デジタル台風：
複合災害の現状認識を目指すビッグデータ解析

北本 朝展（KITAMOTO Asanobu)

国立情報学研究所／総合研究大学院大学

http://agora.ex.nii.ac.jp/digital-typhoon/

@KitamotoAsanobu

気象予報士第5439号

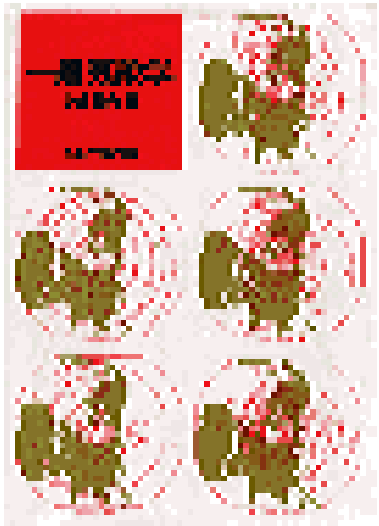
謎のウェブサイト？

- 一体、誰が、何のためにやっているウェブサイトなのか？
- 構造が普通の気象ウェブサイトと違う。
- やたらデータが多く、変なデータもある。
- 気象専門家でもない謎の人物が、好き放題にあれこれ書いている。

自己紹介

- 研究分野は情報学。特に画像処理、画像データベース。
- その後、ビッグデータ研究へ。特に気象情報、地球環境情報、人文科学情報などへの適用。
- 最近はオープンサイエンスの研究なども。

気象（台風）とのかかわり



1. 小学生のころ、**テレビの天気予報**が好きでよく見ていた。
2. 大学1年生のとき、「**一般気象学**」という本を見つけて買った。
3. 大学院の研究室では、**気象衛星NOAA（後にひまわり）画像処理**とアーカイブを行っていた。
4. 就職して2年目、**情報学的な観点**から台風を研究対象とした。
5. 2005年夏、気象予報士。

情報学的に見た台風

- **識別子**：気象庁がIDを一元的に管理し、他のメディアもこれを尊重する。
- **異種データ**：画像、テキスト、数値など、あらゆる種類のデータが生成される。
- **リアルタイム性**：最新データの重要性が高く、時々刻々と状況が変化する。
- **社会的に高インパクト**：公式情報だけでなく、SNS等でも話題となりやすい。

「デジタル台風」のそもそもの目的とは？

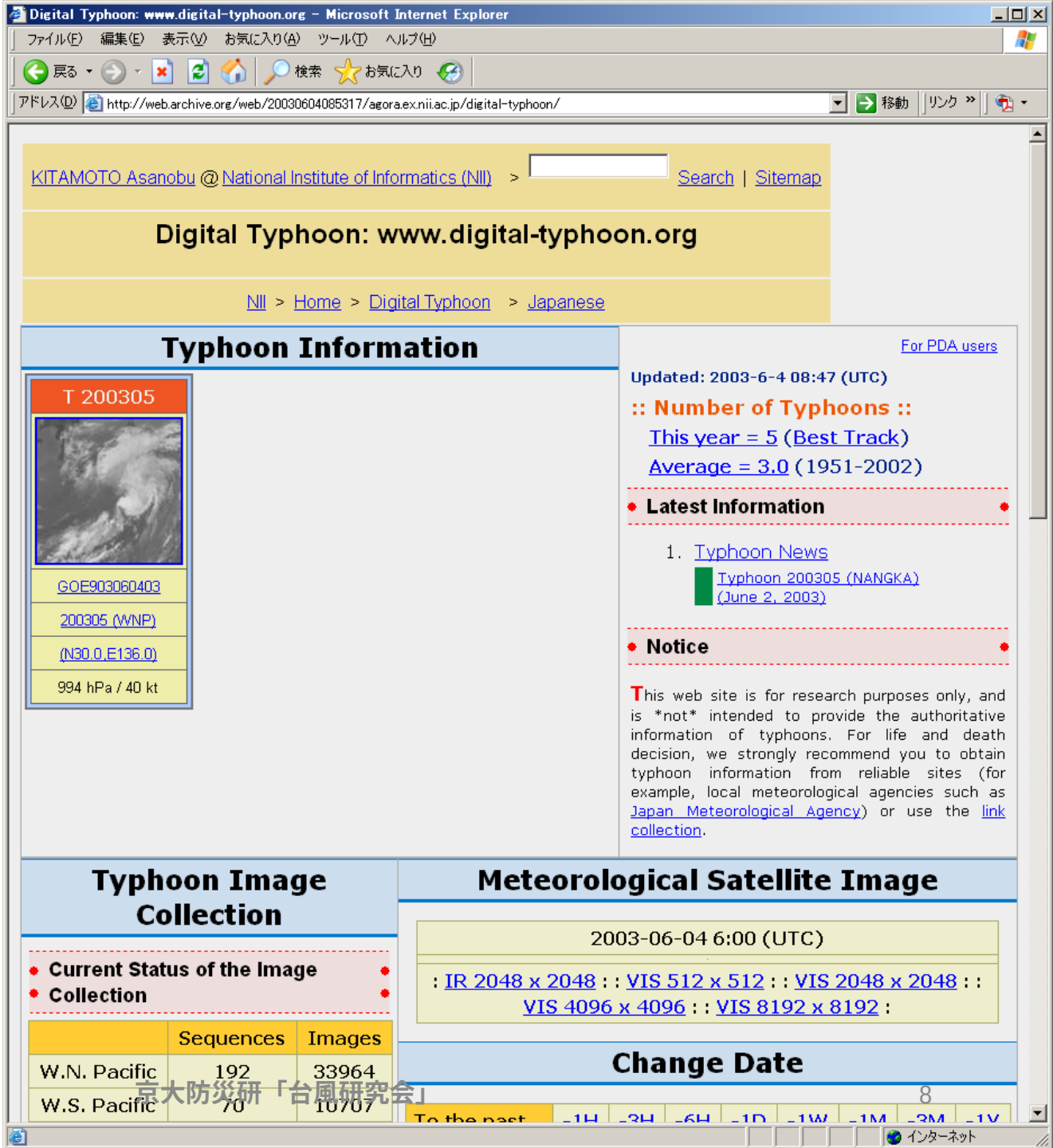
「デジタル台風」の歴史

http://agora.ex.nii.ac.jp/digital-typhoon/history/ten-years/

- 1999年4月22日：アイデアを最初に議論した日。
- 2000年7月4日：digital-typhoon.orgドメイン取得。
- 2001年末：画像検索研究のデモサイトとして公開。当初は過去のデータのみ。
- 2002年秋～2003年1月：リアルタイムデータ提供を検討。
- 2003年6月ごろ：ウェブサイトが現在のデザインに近づく。
- 2003年7月：Yahoo!等に登録し全面公開。
- 2006年7月：ひまわりデータを1981年まで拡大。
- 2007年12月：「台風前線」メディアアート受賞。

2003年06月

Internet Archive
[WayBack Machine]
http://www.archive.org/

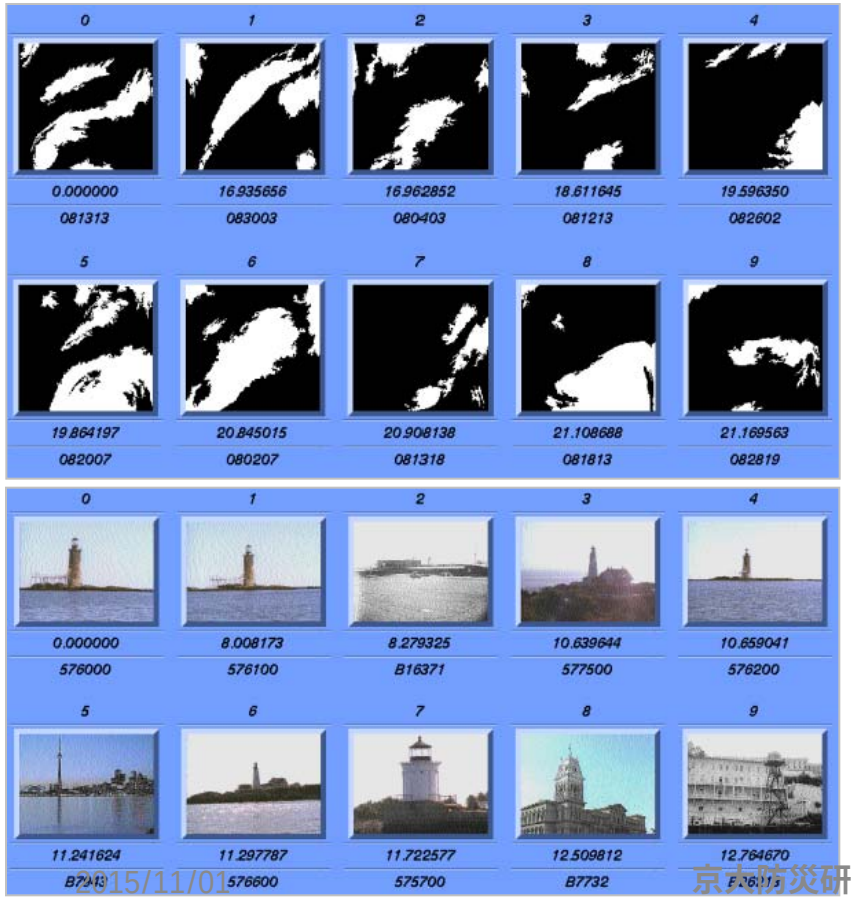


2005年10月

Internet Archive
[WayBack Machine]
http://www.archive.org/

画像内容に基づく検索の研究

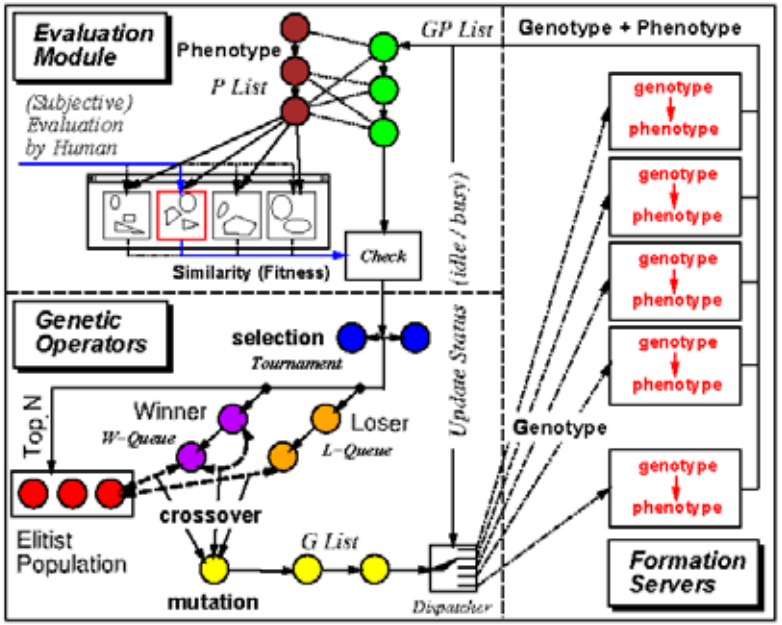
階層モデルに基づく類似画像検索 (1992-1997)



- **大学院**：気象衛星画像の内容に基づく検索の研究に取り組んだ。
- **課題**：検索結果の評価が難しい。何をもって「成功」と言えるか？
- **評価**：通常は「標準データセット」を使うが、それがない。
- 研究に行き詰まった…

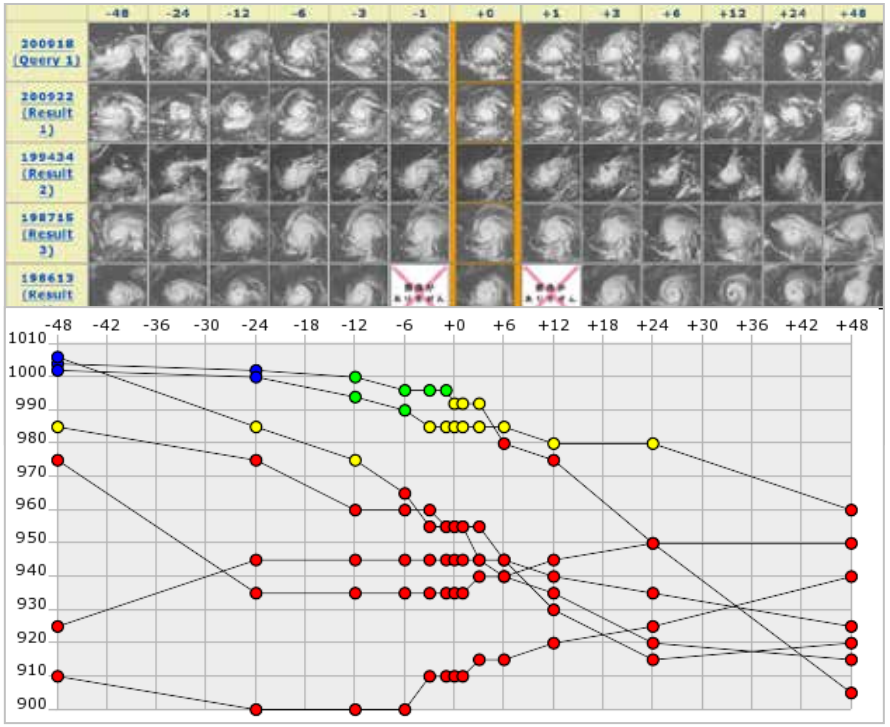
画像検索評価問題の解決策

対話型遺伝的アルゴリズムを用いた画像散策 relevance feedback (1996-1998)



