

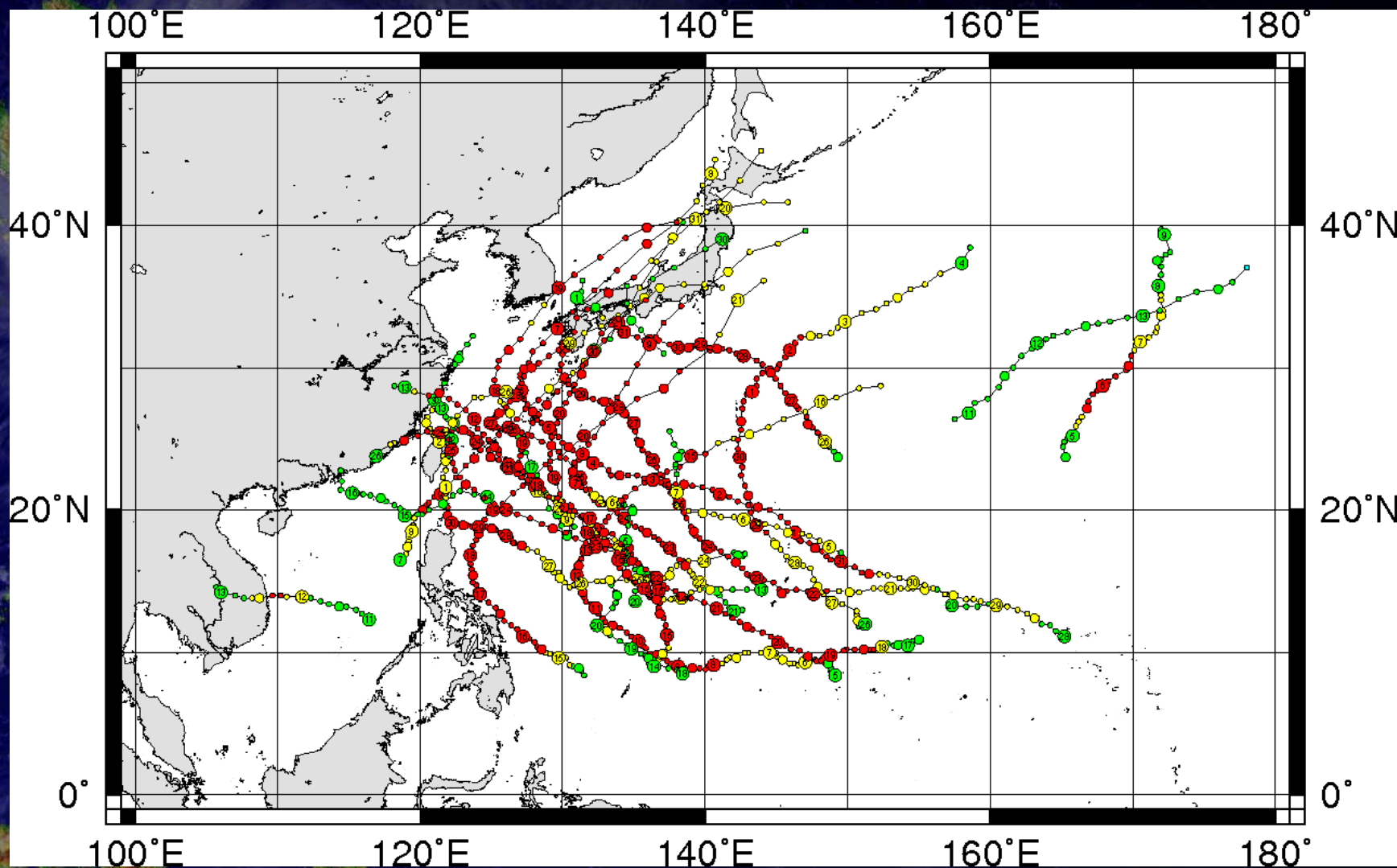
「デジタル台風」プロジェクト：
大規模時系列画像データベースの
マイニングにむけて

国立情報学研究所

北本 朝展

<http://www.digital-typhoon.org/>

2004年の台風経路図(速報)



災害が続発した2004年

- 相次ぐ台風の来襲

- 日本上陸個数：10個（従来記録6個）

- 沖縄接近回数：14個（従来記録12個）

- 台風23号：死者不明89人（兵庫県23人）

- 規模が大きいものも多く、しかも勢力が弱まる前に日本列島に到達した。

- 相次ぐ台風による大量の降雨が、新潟中越地震の土砂災害の原因にもなった。

阪神・淡路大震災から約10年

- 災害時における情報システムの役割とは何か？
- インターネットは飛躍的に普及したが、それによって状況は改善したのか？
- 公共機関が発信する1次情報の充実に比べ、個人が発信するローカルな1次情報の集約・データベース化は？

新潟中越地震における 緊急時情報システム

	1次情報	2次情報
公共的	公共機関からの情報発信は大幅に増加し有用な情報源に成長。	マスコミ報道は強力な情報源。バイアスもあるが改善も。
個人的	安否情報についてはNTT等が活躍した。しかしその他は掲示板等による情報交換が主流。期待のウェブログは十分に役割を果たしているか。	リンク集やコピー・ペーストサイトが主流のよう。10年前からそれほど進歩していないような印象を受ける。

緊急情報のための データベースシステム

- 緊急的な情報の流れを集約し、データベースを即時的に更新。
- 現在の状況と過去の事例とを、即座に比較できるようにする。
- この問題の解決には、マルチメディアデータベース技術が不可欠である。

ウェブサイト「デジタル台風」

	1次情報	2次情報
公共的	気象庁からの台風情報と気象衛星画像の処理とデータベース化。	ニュースサイト記事の自動言語処理と組織化。
個人的	多数の個人が発信するウェブログからのトラックバックを集約したデータベース。将来的には携帯機器対応。	台風ニュース・ウェブログによる台風関連情報の概要。

<http://www.digital-typhoon.org/>

デジタル台風: 台風画像と台風情報

最新台風情報

更新日時: 2004年 10月 26日 19時 20分 (JST)

台風発生数

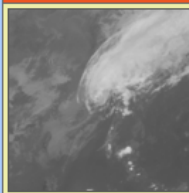
現在 = 24個 (台風経路図)

平年 = 22.5個 (1951-2003)

ニュース

- [台風ニュース・ウェブログ](#)
 - 2004年台風24号(ノックテン | NOCK-TEN)
 - 2004年10月25日
- [台風ニュース・トピックス](#)
 - 地域ごとの台風ニュースと台風情報
- その他のニュース

台風200424号



[GOE904102603](#)

[200424 \(WNP\)](#)

[\(N29.0, E128.0\)](#)

1000 hPa / 45 kt

台風データベース

1. メタデータによる検索

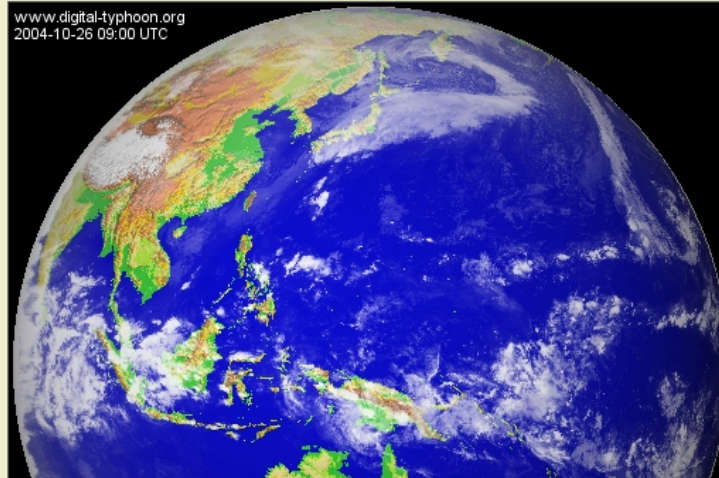
1. [日時・シーズンで検索](#)
2. [名前・番号で検索](#)
3. [地図で検索](#)
4. [地名\(緯度・経度\)で検索](#)
5. [最低気圧で検索](#)
6. [最大風速で検索](#)
7. [日付で検索](#)
8. 台風カレンダー
 - : [北西太平洋](#) :: [南西太平洋](#) ::

2. 画像内容検索

1. ランダム検索
 - : [北西太平洋](#) :: [南西太平洋](#) ::
2. 自己組織化マップ
 - : [北西太平洋](#) :: [南西太平洋](#) ::

