

大規模な気候シミュレーションデータを効率的に探索・取得するシステム（SEAL）を開発 ～都道府県単位の将来予測も簡単表示、温暖化適応策検討にも貢献～

1. 発表のポイント

- ◆ 数ペタバイトの大規模な気候のシミュレーションデータから高速かつ効率的に必要な情報を抽出する技術を開発した。
- ◆ 従来のシステムと比べて約 1%の所要時間で必要な情報を抽出することが可能であり、専門的な知識がなくとも直感的に操作可能なインターフェースを採用している。
- ◆ 将来の気候変動予測や影響評価などの研究が飛躍的に進むと期待されるとともに、地域ごとの温暖化適応策検討にも貢献すると期待される。

2. 概要

国立研究開発法人海洋研究開発機構（理事長 松永 是、以下「JAMSTEC」という。）付加価値情報創生部門情報エンジニアリングプログラムの中川 友進 特任技術副主幹及び国立大学法人京都大学（総長 山極 壽一）学術情報メディアセンタービジュアルライゼーション研究分野の小山田 耕二教授らは、文部科学省委託事業「気候変動適応技術社会実装プログラム（Social Implementation Program on Climate Change Adaptation Technology：SI-CAT）」の元で、数ペタバイトの大規模アンサンブル気候データ（※1）から高速かつ効率的に必要な情報を抽出する技術を開発しました。

「地球温暖化対策に資するアンサンブル気候予測データベース（d4PDF）」の解析は、気候変動予測やそれに伴う不確実性を定量評価するためにとっても重要です。しかし、d4PDF のデータ量は約 3 ペタバイトと膨大です。これまで、従来のシステムではシミュレーションモデル名や変数名などシミュレーション前に決めた情報から検索しなくてはならず、情報の探索・取得には長時間を要し、また大容量の記録装置を備える必要があるなど高いハードルがありました。

そこで本研究では、シミュレーション後に得られる降水量や気温といった物理量を都道府県単位の空間情報にまとめ、時間的に圧縮し、必要なデータを効率的に探索・取得する技術を開発しました。この技術を実装したシステム「SI-CAT 気候実験データベースシステム（System for Efficient content-based retrieval to Analyze Large volume climate data：SEAL、図 1）」は、本プレスリリース後、文部科学省の事業で開発されているデータ統合・解析システム（DIAS）にて公開します。（図 2）。

SEAL の使用によって、従来の Web ベースの検索システムと比べて、ユーザーがダウンロードするデータの容量は 0.5%未満に、また必要なデータを見付けるまでの総時間は 1%未満にまで減ることとなるため（図 3）、今後は過去の顕著な気象イベントに対する要因分析、将来変化予測の不確実性の理解、影響評価などの研究が飛躍的に進むものと期待される（図 4）ほか、ユーザーフレンドリーなインターフェースによって、研究者のみならず、各省庁、自治体、産業界等においても幅広く利用されるものと期待されます。

本成果は、公益社団法人日本地球惑星科学連合が発行する学術雑誌「Progress in Earth and

Planetary Science」に2月26日付（日本時間）で掲載されました。

タイトル: Development of a System for Efficient Content-Based Retrieval to Analyze Large Volume of Climate Data

著者: 中川友進¹、尾上洋介²、川原慎太郎¹、荒木文明¹、小山田耕二³、松岡大祐¹、石川洋一¹、藤田実季子¹、杉本志織¹、岡田靖子¹、川添祥¹、渡辺真吾¹、石井正好⁴、水田亮⁴、村田昭彦⁴、川瀬宏明⁴