評価の対象は画像そのものではなく散策プロセス。

シミュレーションと統合した予測性能評価（1999-）



「客観的」な評価基準、成功の実用的な定義も明確。

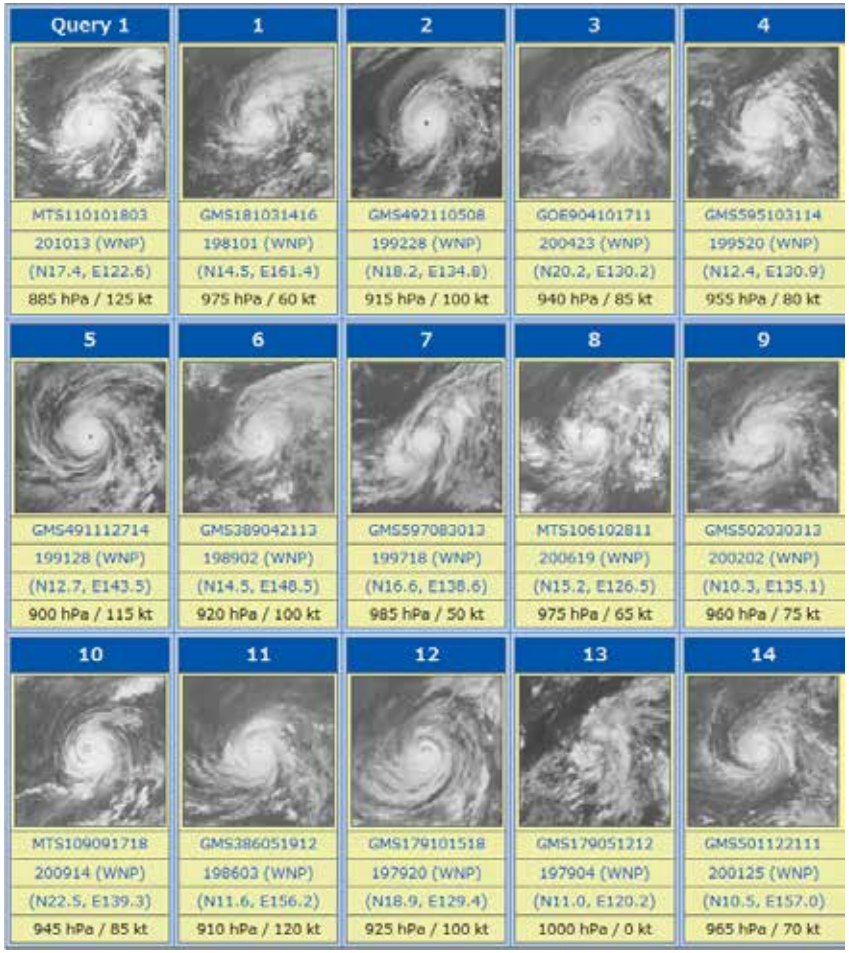
1999年の無謀なアイデア

画像検索の精度評価＝客観的な基準を決めることが難しい。

- **台風解析＆予測：ドボラック法と台風ボーガスという、怪しげな（？）手法を使っている。**
- **精度評価：検索結果を基に台風ボーガスを修正し、予測精度の向上を評価できないか？**

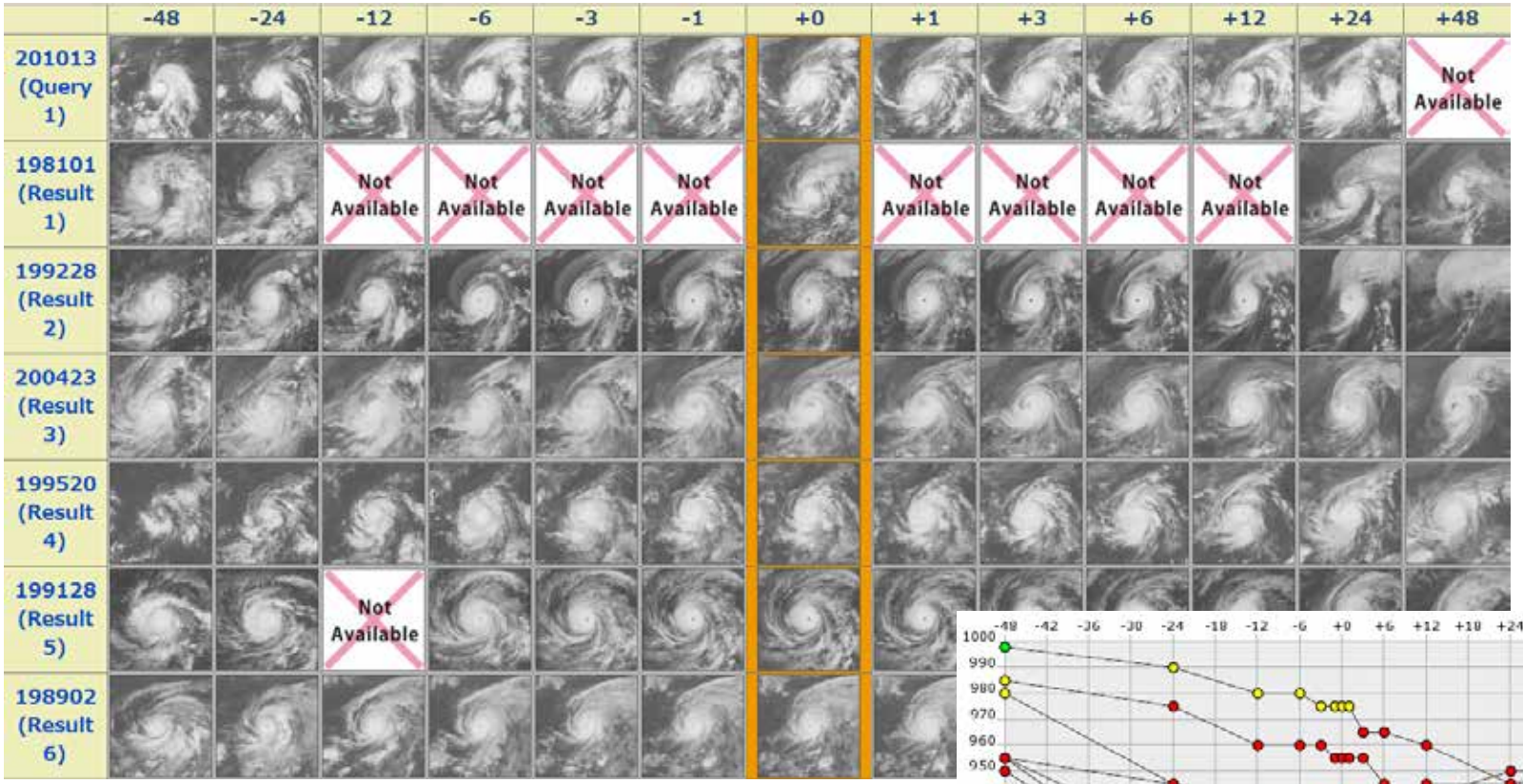
過去事例に基づくスマートな補間により、初期値とデータ同化を改善する。

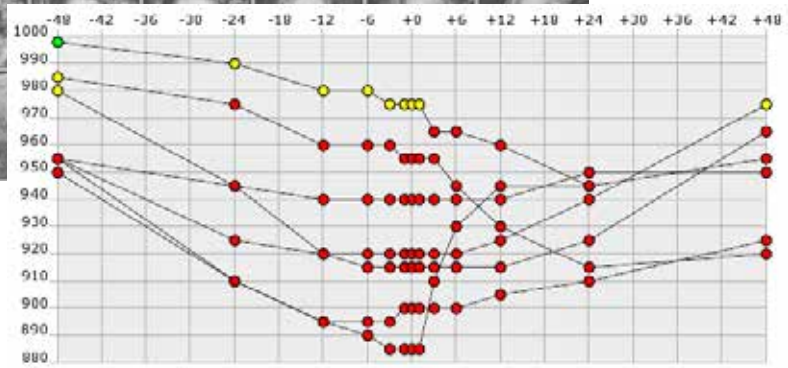
画像類似度に基づく検索



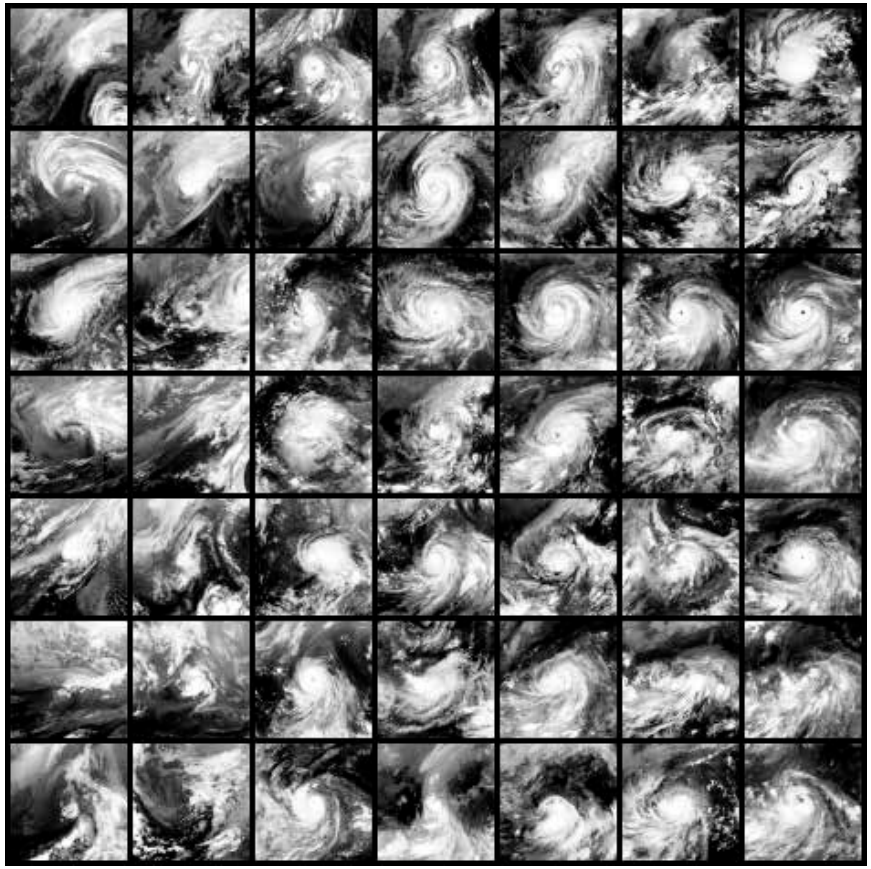
- 台風画像を検索キーとし、**類似したパターンをもつ過去の画像を高速に検索**。
- 本来は複雑な特徴量（**ドボラック法**）とすべきだが、現状では簡単な画像特徴量を利用。

時系列検索と比較


-48
-24
-12
-6
-3
-1
+0
+1
+3
+6
+12
+24
+48
201013 (Query 1)
198101 (Result 1)
199228 (Result 2)
200423 (Result 3)
199520 (Result 4)
199128 (Result 5)
198902 (Result 6)
Not Available

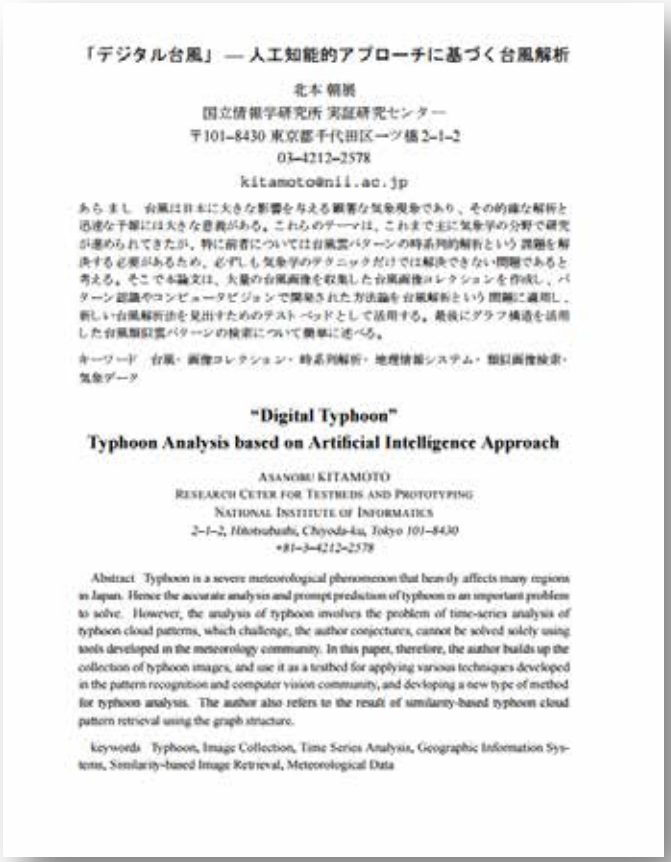

-48 -42 -36 -30 -24 -18 -12 -6 +0 +6 +12 +18 +24 +30 +36 +42 +48
1000
990
980
970
960
950
940
930
920
910
900
890
880