静止気象衛星画像

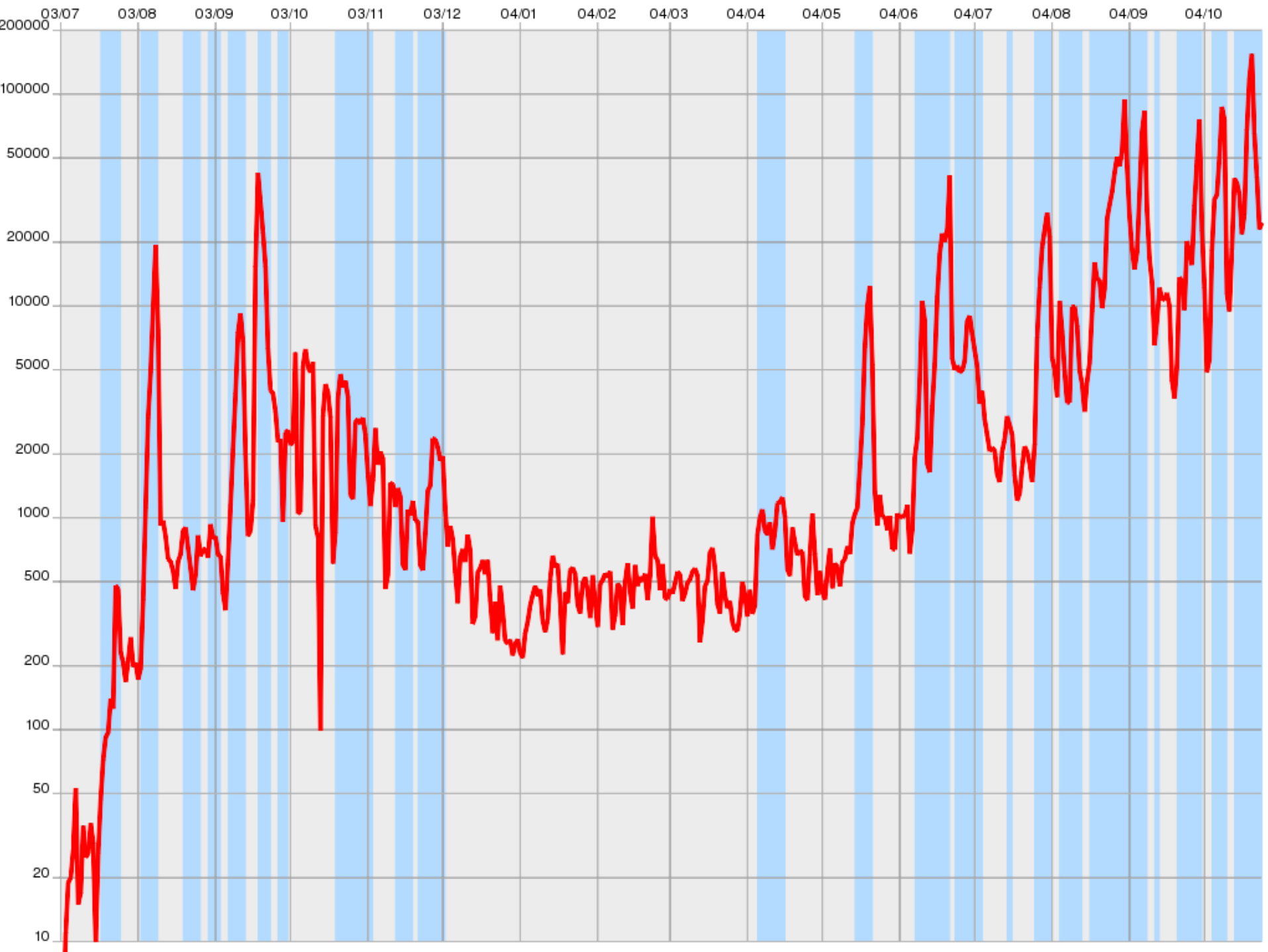
2004年 10月 26日 18時 (JST)



www.digital-typhoon.org
2004-10-26 09:00 UTC

「デジタル台風」の概要

- ・ 台風情報に関するポータルサイト。
- ・ 台風画像と台風ニュースをほぼリアルタイムで更新しデータベース化。
- ・ 大部分の情報を日本語と英語で提供。
- ・ 1日のトップページビュー:
 - Weekday 5,000-20,000
 - Peak 50,000-150,000
 - サイトページビュー 最高420,000/日



なぜ台風なのか？

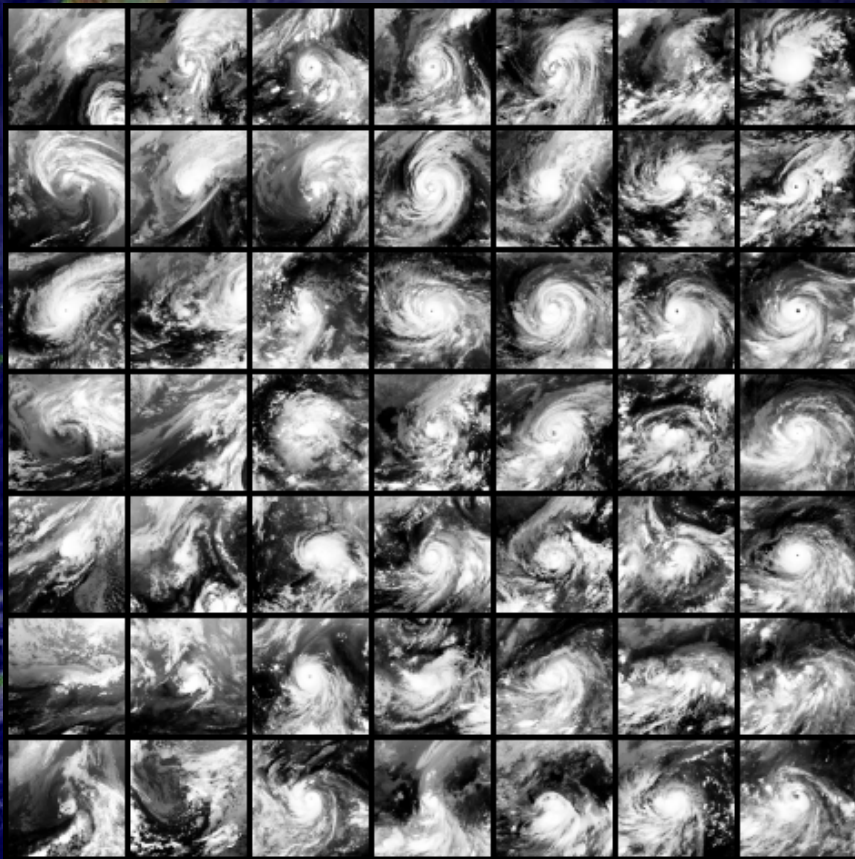
1. 地理的・時間的に複雑なふるまいを見せる実世界の重要現象。メテオインフォマティクスのモデル現象となる。
2. 事例に基づく台風解析と予測に利用。
3. 情報学の研究者としてできることに取り組むが、その結果が気象学的に意味があるかどうかを常に念頭におき、Cross Communityな研究を目指す。

台風画像コレクション

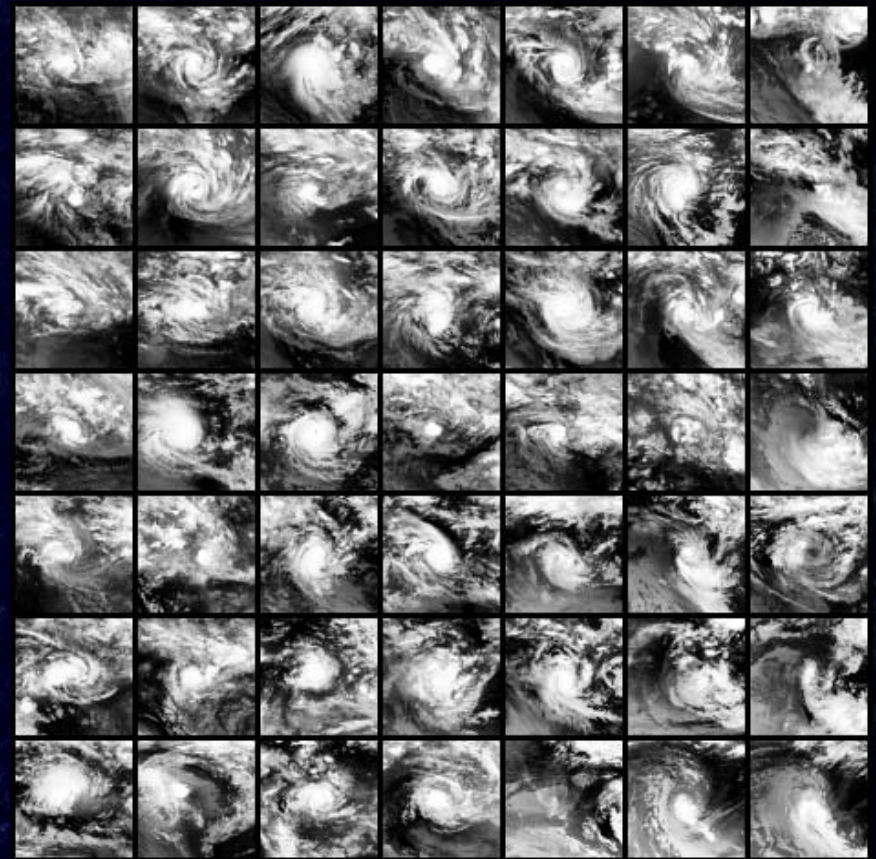
2004年10月現在	北西太平洋	オーストラリア
ベストトラックデータ		
担当機関	気象庁 (JMA)	豪・気象局 (BOM)
緯度範囲	赤道以北	赤道以南
経度範囲	100~180	90~170
台風画像コレクション (累計 53,150)		
シーズン数	9 シーズン	6 シーズン
台風系列数	232	86
台風画像数	41,278	11,872
1系列あたり画像数	53~433	25~513