所属: 1. 国立研究開発法人海洋研究開発機構、2. 学校法人日本大学、3. 国立大学法人京都大学、4. 気象庁気象研究所

3. 背景

コンピュータの性能の向上に伴い、大規模アンサンブル気候データのデータサイズは急激に増加しています。例えば d4PDF のデータサイズは約 3 ペタバイトですが、過去のデータセットと比べて、非常に大きなデータサイズです。このような大規模アンサンブル気候データの系統的な解析は、確率的な気候変動の影響評価を行うために非常に重要です。しかし系統的な解析には、一般に高性能なコンピュータだけでなく大容量のデータ記憶装置が必要となります。ユーザーが d4PDF を自身のローカルコンピュータにダウンロードしようとする、1) ユーザーのディスク容量の不足、2) 長時間のダウンロード、3) データサーバーへの高い負荷、という問題が発生することから、個々の研究者による系統的な解析は困難になりつつあります。

現在、気候の研究分野では DIAS や WDCC (<https://cera-www.dkrz.de/WDCC/ui/cerasearch/>) など従来の Web ベースの検索システムが活用されています。これら全ての従来システムは、シミュレーション後に得られる降水量や気温といった物理量を使用してデータを検索するようには設計されておらず、データファイルそのものに関連付けられているメタデータを使用してデータファイルを検索するように設計されており、前述の問題が発生します。そこで本研究では、物理量などデータファイルの内容に関連付けられているメタデータを使用してユーザーがデータファイルを検索・取得できる技術の開発を行いました。また、その技術を実装したシステムである SEAL を DIAS から公開します。

4. 成果

SEAL は図 1 の概念図で示すようにリレーショナルデータベース（※2）、データ提供機能及びユーザーインターフェースで構成されています。リレーショナルデータベースは空間と時間で圧縮されたシミュレーションデータを格納しており、検索を効率的にする重要な役割を担っています。空間的な圧縮とは、20km 四方で格子状に分割された領域データを 47 都道府県ごとにまとめることを指しています。また、降水量や気温といった物理量は時別値ではなく日別値、月別値または年別値を必要とすることから、47 都道府県ごとに、降水量の積算や気温の平均を計算することで時間的な圧縮を行い格納しています。

データ提供機能は、ユーザーの検索結果に基づいて空間的・時間的に切り出されたデータを提供可能とします。多くのユーザーはテキスト形式や CSV 形式などの人間が読める形式へ変換したデー

タを必要としていることから、バイナリデータ（コンピュータが直接的に処理するためのデータ）をテキスト形式または CSV 形式へ変換する機能を付与しています。

ユーザーは Web から簡易にデータ検索やダウンロードが可能なユーザーインターフェース（Web UI）を通じて本システムを利用可能です。Web UI は、条件に合うデータを見付けることを主な機能とした「SEAL-Finder（SEAL-F）」、データを分析および可視化することを主な機能とした「SEAL-Visualizer（SEAL-V）」で構成されています。図 2 は SEAL の Web UI のトップページのスクリーンショットであり、SEAL-F と SEAL-V へアクセスするためのリンクが貼られています。ユーザーは SEAL-F を使用すると検索の返り値を数値として取得することができ、また SEAL-V を使用すると検索の返り値をヒストグラムなどの画像として取得することができます（図 3）。

空間的・時間的に圧縮されたデータの圧縮率を計算したところ、空間的な圧縮率は北海道（302 個の 20km 格子で構成）では $1/302 \approx 0.33\%$ 、東京都と大阪府（それぞれ 13 個の 20km 格子で構成）では $1/13 \approx 7.7\%$ となりました。また、時間的な圧縮率は、時別値から日別値へ圧縮した場合は $1/24 \approx 4.2\%$ となりました。そのため、空間的圧縮と時間的圧縮の両方を適用した場合、日別値のデータサイズは 0.01~0.3% へと縮小されます。ケーススタディとして従来システムと SEAL を使用した方法で必要なデータファイルを取得するまでの所要時間を定量的に調べたところ、図 4 の条件では 1% 未満となることを確認することができました。

従来、大規模アンサンブル気候データから必要なデータファイルを取得するには「ユーザーのディスク容量の不足」、「長時間のダウンロード」及び「データサーバーへの高い負荷」という問題がありましたが、本成果により、これら全ての問題を解決することが可能です。