網羅的な台風画像コレクションから台風情報の網羅的データベースへ



- **客観的評価のために、統一的な方法で、可能な限り網羅的に、画像を収集する必要がある。**
- 関連データも収集。
- リアルタイムに収集。
- メモをブログ公開。
- もはや当初の目的を踏み外している？？

2000年頃、反応がなかった



「デジタル台風」 — 人工知能的アプローチに基づく台風解析

北本 朝展
国立情報学研究所 実証研究センター
〒101–8430 東京都千代田区一ツ橋 2–1–2
03–4212–2578
kitamoto@nii.ac.jp

あらまし　台風は日本に大きな影響を与える顕著な気象現象であり，その的確な解析と迅速な予報には大きな意義がある。これらのテーマは，これまで主に気象学の分野で研究が進められてきたが，特に前者については台風雲パターンの時系列的解析という 課題を解決する必要があるため，必ずしも気象学のテクニックだけでは解決できない問題であると考える。そこで本論文は，大量の台風画像を収集した台風画像コレクションを作成し，パターン認識やコンピュータビジョンで開発された方法論を台風解析という 問題に適用し，新しい台風解析法を見出すためのテスト ベッドとして活用する。最後にグラフ構造を活用した台風類似雲パターンの検索について簡単に述べる。

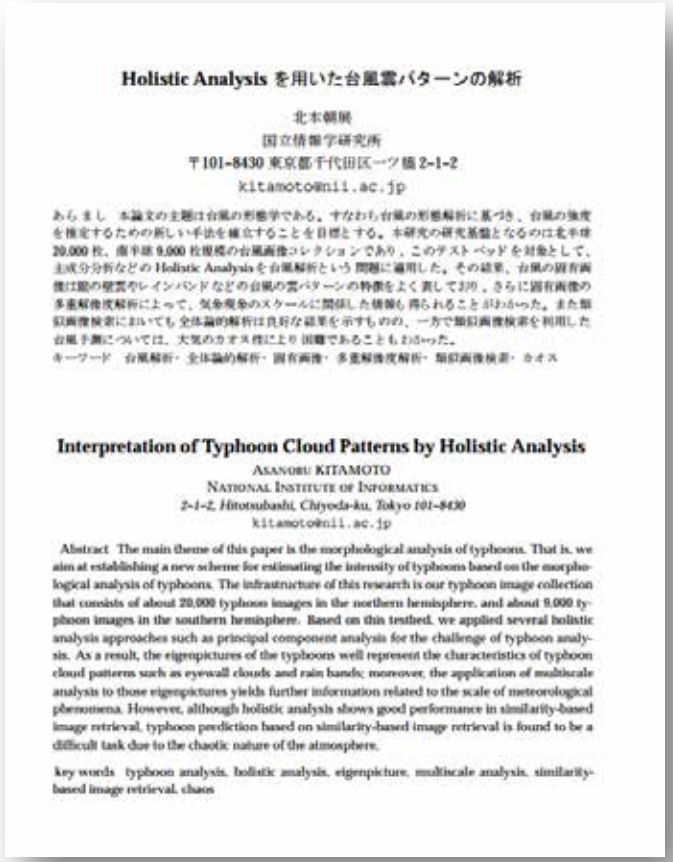
キーワード　台風・画像コレクション・時系列解析・地理情報システム・類似画像検索・気象データ

"Digital Typhoon"
Typhoon Analysis based on Artificial Intelligence Approach

Asanobu KITAMOTO
Research Cuter for Testbeds and Prototyping
National Institute of Informatics
2–1–2, Hitotsubashi, Chiyoda-ku, Tokyo 101–8430
+81–3–4212–2578

Abstract　Typhoon is a severe meteorological phenomenon that heavily affects many regions in Japan. Hence the accurate analysis and prompt prediction of typhoon is an important problem to solve. However, the analysis of typhoon involves the problem of time-series analysis of typhoon cloud patterns, which challenge, the author conjectures, cannot be solved solely using tools developed in the meteorology community. In this paper, therefore, the author builds up the collection of typhoon images, and use it as a testbed for applying various techniques developed in the pattern recognition and computer vision community, and developing a new type of method for typhoon analysis. The author also refers to the result of similarity-based typhoon cloud pattern retrieval using the graph structure.

keywords　Typhoon, Image Collection, Time Series Analysis, Geographic Information Systems, Similarity-based Image Retrieval, Meteorological Data



Holistic Analysis を用いた台風雲パターンの解析

北本朝展
国立情報学研究所
〒101–8430 東京都千代田区一ツ橋 2–1–2
kitamoto@nii.ac.jp

あらまし　本論文の主題は台風の形態学である。すなわち台風の形態解析に基づき，台風の強度を推定するための新しい手法を確立することを目標とする。本研究の研究基盤となるのは北半球 20,000 枚，南半球 9,000 枚規模の台風画像コレクションであり，このテスト ベッドを対象として，主成分分析などの Holistic Analysis を台風解析という 問題に適用した。その結果，台風の固有画像は眼の壁雲やレインバンドなどの台風の雲パターンの特徴をよく表しており，さらに固有画像の多重解像度解析によって，気象現象のスケールに関係した情報も 得られることがわかった。また類似画像検索においても 全体論的解析は良好な結果を示すものの，一方で類似画像検索を利用した台風予測については，大気のカオス性により 困難であることもわかった。

キーワード　台風解析・全体論的解析・固有画像・多重解像度解析・類似画像検索・カオス

Interpretation of Typhoon Cloud Patterns by Holistic Analysis

Asanobu KITAMOTO
National Institute of Informatics
2–1–2, Hitotsubashi, Chiyoda-ku, Tokyo 101–8430
kitamoto@nii.ac.jp

Abstract　The main theme of this paper is the morphological analysis of typhoons. That is, we aim at establishing a new scheme for estimating the intensity of typhoons based on the morphological analysis of typhoons. The infrastructure of this research is our typhoon image collection that consists of about 20,000 typhoon images in the northern hemisphere, and about 9,000 typhoon images in the southern hemisphere. Based on this testbed, we applied several holistic analysis approaches such as principal component analysis for the challenge of typhoon analysis. As a result, the eigenpictures of the typhoons well represent the characteristics of typhoon cloud patterns such as eyewall clouds and rain bands; moreover, the application of multiscale analysis to those eigenpictures yields further information related to the scale of meteorological phenomena. However, although holistic analysis shows good performance in similarity-based image retrieval, typhoon prediction based on similarity-based image retrieval is found to be a difficult task due to the chaotic nature of the atmosphere.

key words　typhoon analysis, holistic analysis, eigenpicture, multiscale analysis, similarity-based image retrieval, chaos

北本 朝展, "**「デジタル台風」 -- 人工知能的アプローチに基づく台風解析**", **情報処理学会技術報告**, Vol. CVIM123-8, pp. 59-66, **2000年09月**

北本 朝展, "Holistic Analysis**を用いた台風雲パターンの解析**", **電子情報通信学会技術報告**, Vol. PRMU2000-240, pp. 129-136, **2001年03月**

2015年、反応があった

北本 朝展, "デジタル台風：熱帯低気圧の実感に向けたデータの文脈化", メソ気象研究会, 2015年10月27日

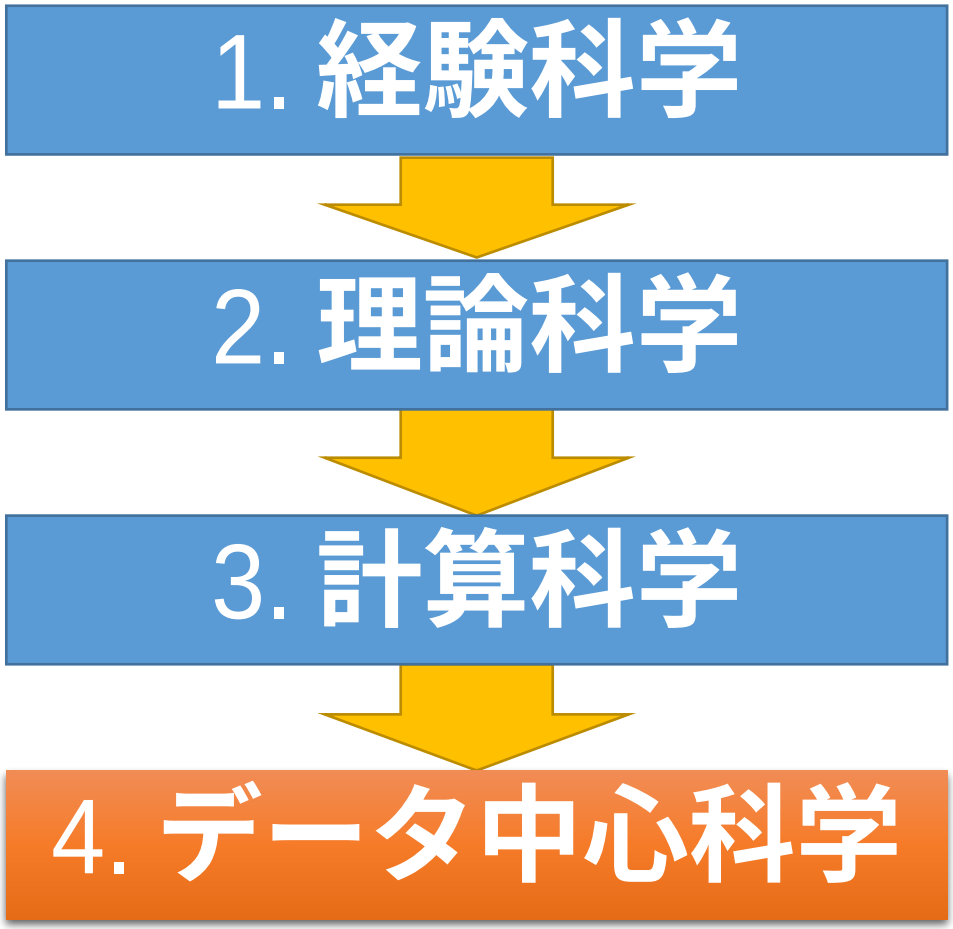
北本 朝展, "デジタル台風：複合災害の現状認識を目指すビッグデータ解析", 平成27年度京都大学防災研究所共同研究集会「台風研究会」, 2015年11月 1日

15年間の歩み

1. **気象衛星「ひまわり」の網羅的データベース。**
2. **アメダスやレーダーなど気象データへの拡大。**
3. **ニュース記事などテキストデータへの拡大。**
4. **ツイッターなどソーシャルメディアへの拡大。**
5. **気象庁防災情報XMLなど防災データへの拡大。**
6. **GPVなど数値予報データへの拡大。**
7. **伊勢湾台風や天気図など歴史データへの拡大。**
8. **コラボレーションデータへの拡大（今後）。**