台風画像コレクション

画像データコレクションのK-平均クラスタリング



北半球



南半球

網羅的画像コレクションとは

- データの重要性に関する予見を排除し、できるだけ大規模にデータベース化。
- 重要なデータは、ユーザがデータベース上で選択する。
- 高品質が望ましいのは当然だが、均質性を保つ品質管理にも留意すべき。
- 地図投影法のような重要な前処理については、事前の慎重な検討が必要。

前処理の一つ: 地図投影法

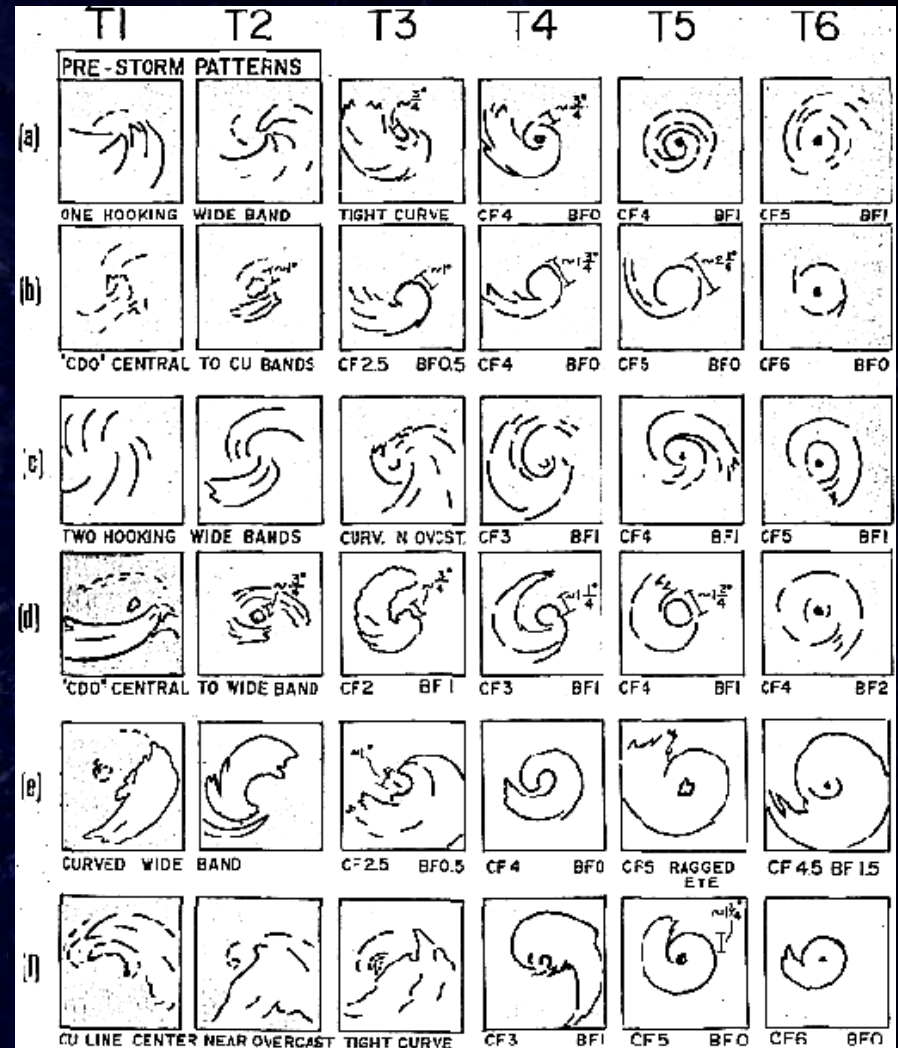
- 地図投影法が満たすべき条件とは？
 1. 地理的位置によらず面積を比較できる。
 2. 円形の物体に対する歪みが小さい。
- 以上の条件を満たす地図投影法として、「ランベルト正積天頂図法」を選択。
- 上記の性質により、複数の台風雲パターンの形態をダイレクトに比較できる。

研究課題

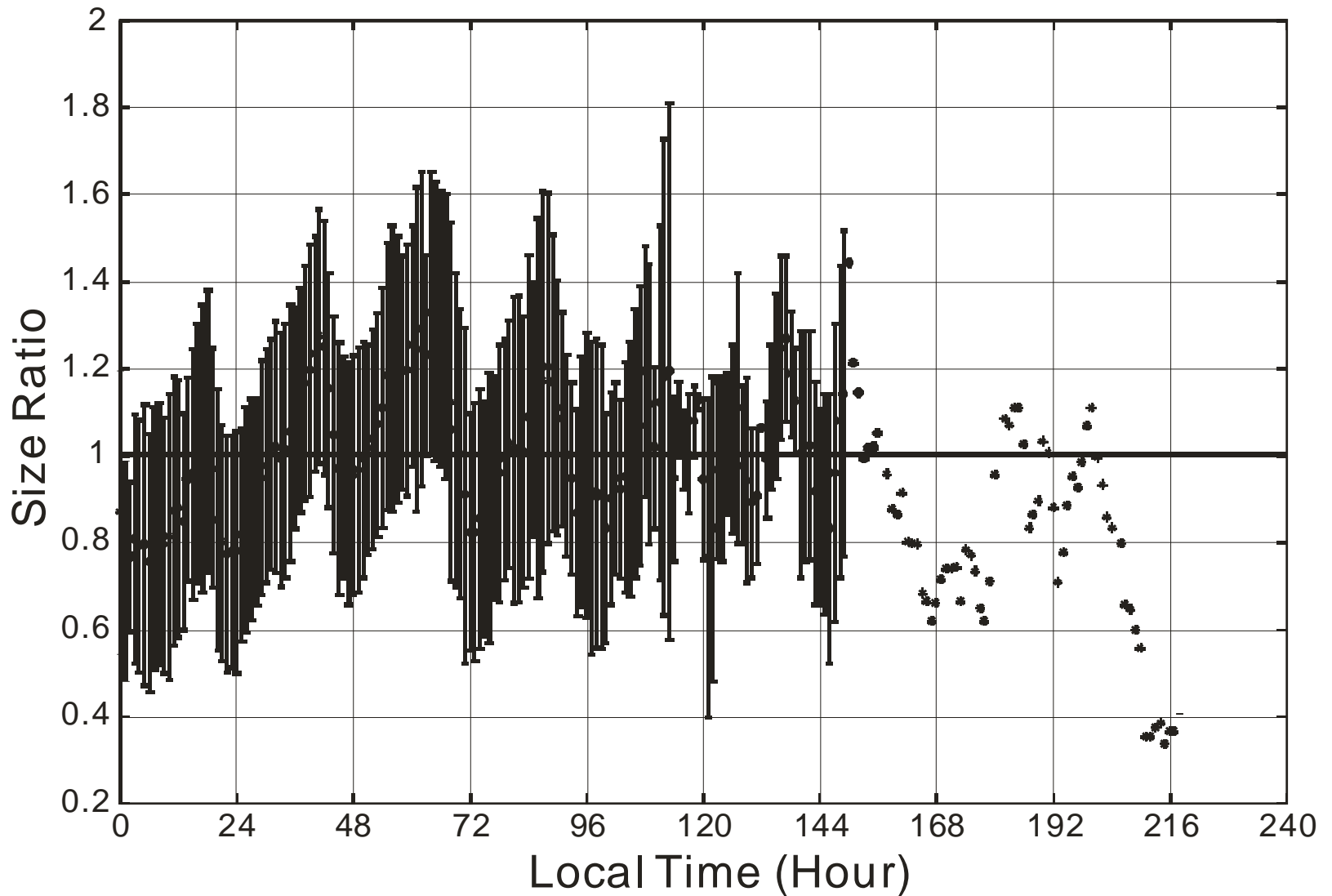
- 「台風の形態学」が重要な研究課題。
- 画像認識および画像データベース、マイニング技術を中心とする。
 1. 画像特徴から台風発達の予兆をどのように発見するか？
 2. 台風の時空間的发展パターンをどのように数理モデル化するか？