5. 今後の展望

SEAL は d4PDF の利用を飛躍的に促進し、学術の発展へ貢献するとともに気候変動に関わる関連企業（コンサルタントなど）が自治体からの業務を請け負う際に SEAL を利用することなども期待されます。

本成果により、複数の格子の物理量を予め積算し空間解像度を粗くすることで高速な検索を実現しましたが、今後は格子のそのままの物理量に対して高速な検索を実現する研究開発の方向性が重要となると考えています。SEAL のために開発された技術は、他の研究分野におけるシミュレーションや観測のデータに対しても非常に有用だと考えられ、今後、様々な分野のビッグデータへの応用も期待されます。

【補足説明】

※1 アンサンブル気候データ：将来気候のシミュレーションを行う際に、初期値に少しばらつきを与えて複数回の計算を行なって作成したデータを意味する。

※2 リレーショナルデータベース：事前定義された相互関係を持つデータ項目の集まり。データ項目は、列と行を持つテーブルの組み合わせとして定義される。リレーショナルデータベースを使用すると、テーブル間の関係を事前に定義することにより複雑なデータの間係を扱うことができる。一般に、リレーショナルデータベースとの通信のインターフェースとして Structured Query Language (SQL) が使用されている。PostgreSQL はリレーショナルデータベースを管理するためのシステムのこと。