ビッグデータ解析

科学のパラダイム

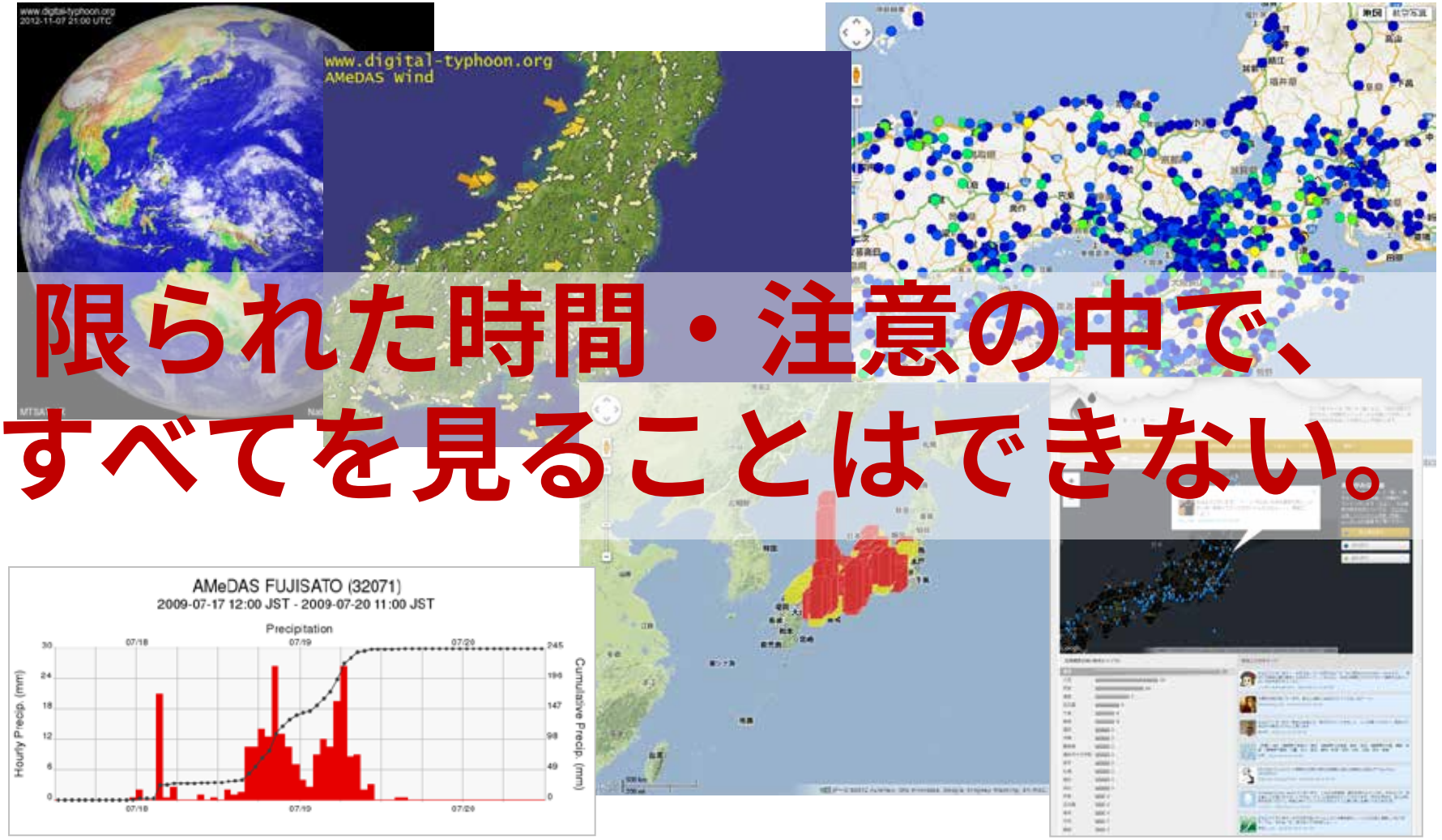


- 気象学は、**計算科学**が中心の研究が多く、**演繹**が強い分野。
- シミュレーションデータはビッグデータか？
- データ中心科学の関心は**帰納**。

ビッグデータとは？

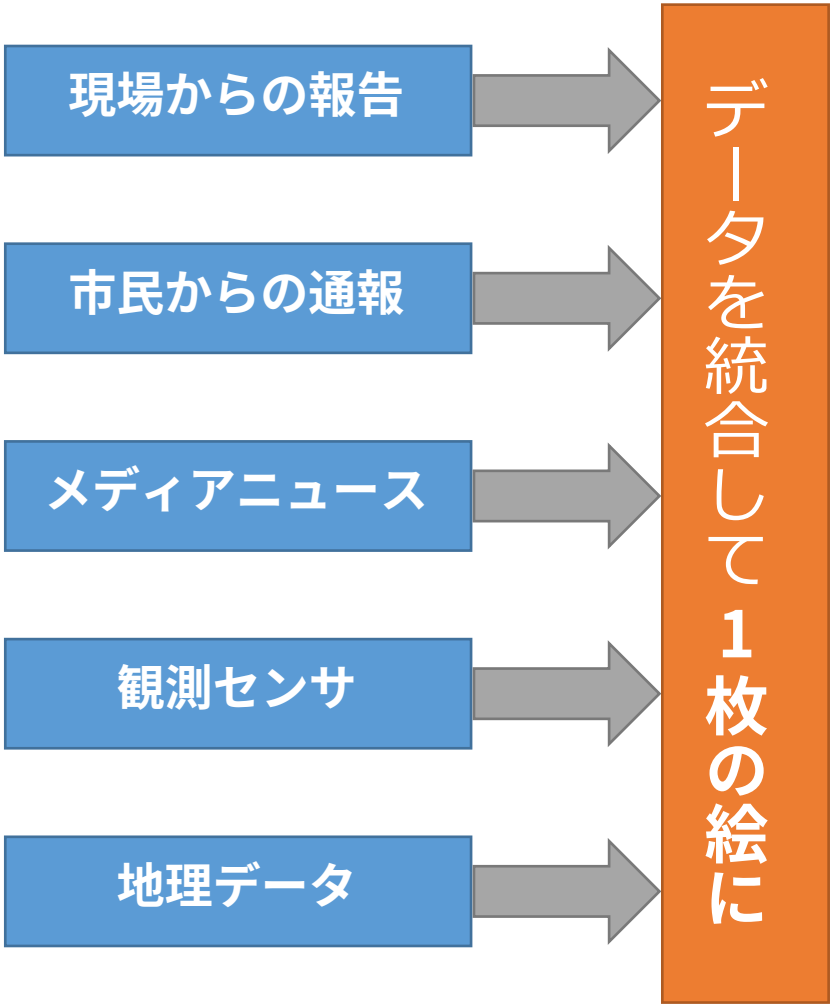
Volume 量	Big Data は「巨大なデータ」だけを意味するわけではない。
Variety 多様性	多様な形式のデータ。テキストなどの非構造データも含む。
Velocity 速度	データの「掛け合わせ」「読み替え」から新しい価値を生み出す。

クライシス時の情報爆発



限られた時間・注意の中で、
すべてを見ることはできない。

状況認識の統一



いつ
どこで
何が
発生？

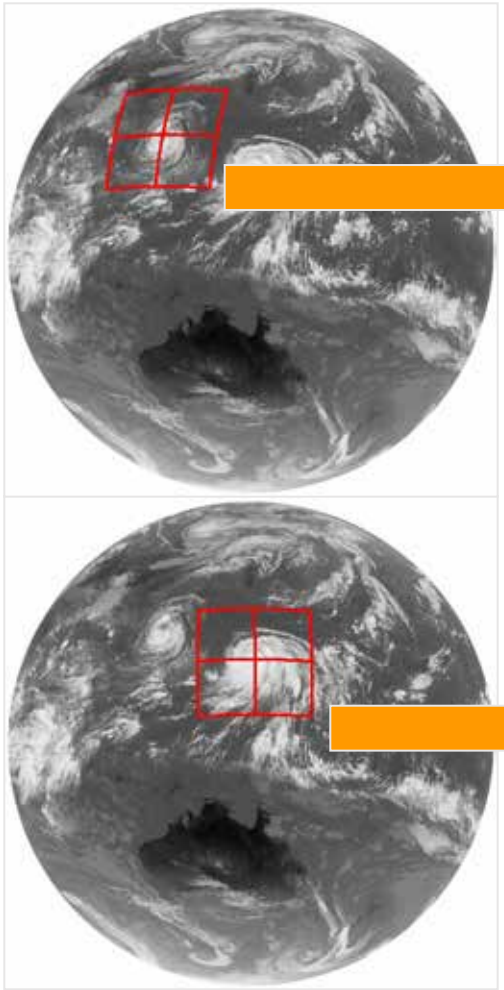
情報を1か所に集約し、1枚の絵を共有して現状を認識する。

気象データ

気象衛星データ

- **1995年以降**：東大生産技術研究所でアーカイブしていたひまわりデータが中心。
- **1981年～1995年**：気象業務支援センターから2005年頃に「大人買い」したデータ。
- **1978年～1979年**：どこかから「発見」されたデータを気象業務支援センターから追加購入。
- **最新データ**：気象業務支援センターから購入。
- **ひまわり8号データ**：NICTからダウンロードして利用。

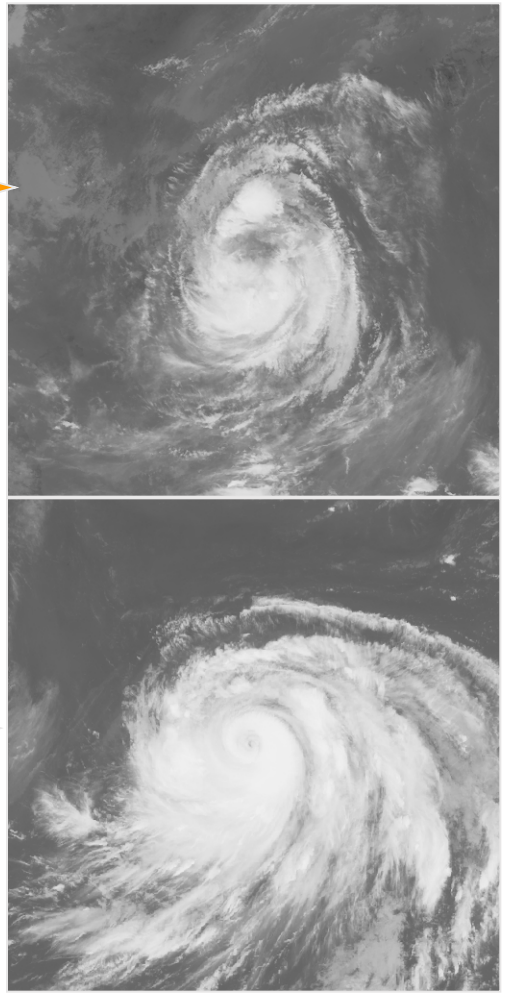
ラグランジアン的台風画像



T 200513
970hPa

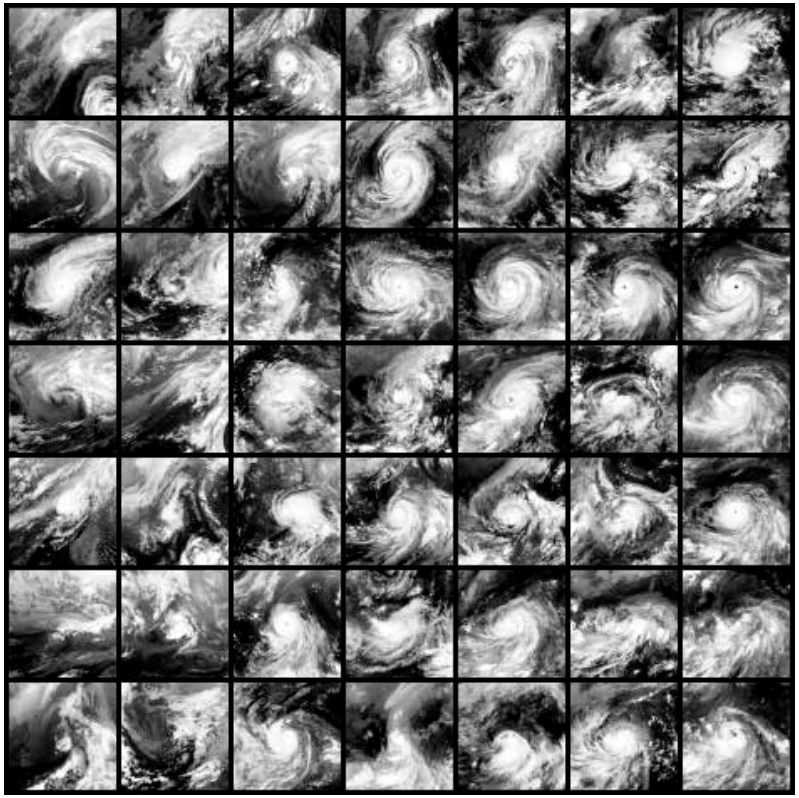
ランバート正積方位図法：円形の歪みが少なく、かつ面積が正しい図法を利用。
サイズは一辺2600km。