台風の状態学

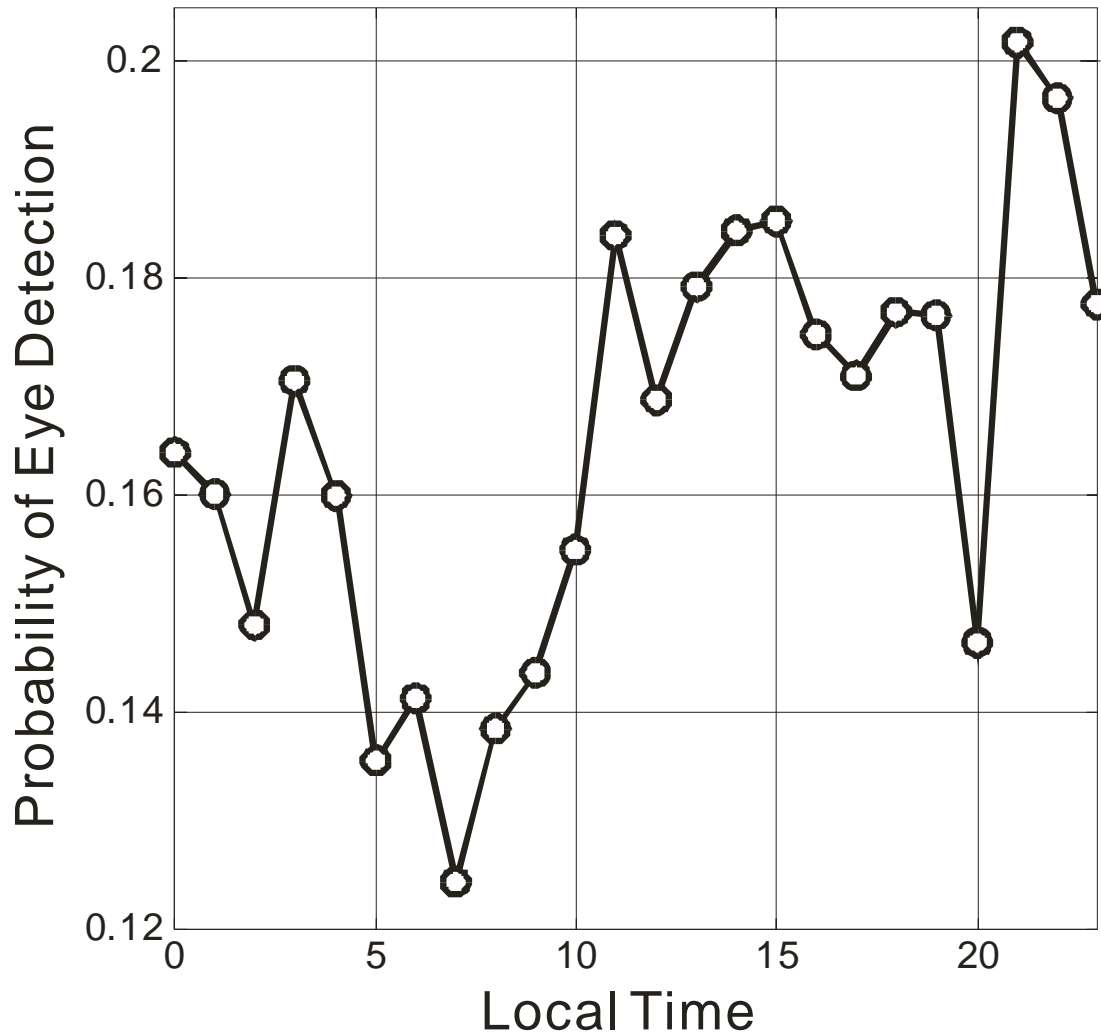
- 状態学的方法による台風解析(ドボラック法)。
- 専門家による気象衛星画像雲パターンの目視解析を基本に、他の観測データを加えて総合判断する。



台風サイズの日周変化



台風眼検出確率の日周変化



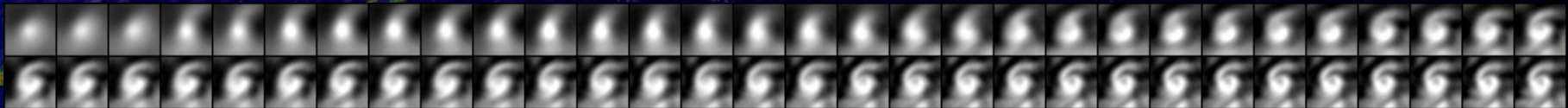
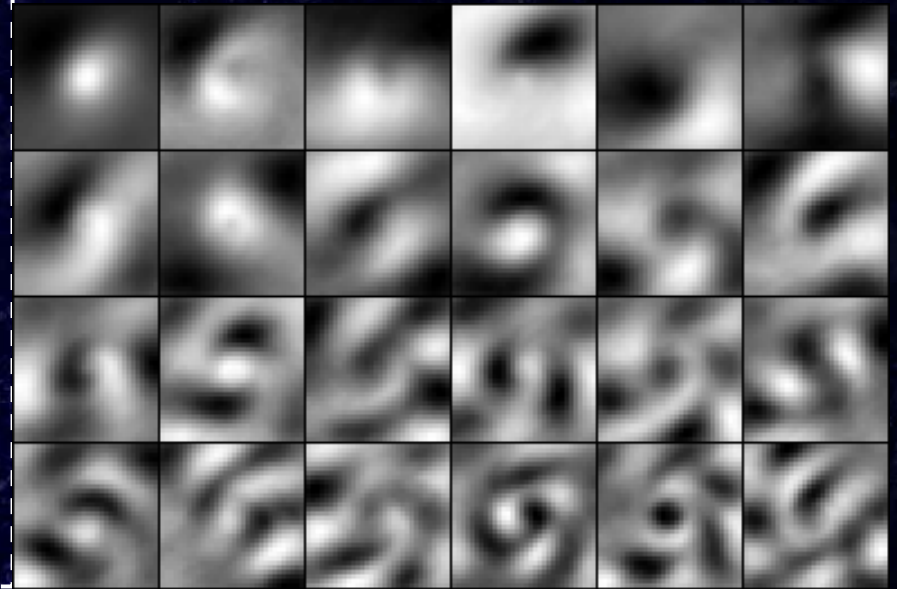
画像特徴空間

- 画像そのものをデータとするのではなく、**重要な画像特徴を抽出して空間を構成し、その空間で検索などをおこなう。**
- **画像特徴空間の次元を小さくし、かつよい距離尺度を定義できれば、そこから有用な情報を発見することができる。**
- この部分は画像解析の**核心的な問題**。

固有表現に基づく画像特徴空間

- 枠取りした台風画像の地図投影
- 雲画像分類から雲量ベクトルの計算
- 主成分分析による次元削減
- 特徴ベクトル算出

北半球の台風に対する固有台風表現



類似性検索

- 過去の事例を類似性に基づき検索。
- 画像特徴空間の選択により、適切な事例が検索できるかが決まる。

デジタル台風:類似画像ビュー - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://agora.ex.nii.ac.jp/cgi-bin/dt/search_image2.pl?prefix=GOE904101902

