
16			0 Processed / Diurn	Reference Site / E.	2002/01/14 00:00	Mon May 02 14:00
17			1 In-Situ / Meteorolo	Reference Site / E.	2002/01/01 00:00	Thu May 19 14:00
18			0 MOLTS / JMA / G4	Reference Site / E.	2002/01/01 00:00	Thu May 19 14:00
19			0 Processed / Diurn	Reference Site / E.	2002/01/01 00:00	Thu May 19 14:14
20			1 Processed / Diurn	Reference Site / E.	2002/01/01 00:00	Thu May 19 14:14

図3 データ管理

上に格納するのは困難であるため、階層ファイルシステム上のファイルとする必要があるが、DBMSによって提供されている BLOB は階層ファイルシステム上のファイルには必ずしも最適化されておらず、十分な性能が発揮されないことが考えられるためである。

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る。これらの処理によって新たに作成されたデータも内部表現形式によって表現されており、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。

3.4. データ検索・解析インタフェース

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、データサーバへリクエストを発行する。

図2は、データ検索条件を入力するためのウィンドウである。データの検索は、データの位置、種類、期

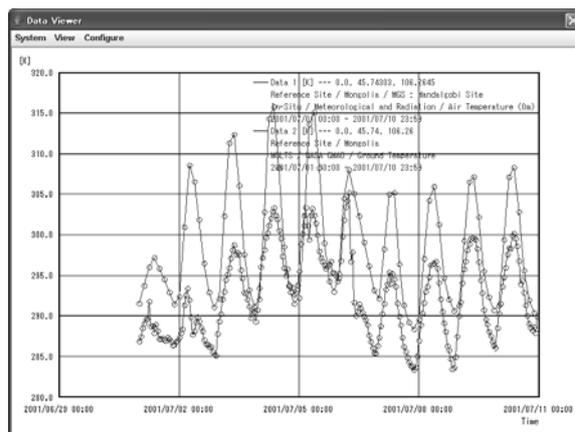


図4 2次元グラフ表示



図5 2次元画像表示

間に関する条件を指定することにより行われる。各条件はメニューにより提供されており、利用者は必要な条件をメニューから選択するのみである。また、データの期間に関しては、メニューによる選択以外に直接数値を入力することも可能である。条件を入力し、検索ボタンを押すと、サーバに対してリクエストが発行されてサーバ上で検索メソッドが実行され、検索結果として切り出されたデータが利用者領域に格納される。

図3はサーバの利用者領域上のデータの管理ウィンドウである。検索、処理されたデータの簡単な情報が表形式により表示されている。このウィンドウから、より詳細な情報の表示や、不要データの削除を行うことができる。また、統計量の計算や演算、表示の対象とするデータの選択も、このウィンドウを通じて行われる。操作の対象とする1つ以上のデータを表から選択し、上部のメニューバーから操作を選択することにより、必要なリクエストがサーバへ発行されて、処理が実行される。

図4は利用者領域上のデータを折れ線グラフで表示した例である。2つのデータを同時に1つのグラフ上

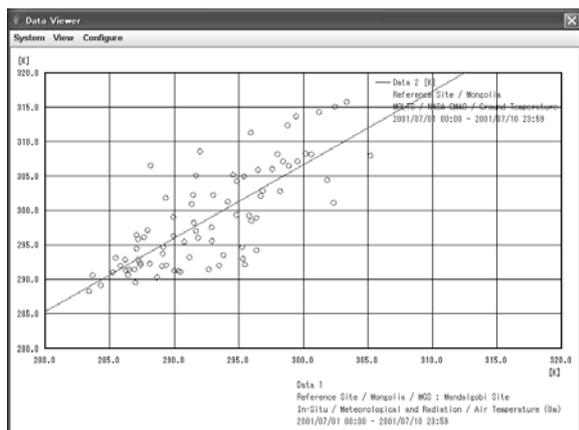


図 6 散布図表示

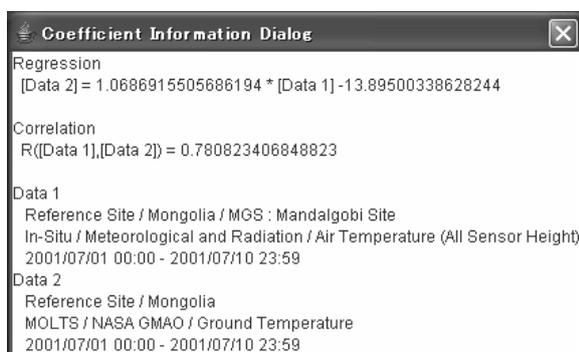


図 7 回帰式・相関係数表示

に表示している。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できるが、この例では、1 地点における物理量の時系列値である現地観測データと MOLTS データを選択しているため、x 軸を時間軸, y 軸を物理量の値としている。図 5 は利用者領域上のデータを 2 次元ビットマップ画像として表示した例である。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。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。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, 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 5 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ダーが表示され, 任意にそれらの軸の値を変更することができる。また, 複数のグラフ, 画像を表示させ, それらのスライダーを同期させて 1 つのスライダーを移動させることで複数のグラフ, 画像の指定

された軸を変更することも可能である。

データの表示は, クライアントからデータの転送リクエストが発行されてサーバからクライアントへ内部表現形式の物理量データが転送され, クライアント上にてグラフ, 画像化が行われる。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。

図 6 は, 複数のデータを散布図として表示した例であり, 図 7 はそれらのデータ間の回帰式, 相関係数を表示した例である。図 3 のデータ管理画面においてデータを選択し, メニューより機能を選択することで, 図 6 のような散布図が表示され, さらに散布図が表示されているウィンドウのメニューで指定することにより, 図 7 の回帰式, 相関係数が表示されというように, 簡単な操作で, 解析作業を進めることが可能となっている。

4. 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関して述べた。まず, アーカイブ対象のデータ, 利用者からの要求機能についてまとめ, その後, 構築中のシステムの構成を説明するとともに, データ検索・解析インタフェースの紹介を行った。利用者が解析を行う際に頻りに利用される機能のすべてが実装されてはならず, 今後, 実装を進め, 機能の充実を図る予定である。

文 献

- [1] T. Nemoto, E. Ikoma and M. Kitsuregawa, "Design of data server for CEOP data," Proc. 2nd Asia Pacific Association of Hydrology and Water Resources Conference, no.2, pp.558-565, Singapore, July 2004.
- [2] E. Stolte, C. Praun, G. Alonso and T. Gross, "Scientific Data Repositories – Designing for a Moving Target," Proc. ACM SIGMOD International Conference on Management of Data, pp.349-360, San Diego, California, USA, June 2003.
- [3] T. Koike, "The Coordinated Enhanced Observing Period – an initial step for integrated global water cycle observation," WMO Bulletin, vol.53, no.2, pp.115-121, April 2004.