T 200514
935hPa

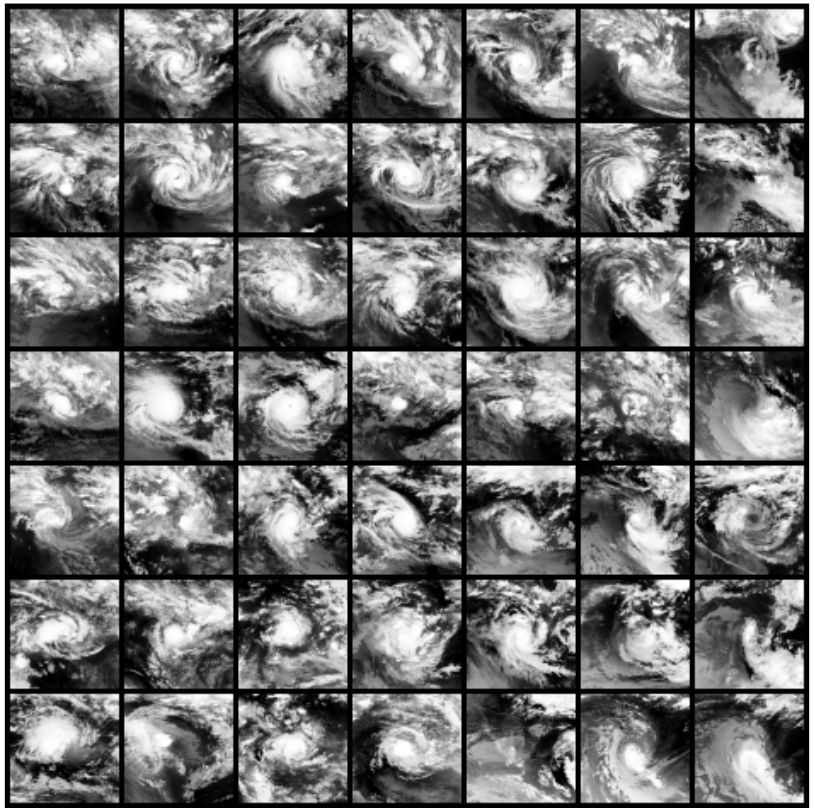


熱帯低気圧画像コレクション

1978年以来の画像数は、154,000件（北半球）／35,000件（南半球）



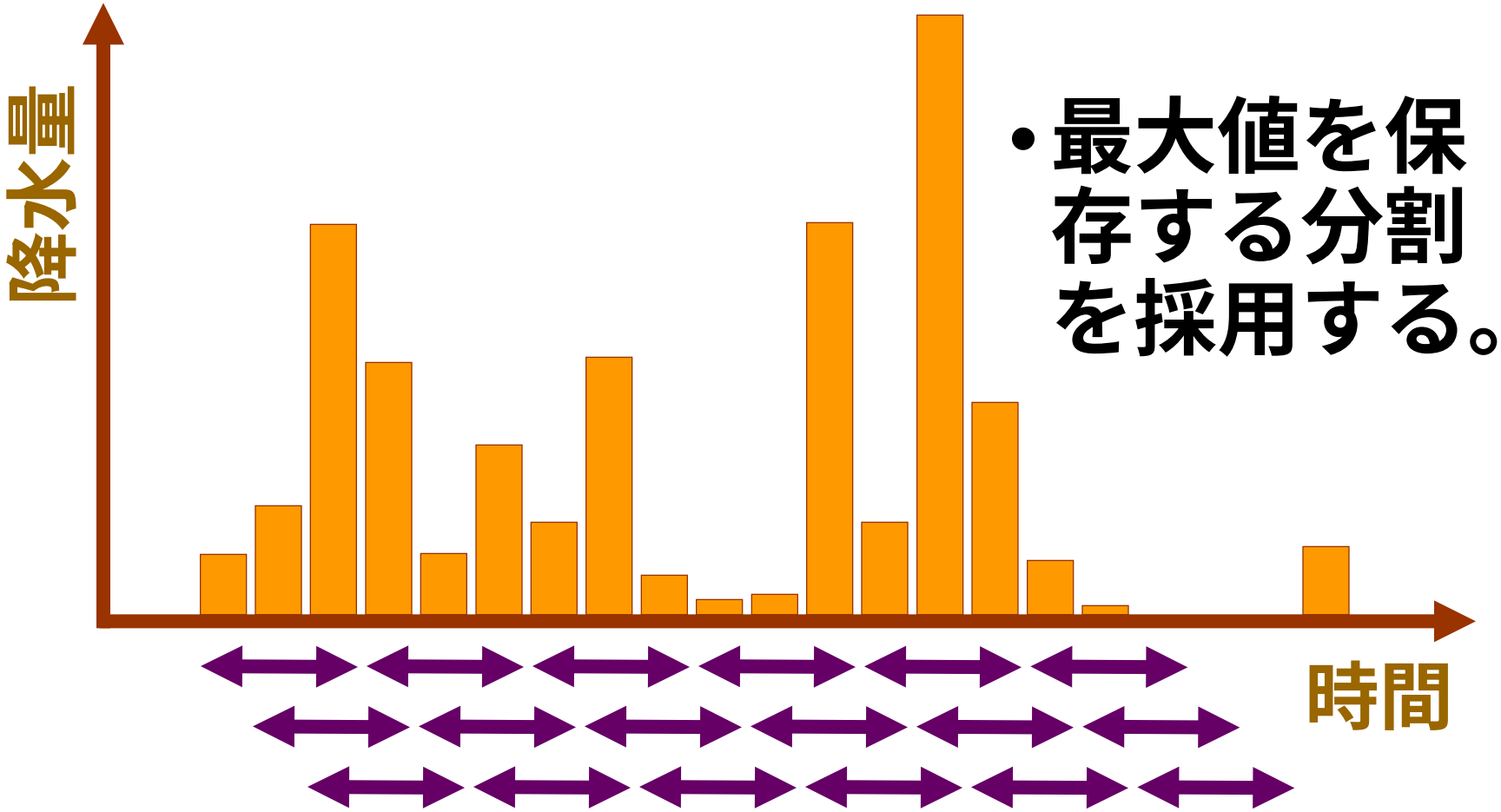
北半球（台風）



南半球（サイクロン）

N時間降水量ランキング

http://agora.ex.nii.ac.jp/digital-typhoon/heavy_rain/



- 最大値を保存する分割を採用する。

アメダス・ランキング・サービス

http://agora.ex.nii.ac.jp/digital-typhoon/ameras/

n 史上ランキング（その地点で史上何位の降水量か？）

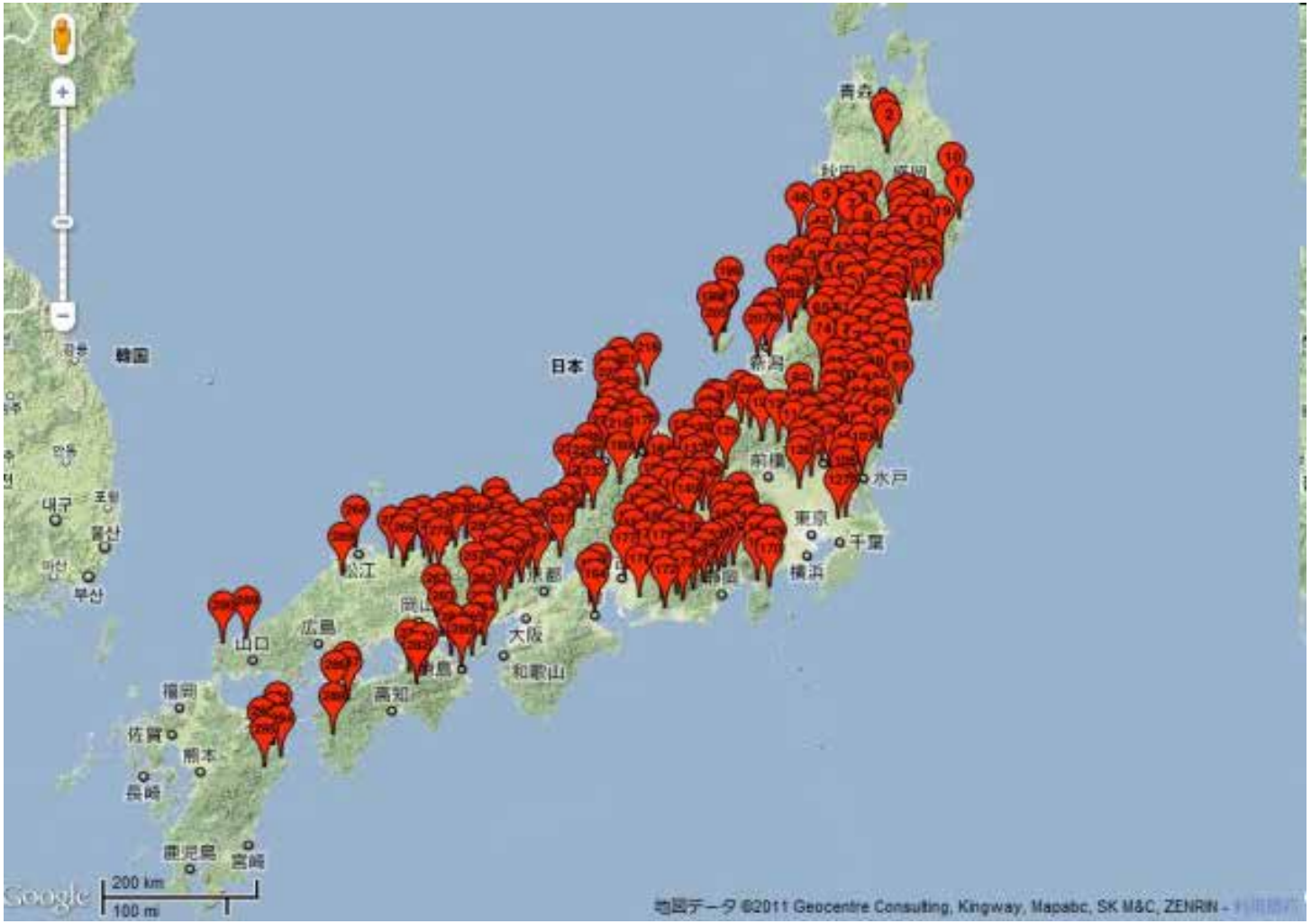
番号	名前	所在地	観測値	ランキング
92006	北大東	沖縄県島尻郡北大東村	119.5	3位(182件中)
11176	豊富	北海道天塩郡豊富町	42.5	129位(2,575件中)
21276	鵡川	北海道勇払郡むかわ町	45	145位(2,095件中)
12011	中川	北海道中川郡中川町	35.5	199位(2,591件中)
92012	旧東	沖縄県島尻郡南大東村	53	15位(180件中)

n 同時ランキング（その時点で全国何位の降水量か？）

番号	名前	所在地	観測値	ランキング
92006	北大東	沖縄県島尻郡北大東村	119.5	1位(1,321件中)
92012	旧東	沖縄県島尻郡南大東村	53	2位(1,321件中)
92011	南大東	沖縄県島尻郡南大東村	49	3位(1,321件中)
23321	知内	北海道上磯郡知内町	47	4位(1,321件中)
31482	酸ケ湯	青森県青森市	45.5	5位(1,321件中)

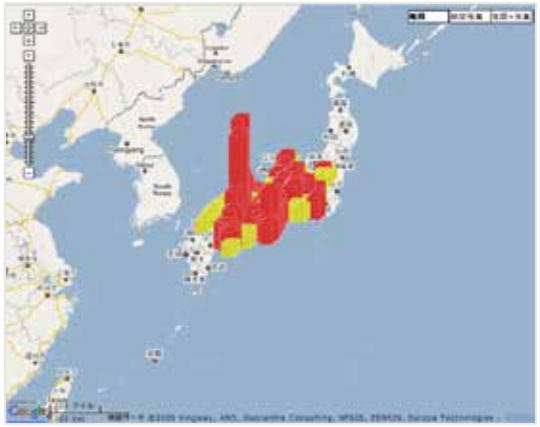
アメダスイベント検出

http://agora.ex.nii.ac.jp/digital-typhoon/ameras/event/



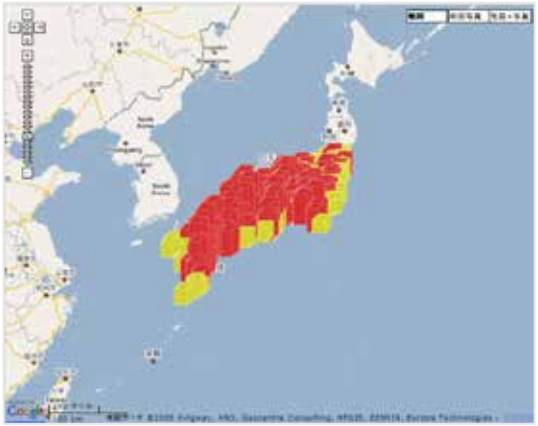
豪雨事例の類似検索

http://agora.ex.nii.ac.jp/digital-typhoon/heavy_rain/clustering/

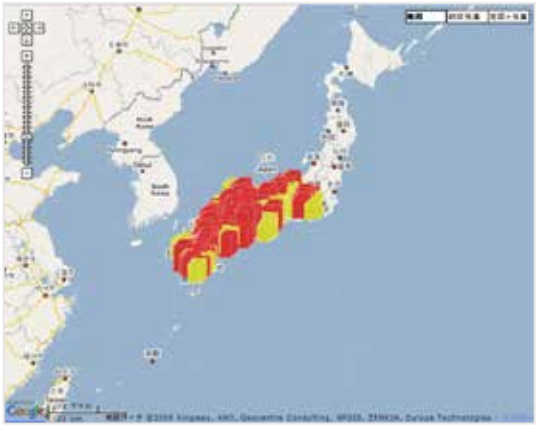


1998/09/20-1998/09/25

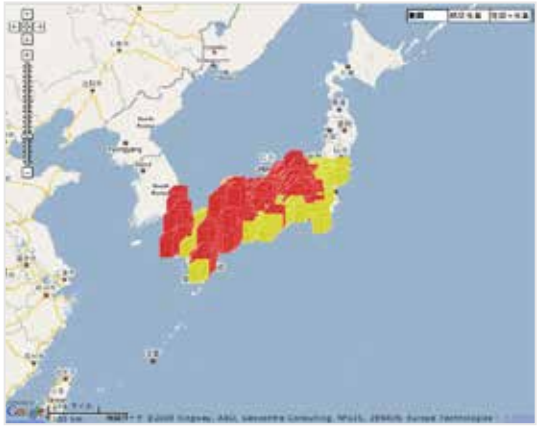
- 大雨地点分布をキーとして、類似豪雨事例を検索する。
- 豪雨事例期間をキーとして、気象災害記録を検索する。