北本 朝展 @ 国立情報学研究所 (NII) 検索 サイトマップ

デジタル台風:類似画像ビュー

agora ホーム デジタル台風 ヘルプ English

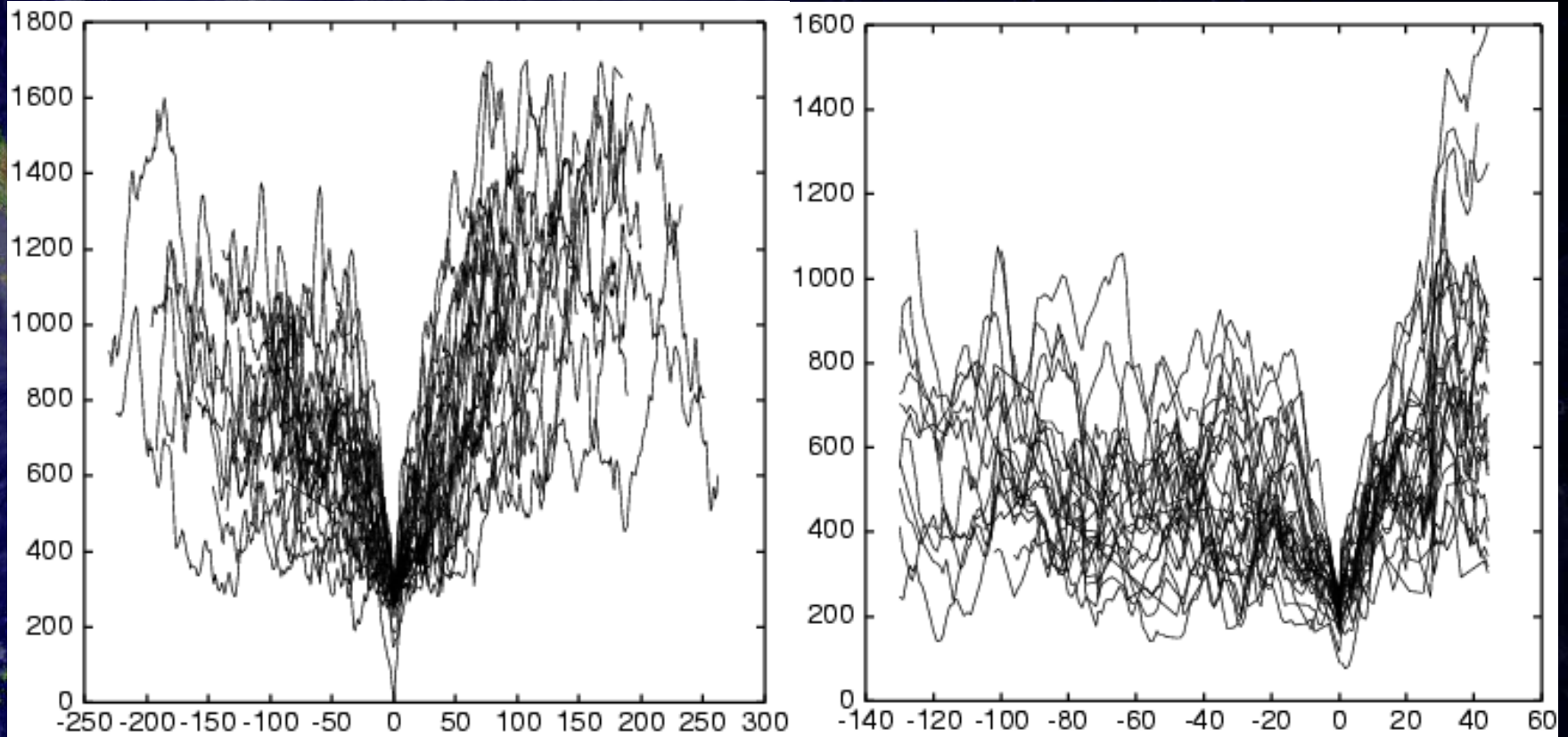
画像類似度による検索

Query 1	1	2	3	4
GOE904101902	GOE904083011	GMS502083023	GOE903113013	GOE903091201
200423 (WNP)	200416 (WNP)	200215 (WNP)	200321 (WNP)	200314 (WNP)
(N25.1, E127.4)	(N34.8, E132.6)	(N32.4, E127.2)	(N25.0, E136.0)	(N30.8, E126.6)
950 hPa / 80 kt	965 hPa / 70 kt	955 hPa / 70 kt	945 hPa / 80 kt	930 hPa / 95 kt

5	6	7	8	9
GMS507000402	GMS508002208	GMS507001000	GMS502120008	GMS506002023

Done

カオス的性質の現れ



- 適切な画像特徴空間をとってもなお、データ間の距離は初期値から拡大する。

Technical Result (1)

- Problem Statement:
Classification of Developing / Weakening Typhoon
- Goals:
 - Recognize some “**signs**” of cloud patterns before it actually occurs?
 - Solve a relaxed problem → **classify moderately developing / weakening typhoons from cloud patterns.**
 - Threshold = $|10\text{hPa}| / 24\text{hour}$.

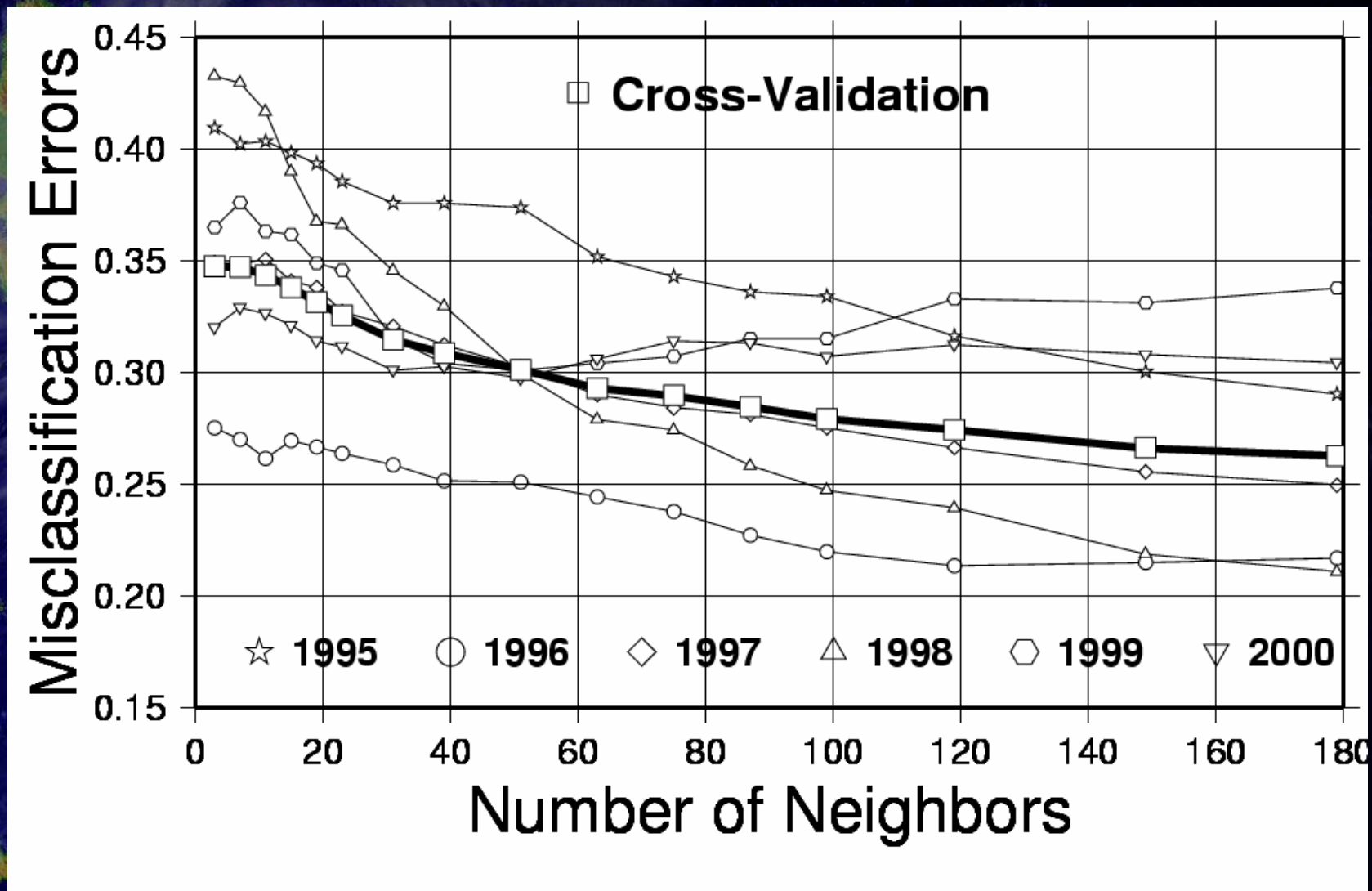
Rapid Deepening

- **Rapid deepening** – intensifying +30-40 hPa in 24 hours.
- A very significant event from the viewpoint of **disaster management**.
- Meteorologically **the prediction of rapid deepening** is difficult.

Our Approach

- **Kernel PCA** -> Extraction of non-linear principal components.
 - The effectiveness is still not clear.
-
- **Support Vector Classifier** -> This part is discussed in this presentation.
 - Many thanks to Libsvm (V2.35) from Prof. Lin@National Taiwan University.

Classification by K-NN



Classification by nu-SVM (1)

Kernels	Nu	6-fold CV
POLY d=1	0.3	33.9
	0.5	26.9
	0.7	25.4
POLY d=2	0.3	33.1
	0.5	32.6
	0.7	34.0
POLY d=4	0.3	34.6
	0.5	35.0
	0.7	37.6

Classification by nu-SVM (2)

Kernels	Nu	6-fold CV
RBF $g=0.1$	0.3	28.4
	0.5	26.3
	0.7	25.6
RBF $g=1$	0.3	26.7
	0.5	25.8
	0.7	25.5
RBF $g=10$	0.3	26.2
	0.5	24.8
	0.7	25.8

Comparison

- From the result of K-NN, **the baseline classification error is about 25%**.
- The polynomial kernel achieved this performance with only a limited combination of parameters.
- The RBF kernel is relatively stable and showed **comparable performance**.
- This error rate (25%) may be close to **Bayes error rate?**

Summary

- We can classify and capture the signs of typhoon development from image features.
- Advanced image features such as
 - Shape, texture-based image features.
 - Dynamic (temporal) features.
- **Advanced kernel design** such as a HMM-based kernel.