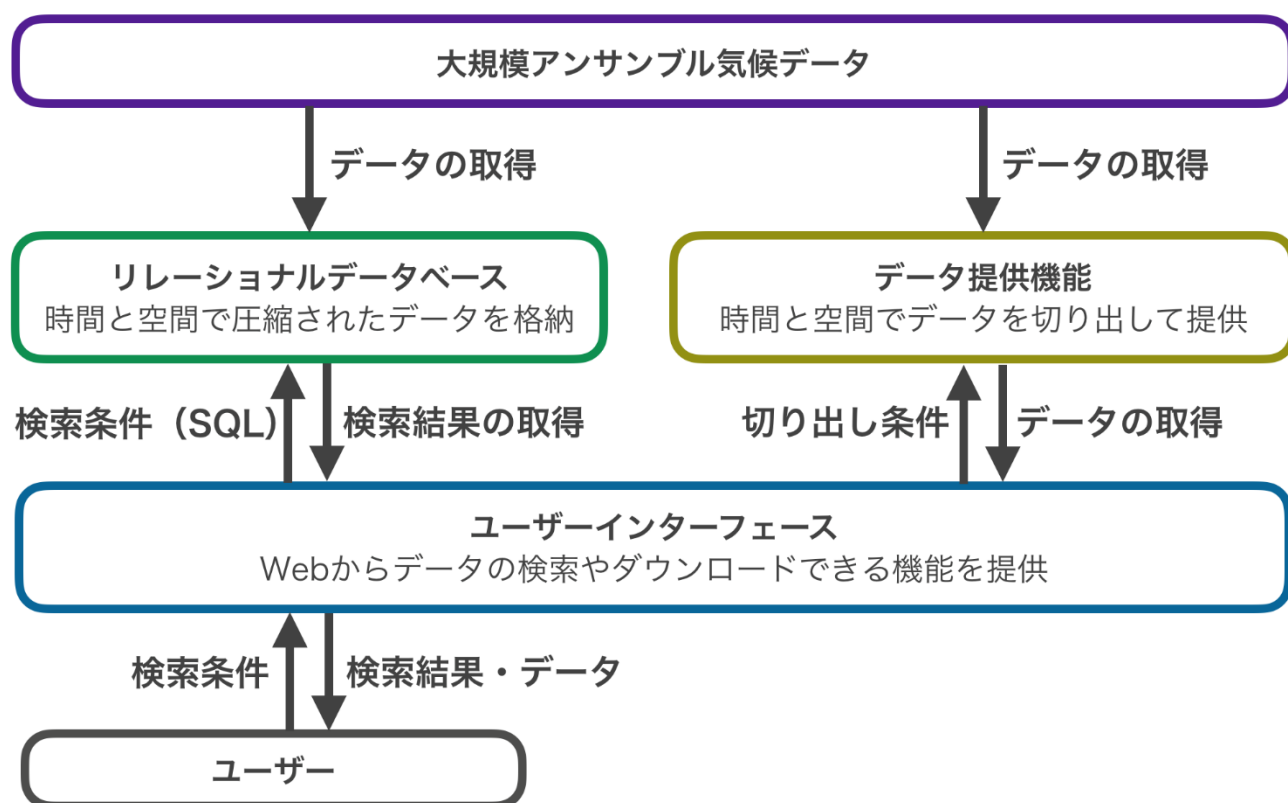


図1 SEALの概念図。



SI-CAT 気候実験データベースシステム

System for Efficient content-based retrieval to
Analyze Large volume climate data (SEAL)

概要

SI-CAT気候実験データベースシステム (System for Efficient content-based retrieval to Analyze Large volume climate data; SEAL) は、気候変動適応技術社会実装プログラム (Social Implementation Program on Climate Change Adaptation Technology; SI-CAT) において開発されました。

SEALは地球温暖化対策に資するアンサンブル気候予測データベース (database for Policy Decision making for Future climate change; d4PDF) のデータから、高速かつ効率的に必要なデータを見付けるためのシステムです。d4PDFのデータは、[データ統合・解析システム \(Data Integration and Analysis System: DIAS\)](#) からダウンロードすることができます。当ページの下方にある「[利用の手引き](#)」を参照の上、ご利用ください。



SEAL-F (SEAL-Finder) について

SEAL-Fは、d4PDFから条件に合ったデータを絞り込むためのWebユーザーインターフェースです。降水量や気温について、都道府県などの条件を入力することで、極端現象などの必要なデータを検索することができます。また、検索結果に基づき、オリジナルデータをダウンロードすることができます。

SEAL-V (SEAL-Visualizer) について

SEAL-Vは、d4PDFのデータを可視化するためのWebユーザーインターフェースです。降水量や気温について、ヒストグラムなどを描くことができます。

利用の手引き

利用の手引きは、SEALの利用方法を説明したものです。Webユーザーインターフェースのスクリーンショットを見ながら、利用方法を知ることができます。とりあえず利用したい方は、第3章をご覧ください。より詳細な利用方法を知りたい方は、第2章をご覧ください。利用の手引きは現在作成中です。間違いを見付けられたり、質問・コメントが御座いましたら、umineko_support@jamstec.go.jpまでご連絡下さい。

[利用の手引き \(初版\) \(最終更新日: 2018年12月26日\)](#)

謝辞

- SEALでは、気候変動リスク情報創生プログラムのもとで作成された、地球温暖化施策決定に資する気候再現・予測実験データベース (d4PDF; [Mizuta et al. 2016](#)) を使用した。
- SEALでは、気候変動適応技術社会実装プログラム (SI-CAT)のもとで作成された、将来2°C昇温実験 ([Fujita et al. 2018](#)) を使用した。
- SEALでは、気象庁気象研究所により作成された、台風トラекデータ ([Yoshida et al. 2017](#)) を使用した。

論文等での引用

- SEALを利用した場合は、Nakagawa et al. 2020, accepted by PEPSを引用して下さい。
- d4PDFを利用した場合は、[Mizuta et al. 2016](#), [Fujita et al. 2018](#)を引用して下さい。
- 台風トラекデータを利用した場合は、[Yoshida et al. 2017](#)を引用して下さい。

問い合わせ先

JAMSTEC 海洋研究開発機構 umineko_support@jamstec.go.jp

図2 SEALのWeb UIのトップページのスクリーンショット。

SEAL-F System for Efficient content-based retrieval to Analyze Large volume climate data (SEAL) - Finder