2004/10/18-2004/10/21

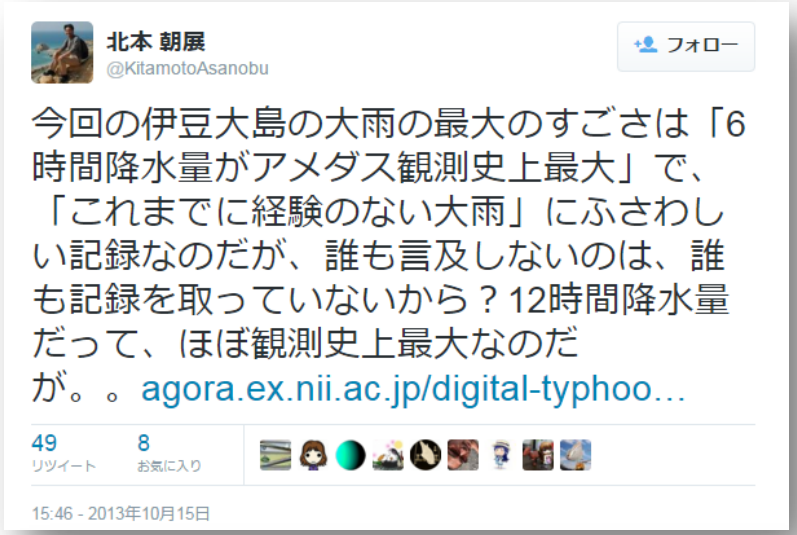


1998/10/14-1998/10/18



1983/09/25-1983/09/29

中期間降水量記録の不備



順位	観測所名	都道府県	降水量
1	大島	東京都	549.5
2	多良間	沖縄県	502.0
3	尾鷲	三重県	497.0
4	登別	北海道	467.0
5	尾鷲	三重県	463.5

- 2013年台風26号による大島豪雨。日本の全アメダス記録において、6時間では最大値、12時間でも最大値にほぼ並ぶ記録だった。
- 計算しようと思ってもすぐにはできない。
- 中期間の大雨に対する備えが不十分だった？

テキストデータ
ソーシャルメディア

固有表現抽出

富士小目井線：２年８カ月ぶり、観光ロード復旧－－日南の市道　／宮崎
http://headlines.yahoo.co.jp/hl?a=20070628-00000300-mailo-l45

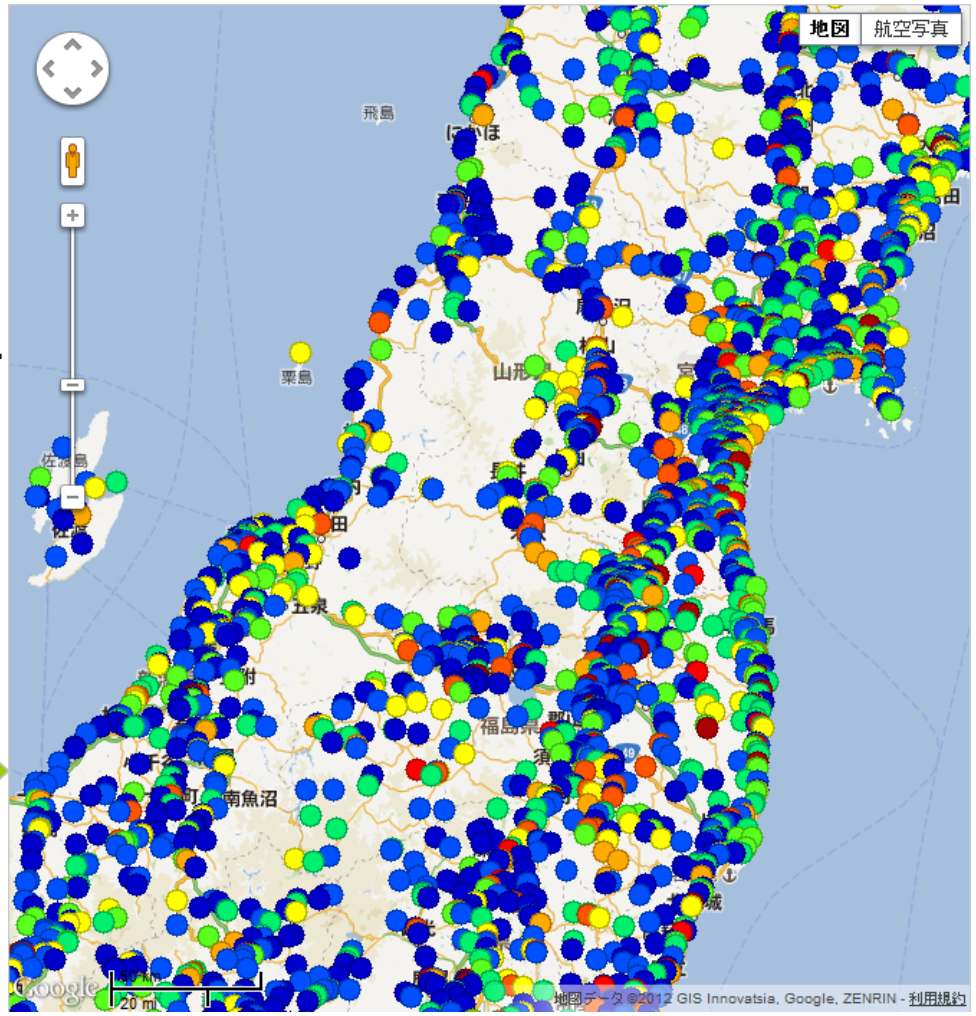
０４年１０月の台風２３号の豪雨による大規模な地滑りで崩壊した日南市富士の市道富士小目井線がこのほど、２年８カ月ぶりに復旧した。富士トンネルが開通するまで国道２２０号として使われ、沿線に観光名所のサボテンハーブ園（休園中）がある風光明媚（めいび）な道路として知られていた。市道は総延長約４キロ。このうち園から日南方向に岬を曲がった所で地滑りが発生し、橋げたを含む約４００メートルが土砂に埋まった。当初の査定では崩壊した橋げたの撤去費用など１３億７０００万円の工事費が見込まれたが、翌年９月の台風１４号の豪雨で橋げたがさらに海側に押し流されて復旧工事の邪魔にならなくなり、工事費は４億５０００万円に縮小された。（以下略）

- 台風200423号の抽出は容易だが、台風200514号の抽出はなかなか難しい。

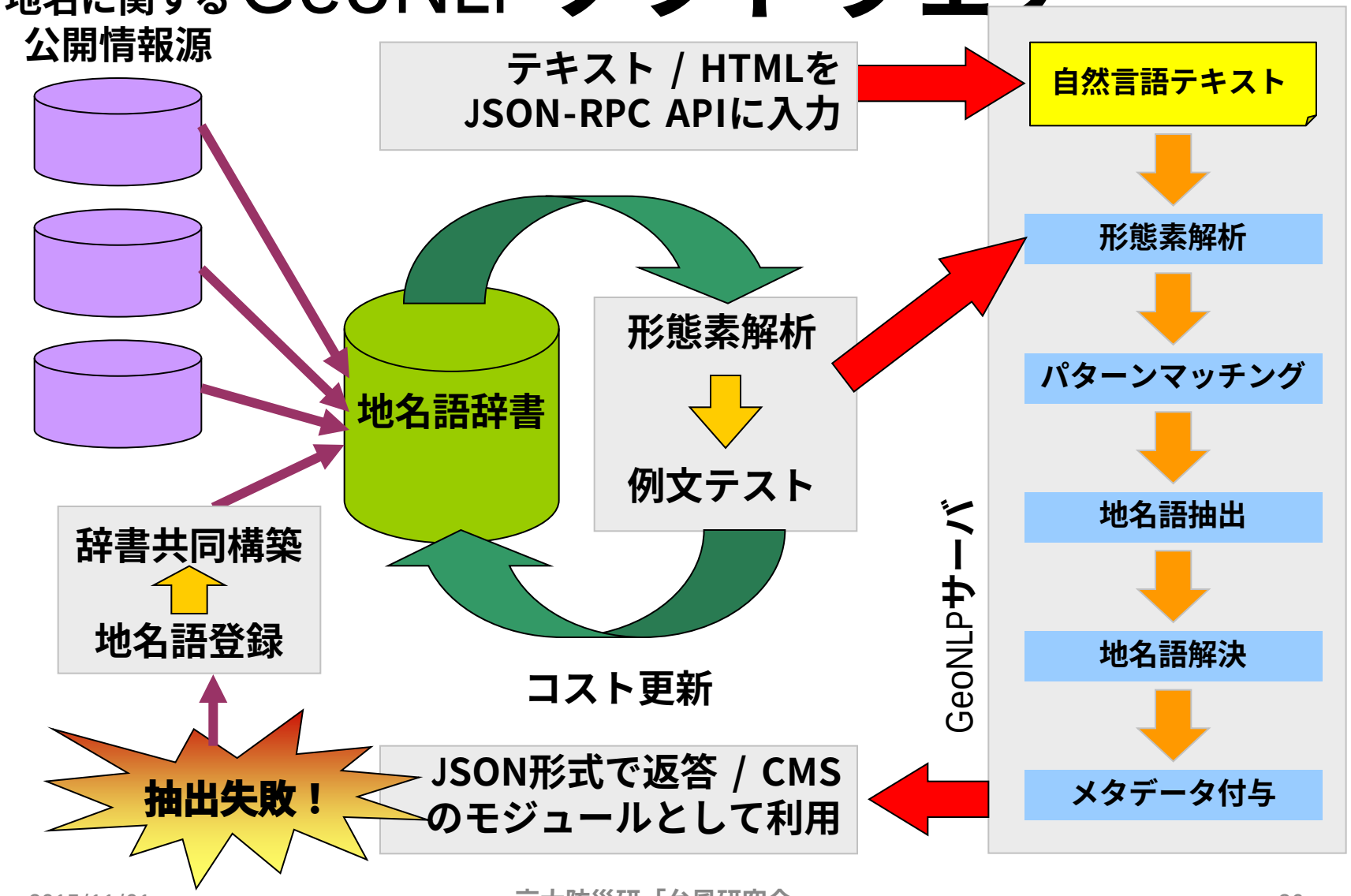
ジオタギング

宮城県などによると、気仙沼市や多賀城市などで大規模な火災が発生。気仙沼市は津波で市街地の3分の1が水没し、気仙沼湾に浮かぶ大島の4集落が壊滅状態となった。女川町もほぼ壊滅という。岩手県では陸前高田市がほぼ壊滅し、山田町や宮古市の一部が水没。福島県では南相馬市の約1800世帯が壊滅状態という。宮城県警は東松島市のJR野蒜駅付近で、この脱線した列車から乗客ら9人、別の列車からも11人を救出した。

- テキストから地名の候補を抽出。
- 同綴異義地名をさらに解析（例えば岩手県「宮古」市と沖縄県「宮古」島）。
- テキストの周囲に現れる他の地名を手掛かりに、複数候補の中から最適候補を決定。
- 地名辞書と照合して地名の緯度経度を検索。
- 地図上にマッピング。



GeoNLPソフトウェア



地名解析の問題

横浜は雨だよ。川崎は雪？

横浜は好きだよ。川崎は好き？

- 神奈川県横浜市と神奈川県川崎市？
- 青森県横浜町と福岡県川崎町？
- 横浜さんと川崎さん？
- 各種の曖昧さを解消し、実世界と紐付ける。

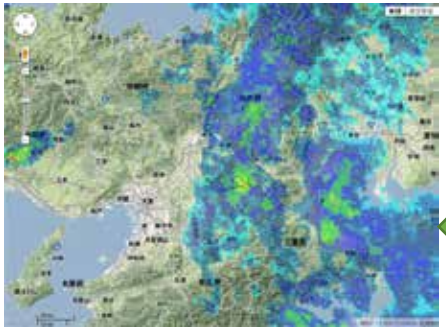
ソーシャル気象観測

キーワード
で検索

ソーシャル
データ

GeoNLPを用いた
地名ベースのジオタギング

(time, latitude, longitude, situation)