Technical Result (2)

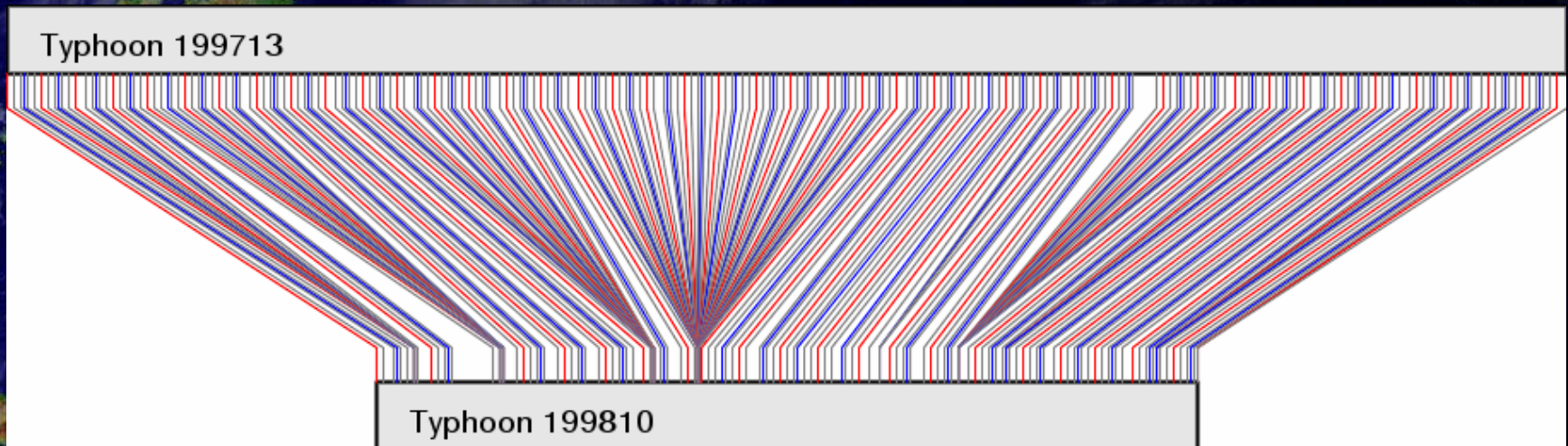
- Problem Statement: Multiple alignment and meaningful fragment extraction of typhoon image sequences.
- Goal:
 - Extract meaningful temporal fragments based on pattern similarity.
 - Natural boundaries between fragments are usually unknown.

Our Approach

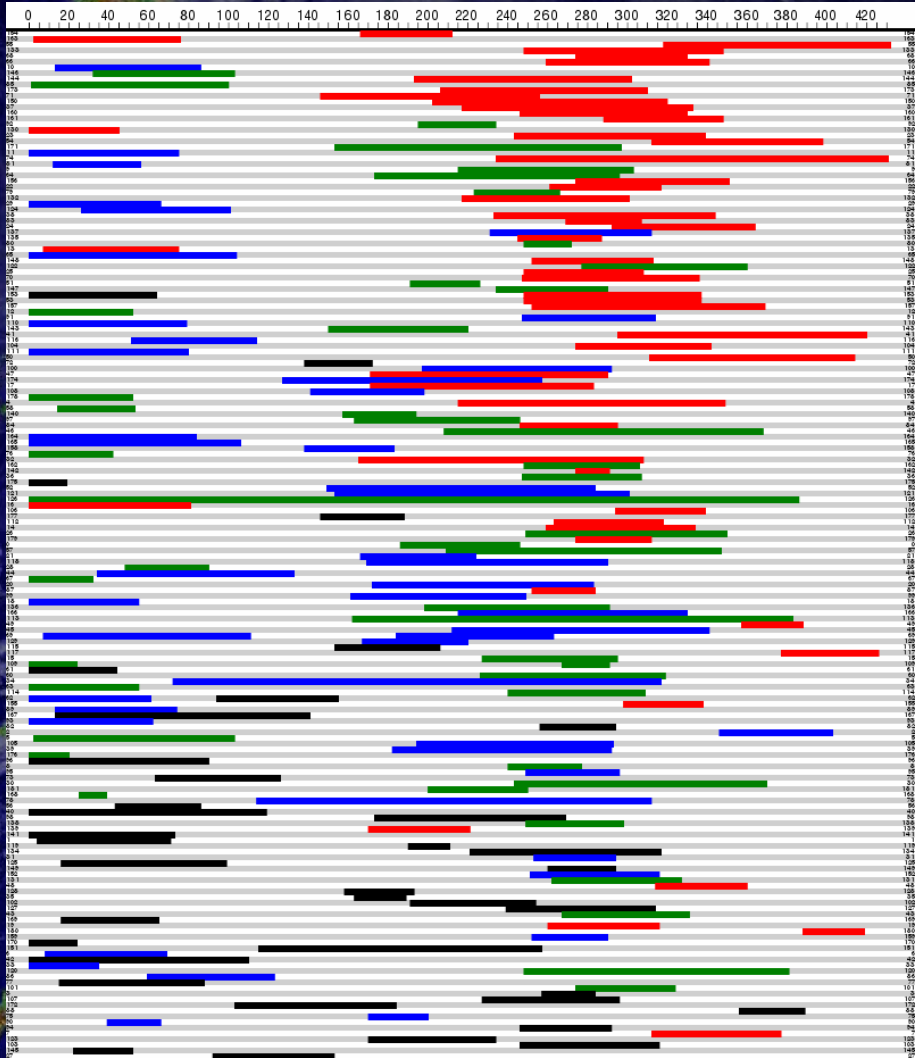
1. Apply **global alignment** between the pivot sequence and another sequence.
2. Apply **multiple sequence alignment** to find similar temporal patterns.
3. **Dynamic programming** and **clustering** are key components.

Pair-wise Alignment

1. **Global alignment** of two typhoon sequences and extract fragments.
2. **Dynamic programming** is a typical solution for computing the **minimum matching score**.

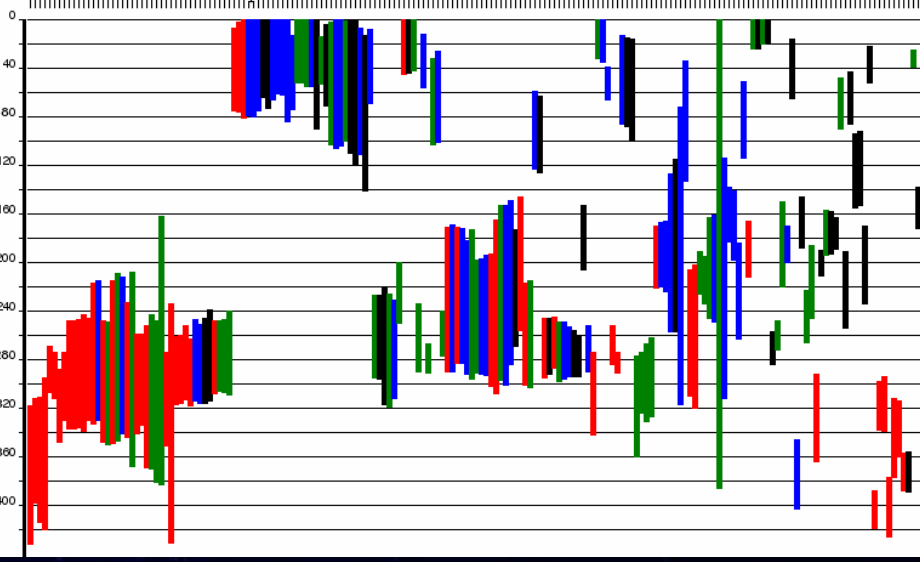
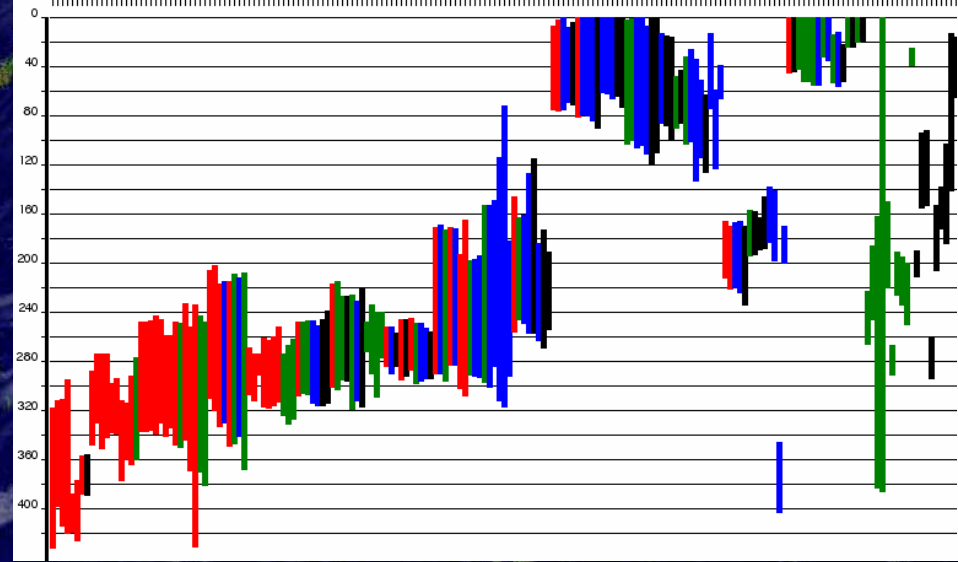
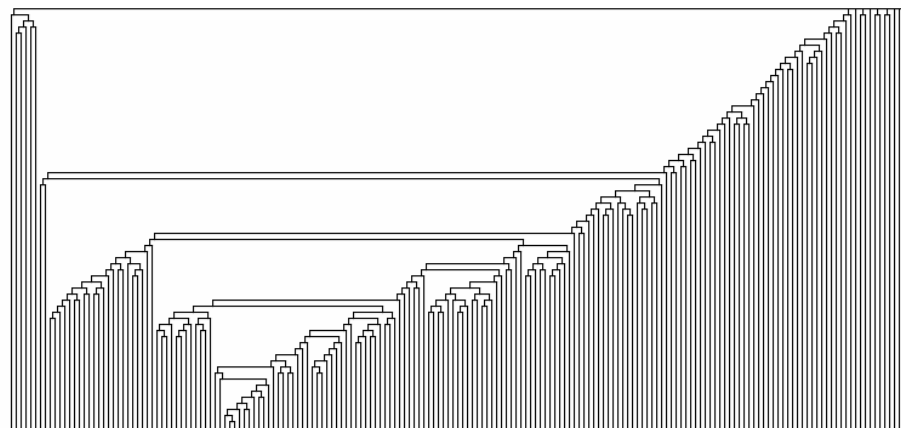
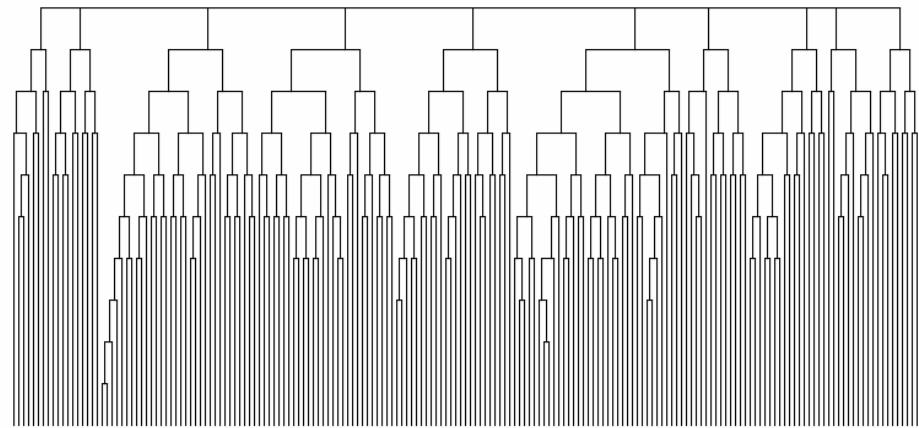