データセット：

変数：

実験の種類：

検索の種類：

行政区域：

閾値 (日降水量) (X_{RAIN}) : (mm)
 閾値 (連続降水日数) (X_{days}) : (日)
 開始年月日 (T_s) : (YYYY-MM-DD)
 終了年月日 (T_e) : (YYYY-MM-DD)
 閾値 (積算降水量) (X_{tot}) : (mm)

検索の種類の詳細：ある行政区域の指定期間 ($T_s > \text{年月日} \geq T_e$) において、連続降水日数 (日降水量が X_{RAIN} mm を超える日を降水日とする) が X_{days} 日以上、かつ積算降水量が X_{tot} mm 以上のケースを検索する。

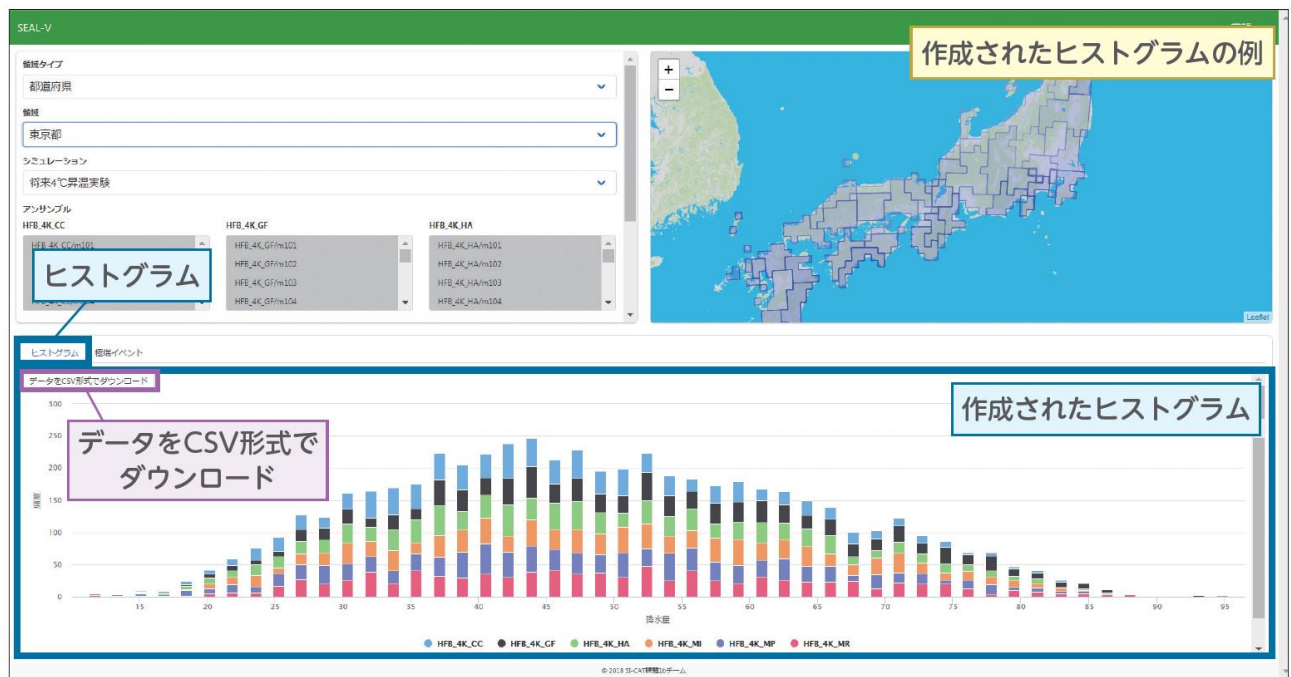
年月日の指定可能範囲 (領域モデル実験)

将来4℃昇温実験：2050-09-01 ~ 2111-09-01 or 2050-09 ~ 2111-09
 将来2℃昇温実験：2030-09-01 ~ 2091-09-01 or 2030-09 ~ 2091-09
 過去実験：1950-09-01 ~ 2011-09-01 or 1950-09 ~ 2011-09