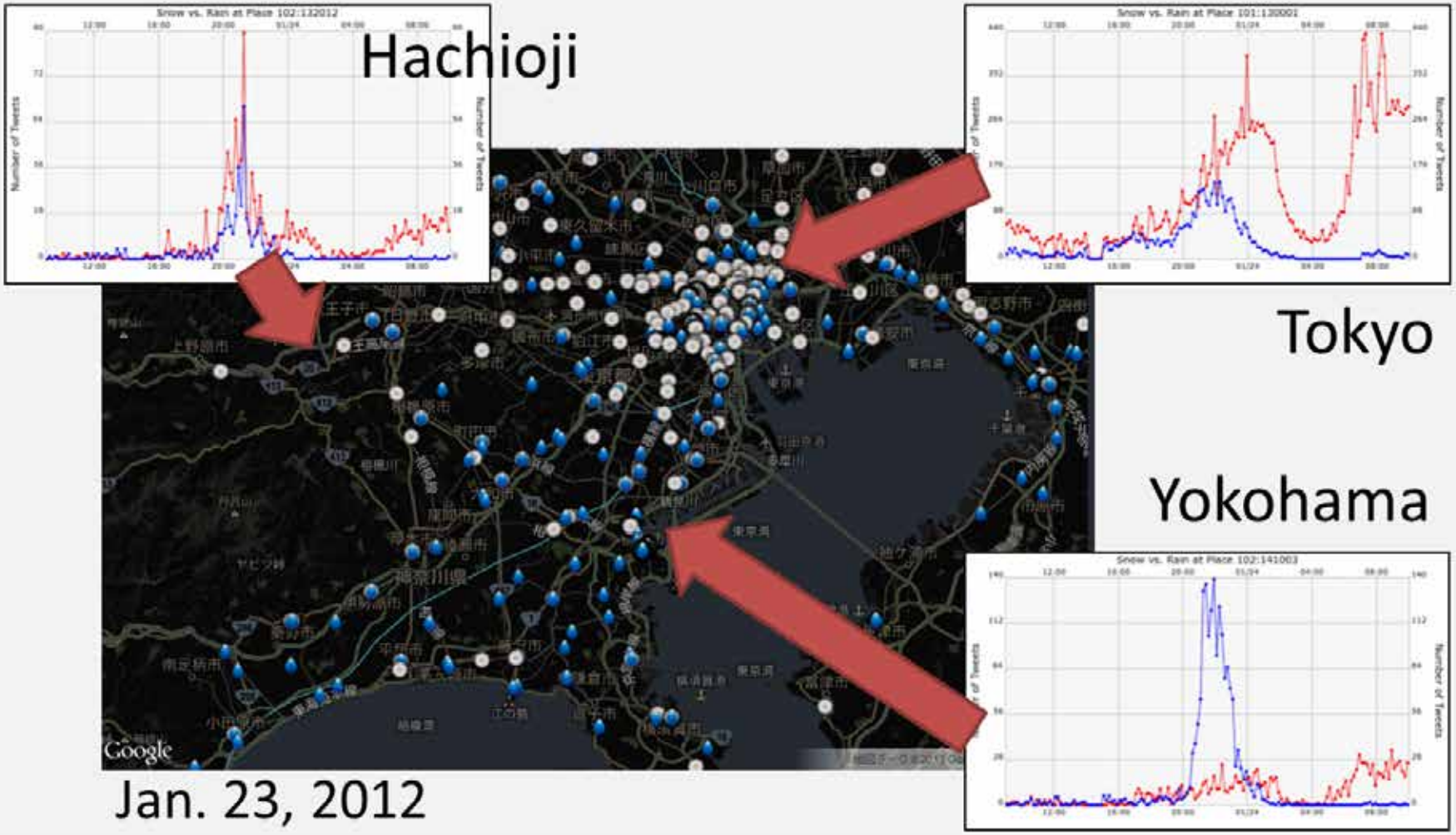


科学データ

ソーシャルおよび科学データストリームの降水状況を比較。

- 天気一般と比べると、降水の有無は目視でも確認しやすい。
- 気象レーダーは客観的な「グランドトゥルース」データとして使える。

Snow (white/red) and Rain (blue)



ふってきったー: http://agora.ex.nii.ac.jp/futtekitter/

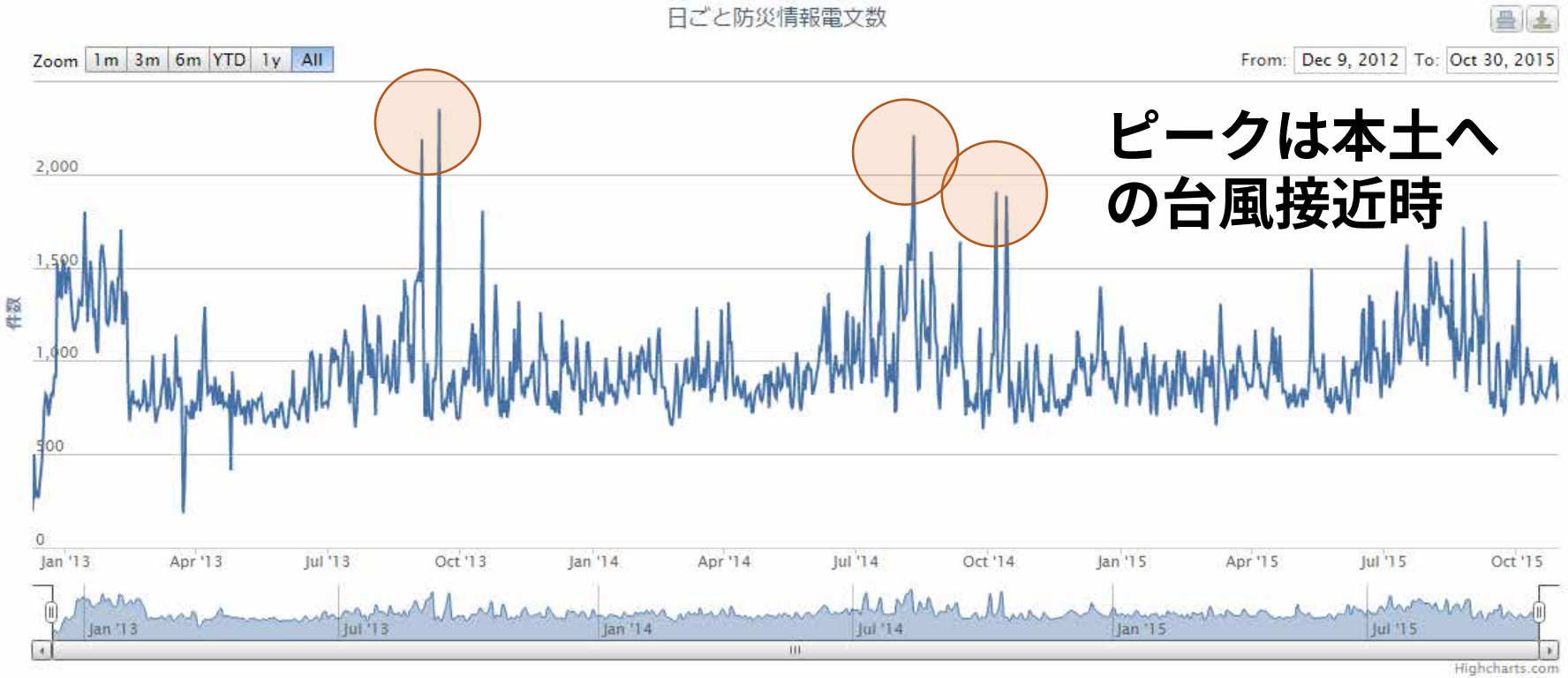
災害・防災データ

気象庁防災情報XML

http://agora.ex.nii.ac.jp/cps/weather/report/

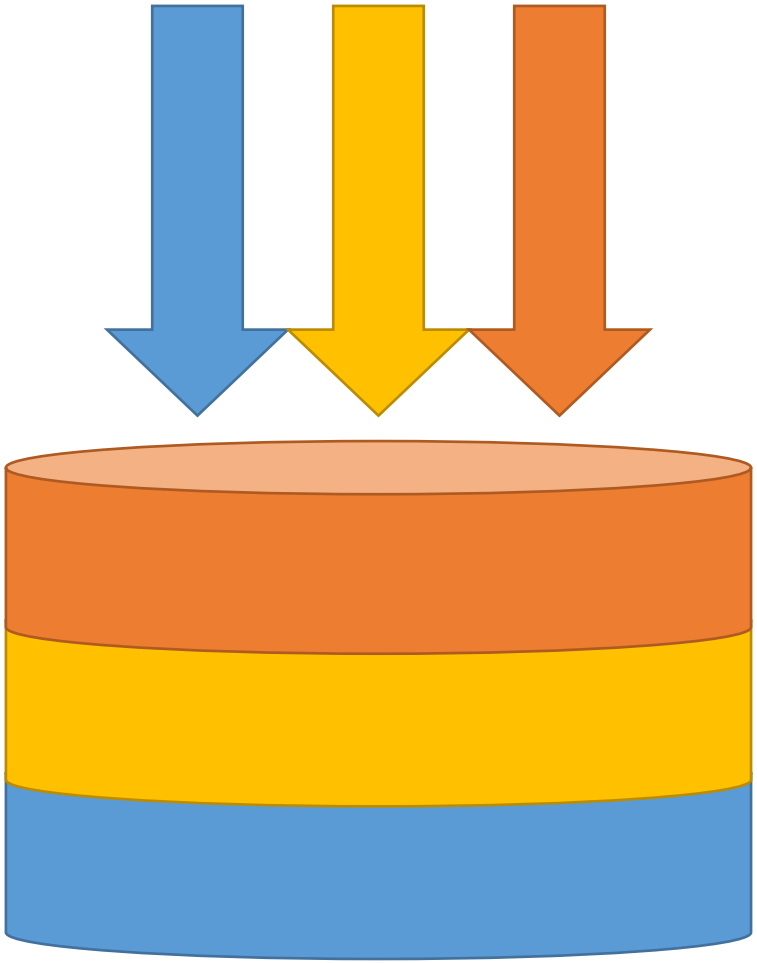
- XML形式を用いて、詳細で高度化された防災情報を提供。
- 2011年5月に本格運用開始。2012年12月に一般向けの試行的公開。
- PubSubHubbubを用いたプッシュ送信により、ほぼ**リアルタイムで防災情報を取得可能**。

防災情報電文の推移



2012年12月の試験配信開始以来、電文数は約101万件。最大流量は1日2000件超。平均流量は1日1000件弱。

フローからストックへの変換



- フロー情報を垂れ流すだけでは物足りない。
- **消えやすいフロー情報**をすくい上げ、**永続化してストック情報**へ。
- データベース化（検索）、頻度比較調査。
- モデルのバイアス等、**過去の検証**に利用。