Alignment with the pivot sequence



- Alignment of sequences with the pivot sequence.
- Parallel matching fragments are extracted.
- Similar sequences are shown upper.

Hierarchical Clustering



Ward method

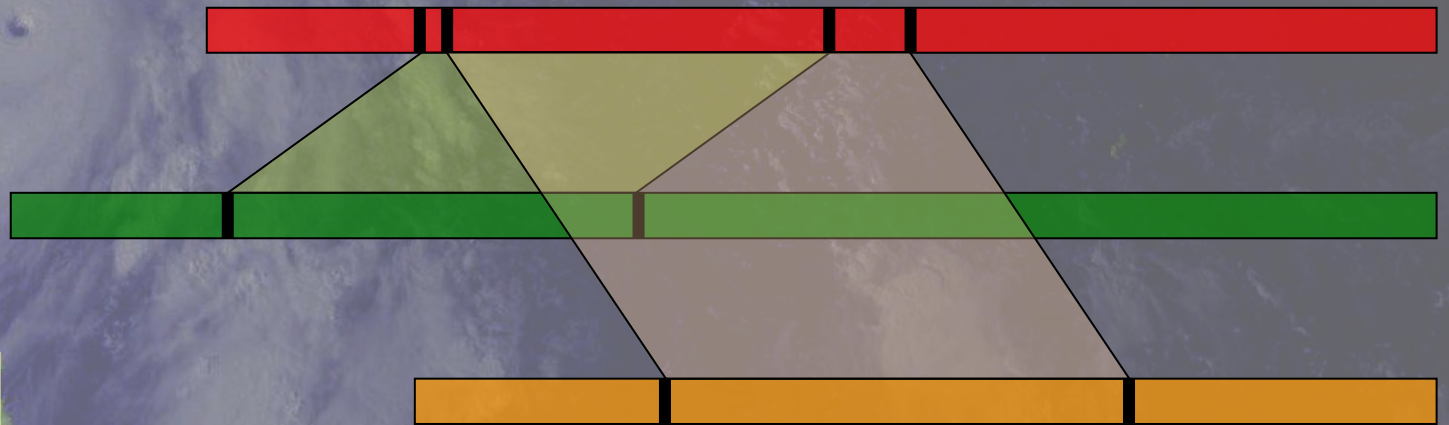
Min-Dist Method

Distance Measure between Fragments

Pivot i

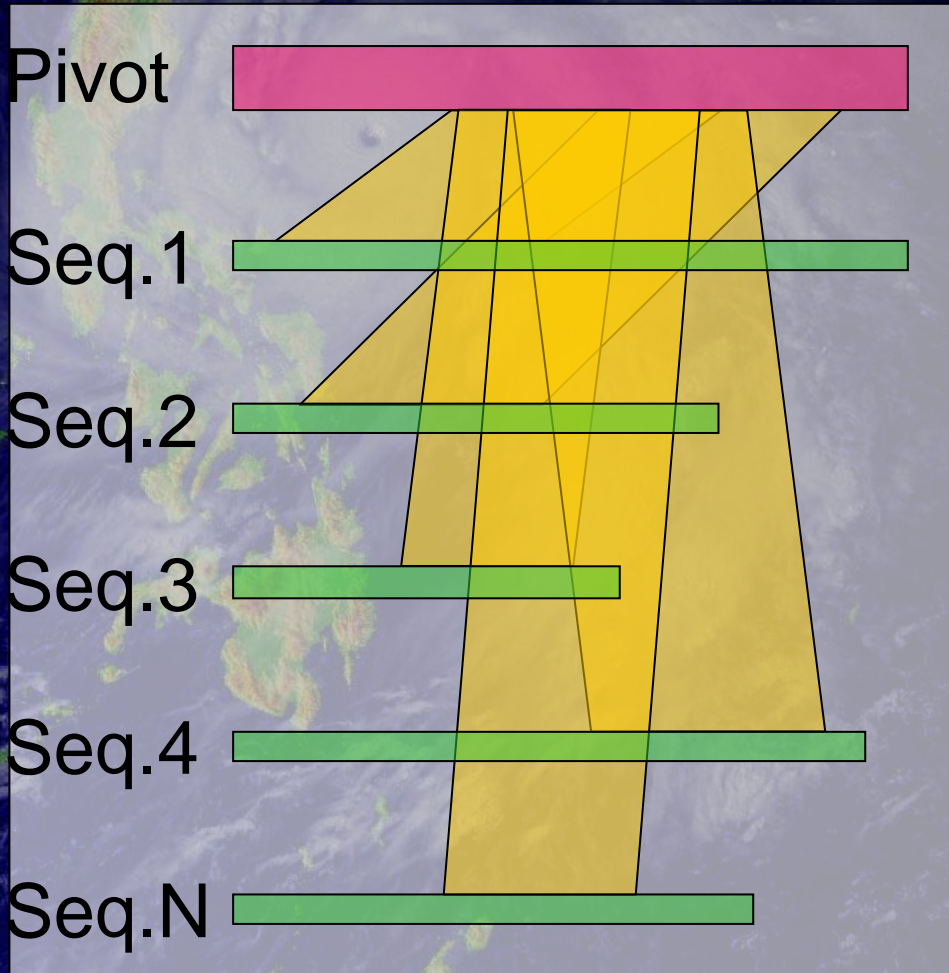
Seq. j

Seq. k



1. Large overlap on the pivot sequence?
2. Ratio of matching scores is small?

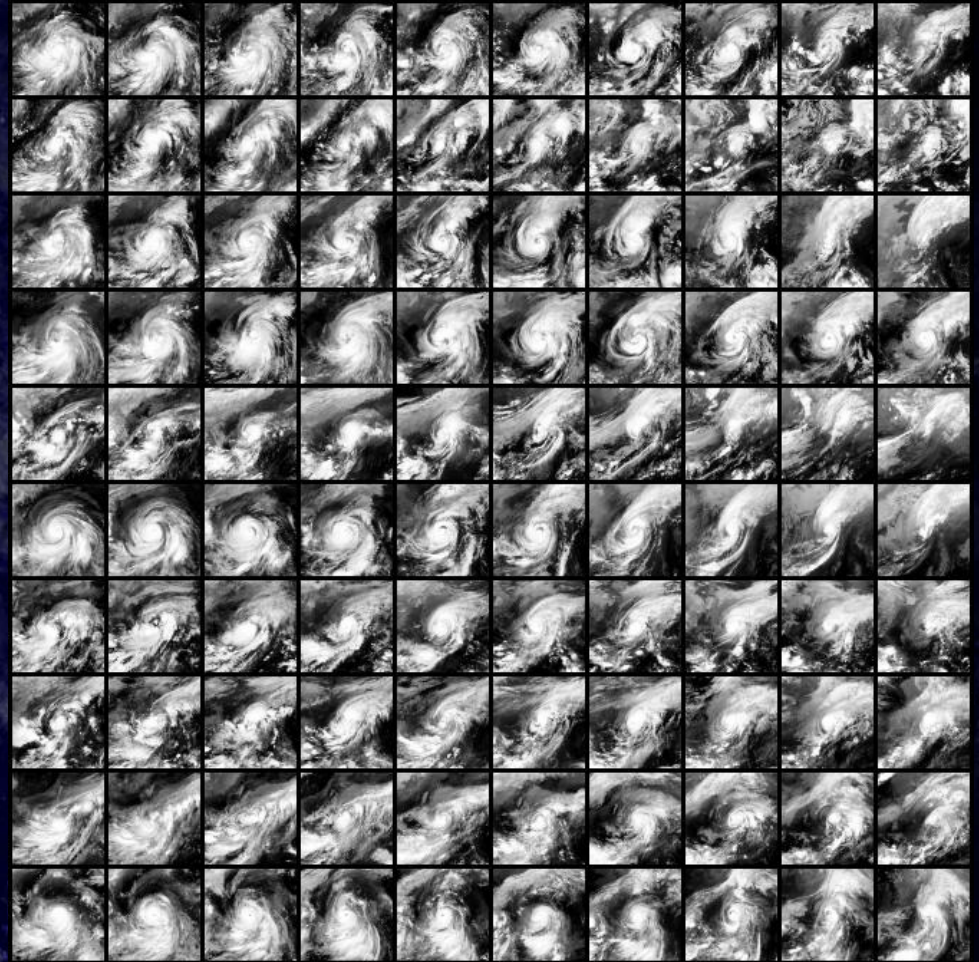
Consensus Fragments



- **Global alignment** results are combined.
- Average position is computed as **consensus fragments**.

Matching Results

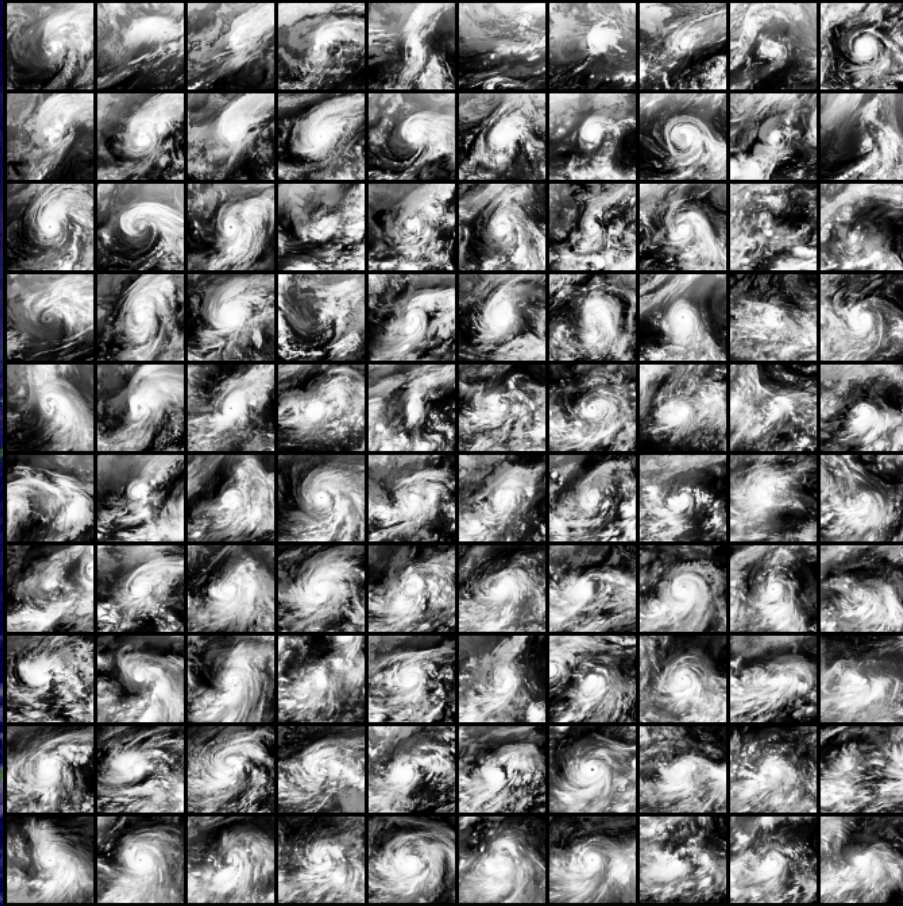
- **Fragment sets** that illustrate the diversity of **typhoon life cycle**.
- **Comparison of cloud patterns** is intuitive after alignment.



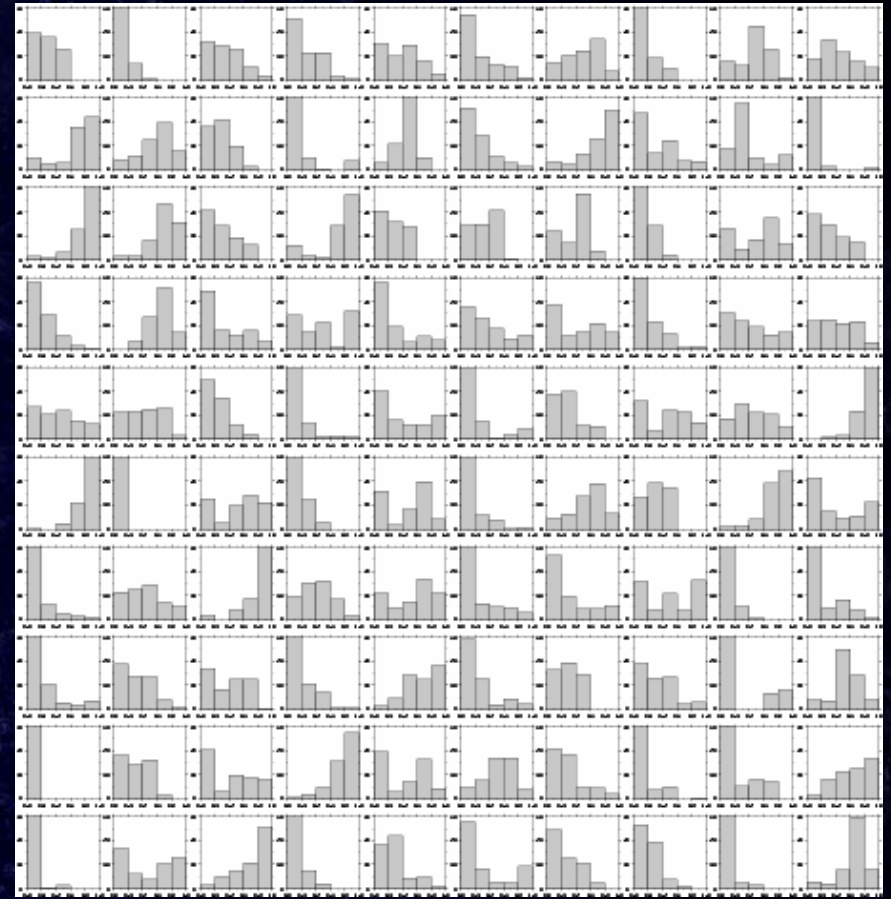
Summary