年月日の指定可能範囲 (全球モデル実験)

将来4℃昇温実験：2051-01-01 ~ 2110-12-31
 過去実験：1951-01-01 ~ 2011-12-31
 非温暖化実験：1951-01-01 ~ 2010-12-31

図 3-1 SEAL-F の検索ページ。データセット (領域モデル実験、全球モデル実験)、変数 (降水量、気温、台風トラックデータ (全球モデル実験のみ))、実験の種類 (領域モデルは将来 4℃昇温実験・将来 2℃昇温実験・過去実験の 3 種類、全球モデルは将来 4℃昇温実験・過去実験・非温暖化実験の 3 種類)、検索の種類、および行政区域 (47 都道府県) やシミュレーション期間などの検索の種類ごとに設定された条件を指定する。検索の種類は、降水量ならば例えば 3 日間積算降水量、日降水量の年最大値や連続無降水日数といったものをリストから選択可能。気温ではある期間の平均気温、猛暑日日数や真冬日日数などを選ぶことができる。検索結果は SEAL-V を用いると以下のように簡単にグラフ化できる。



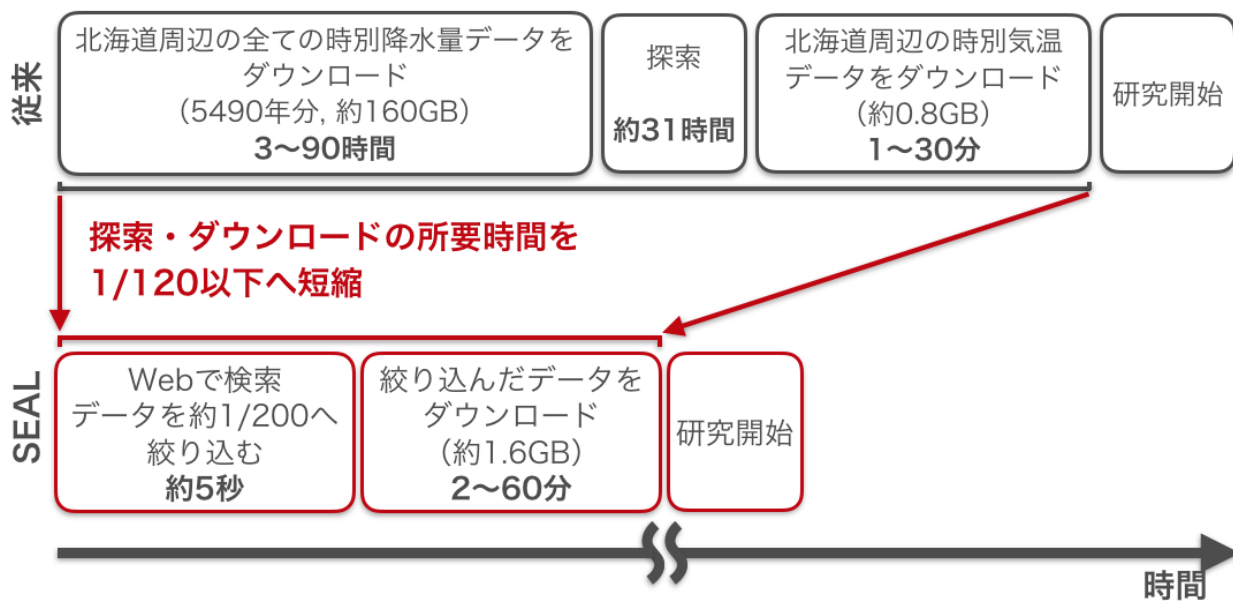


図4 SEALの予想される利点の概略図。図に示されている経過時間は、DIASからJAMSTECへのダウンロード速度（4～120 Mbps）に基づく、将来4度昇温実験のデータを使用した際の実測値。