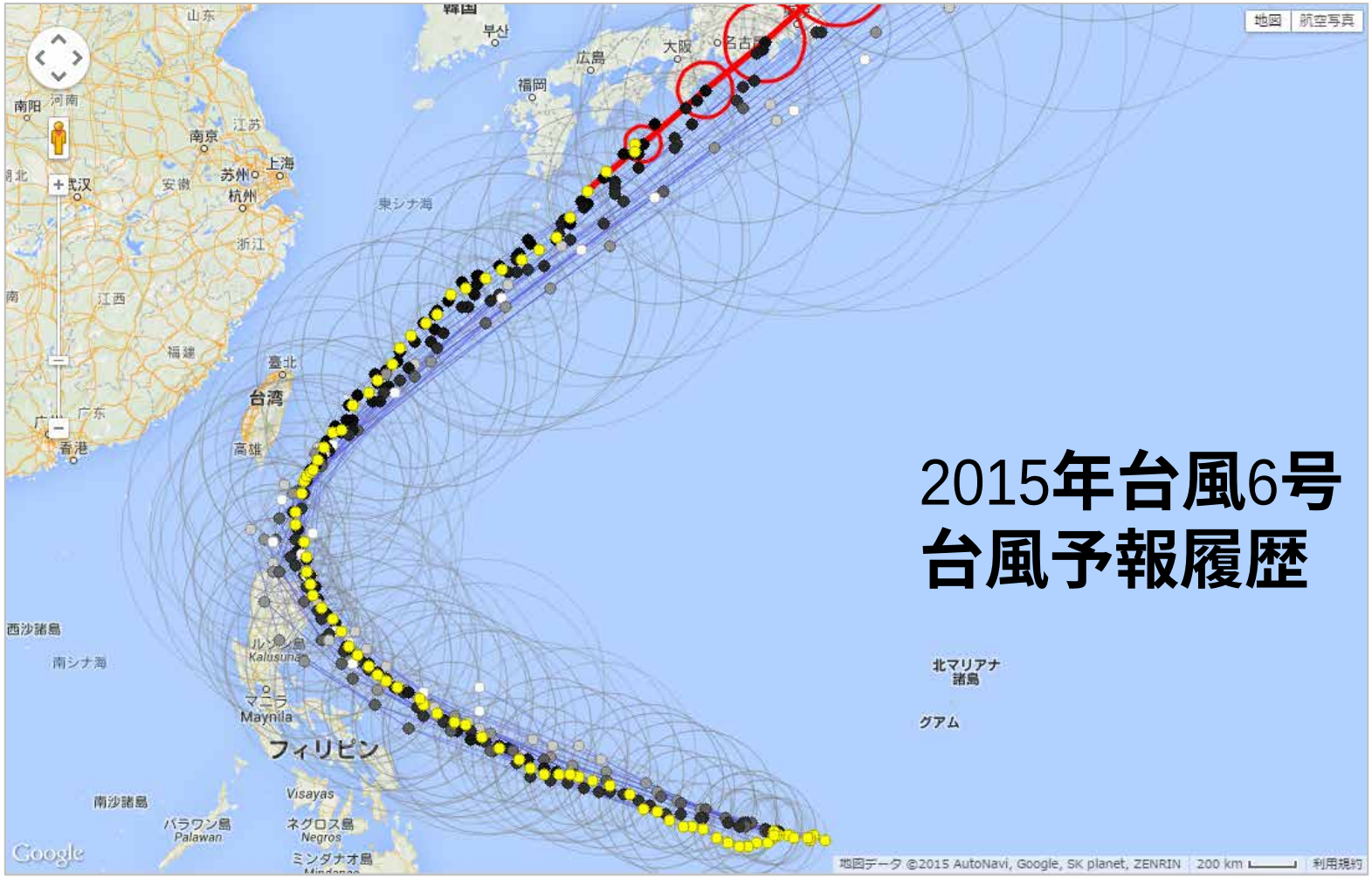
気象警報データベース

http://agora.ex.nii.ac.jp/cps/weather/warning/

- 全国を網羅的にカバーする気象警報データベースを構築し、**気象警報の再解析を可能とする**。
- 新しい電文を取得 ⇒ 現在の警報の状態と照合 ⇒ 警報ごと・地域ごとの発表・継続・解除を正しく対応付け ⇒ データベースを更新。
- **警報の継続には複数の電文が関係**するため、個々の電文を処理するだけでは対応できない。
- **対象電文数約24万件 ⇒ 発表（継続・解除）件数約1831万件 ⇒ 警報件数約230万件。**

台風予報履歴の可視化

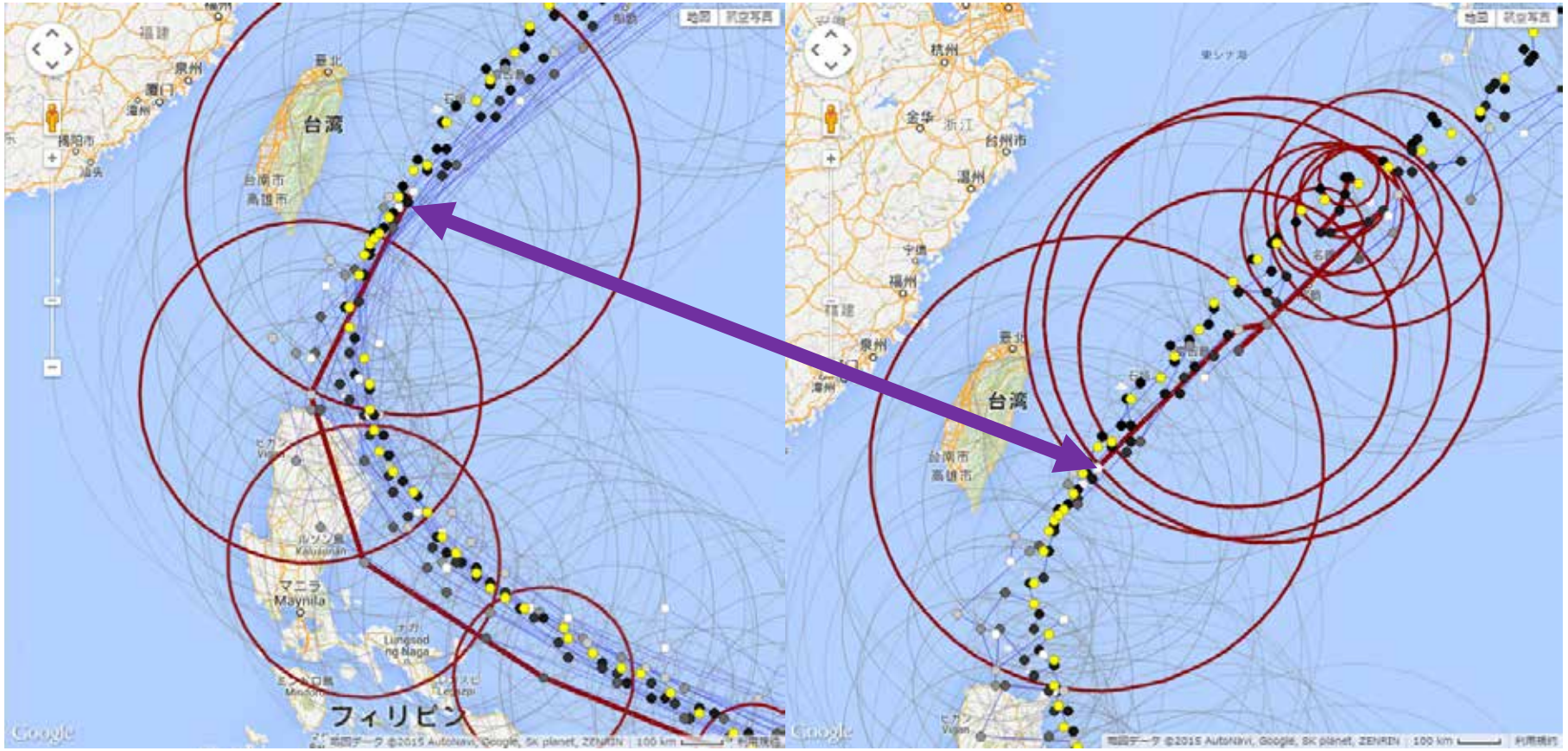
http://agora.ex.nii.ac.jp/digital-typhoon/summary/wnp/f/201506.html.ja



気象情報の2つの時間軸

発表時刻5月7日9時の120時間予報

予報時刻5月12日9時の予報



発表時刻固定

予報時刻固定

今後の研究課題

１．情報過多と情報過少

情報過多

- **問題**：情報量と複雑さの増加は、人間の処理能力を超える。
- **解決**：情報に意味的なタグを付与し、自動的に仕分ける。
- **手法**：時間・空間・テーマ軸での仕分け。

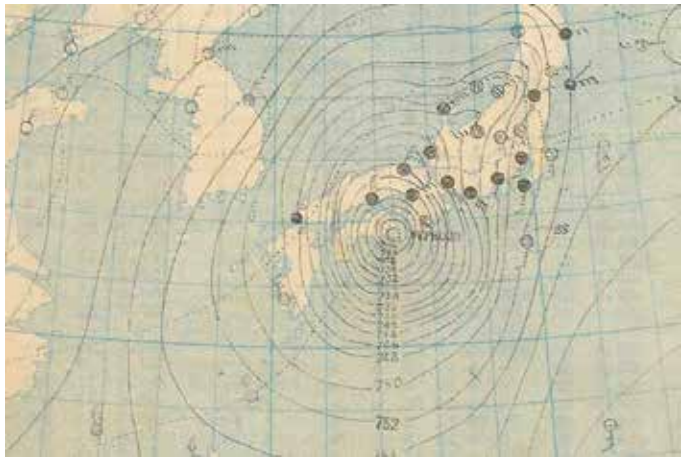
情報過少

- **問題**：一見多そうな情報も、細分化すると密度は高くない。
- **解決**：不特定多数の人々に情報収集を依頼し、不足を補う。
- **手法**：（空間的）クラウドソーシング。

クラウドソーシング

過去データのDB化

- 過去の天気図から台風などの情報を読み取り、データベース化するプロジェクト。



最新データの定点観測

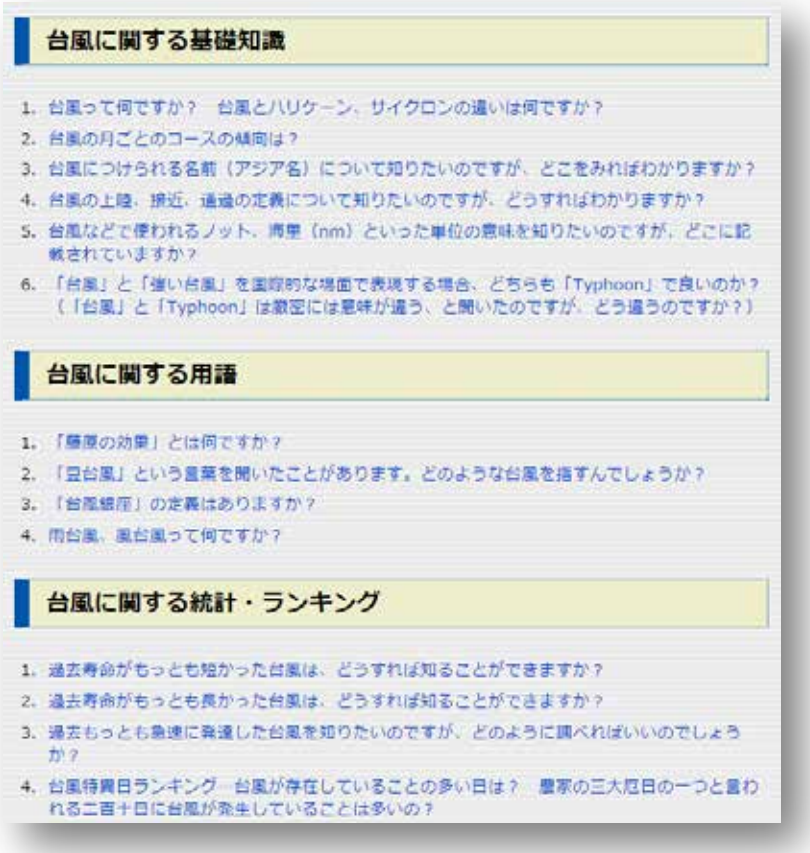
- スマホカメラで同じ場所を同じ構図で定点観測し、変化を計測するプロジェクト。

２．オープンサイエンス

	利便性	透明性	参加
オープンデータ	○	○	○
オープンアクセス	○	○	
オープンイノベーション	○		○
オープンガバメント		○	○
シチズンサイエンス	○		○
研究再現性		○	
データサイテーション	○		

コラボレーション

デジタル台風「レシピ集」



台風に関する基礎知識

1. 台風って何ですか？　台風とハリケーン、サイクロンの違いは何ですか？
2. 台風の月ごとのコースの傾向は？
3. 台風につけられる名前（アジア名）について知りたいのですが、どこをみればわかりますか？
4. 台風の上陸、接近、通過の定義について知りたいのですが、どうすればわかりますか？
5. 台風などで使われるノット、海里（nm）といった単位の意味を知りたいのですが、どこに記載されていますか？
6. 「台風」と「強い台風」を国際的な場面で表現する場合、どちらも「Typhoon」で良いのか？（「台風」と「Typhoon」は厳密には意味が違う、と聞いたのですが、どう違うのですか？）

台風に関する用語

1. 「藤原の効果」とは何ですか？
2. 「豆台風」という言葉を聞いたことがあります。どのような台風を指すんでしょうか？
3. 「台風銀座」の定義はありますか？
4. 雨台風、風台風って何ですか？

台風に関する統計・ランキング

1. 過去寿命がもっとも短かった台風は、どうすれば知ることができますか？
2. 過去寿命がもっとも長かった台風は、どうすれば知ることができますか？
3. 過去もっとも急速に発達した台風を知りたいのですが、どのように調べればいいのでしょうか？
4. 台風特異日ランキング　台風が存在していることの多い日は？　農家の三大厄日の一つと言われる二百十日に台風が発生していることは多いの？

気象キャスターネットワークとのコラボ

- 研究データ管理と永続的な保存に向けた要望の高まり。
- デジタル台風に組み込む形での研究成果の公開。
- デジタル台風の本当の役割とは何？

まとめ

1. デジタル台風の最初の目的は、**画像検索評価のためのデータセット作成**だった。
2. 現在の方向性は**ビッグデータ解析**。大規模異種データ統合等の問題に取り組む。
3. 気象データ、テキストデータ、災害・防災データ等の具体的な事例をまとめた。
4. 今後の課題として、クラウドソーシングやオープンイノベーションを挙げた。

関連サイト

- **デジタル台風**
  - http://agora.ex.nii.ac.jp/
- GeoNLP
  - https://geonlp.ex.nii.ac.jp/
- **気象庁防災情報**XML
  - http://agora.ex.nii.ac.jp/cps/weather/report/
- Researchmap
  - http://researchmap.jp/kitamoto/