- Similar temporal fragments can be extracted using **dynamic programming** and **clustering techniques**.
- Advanced **distance measures** and **alignment schemes**.
- Advanced **clustering methods** for “**semantically**” similar fragments.

クラスタリングと時間分布



K-平均クラスタリング



時間分布

データベースに基づく科学研究

- 生物情報学

- ヒューマンゲノムプロジェクト

- 天文学

- スローンデジタルスカイサーベイ

- X-Informatics と呼ばれる分野を形成。

- 計算的アプローチと情報学的アプローチ

メテオ・インフォマティクス

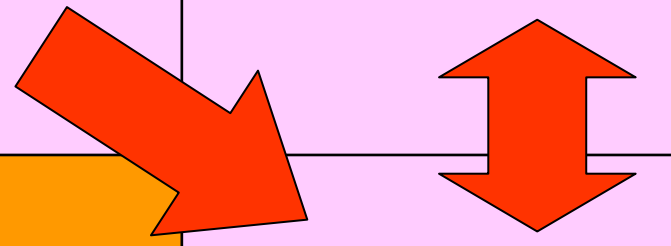
- **メテオ・インフォマティクス(気象情報学)**があってもよいのでは？
- データの検索や比較を繰り返すことにより、**知見を発見し蓄積していく。**
- **どんなデータベースが必要か、データからどんな知識が発見できるか？**
- 数値シミュレーション等の**演繹的アプローチ**に対し**帰納的アプローチ**を追究。

計算的アプローチと 情報学的アプローチ

- 気象学は例外的に(?)計算的アプローチ(数値シミュレーション)が非常に発達し、しかも成功した分野。
- ENIAC以来、計算的アプローチが他を圧倒し発展してきた歴史がある。
- しかし、それだけでいいのか？

メテオインフォマティクスの領域

	データ	モデル
説明	事例解析 統計解析	数値天気予測
一般化	メテオ・イン フォマティクス	大気(地球)シ ミュレーション



「事例」概念の拡張

- シミュレーション結果は厳密には「事例」ではないが、データに基づく方法の根本的な問題である事例不足を軽減できる可能性がある。
- リアルデータとシミュレーションデータを比較できるような、巨大科学データベースの設計について考えるべき？

二つのアプローチの融合

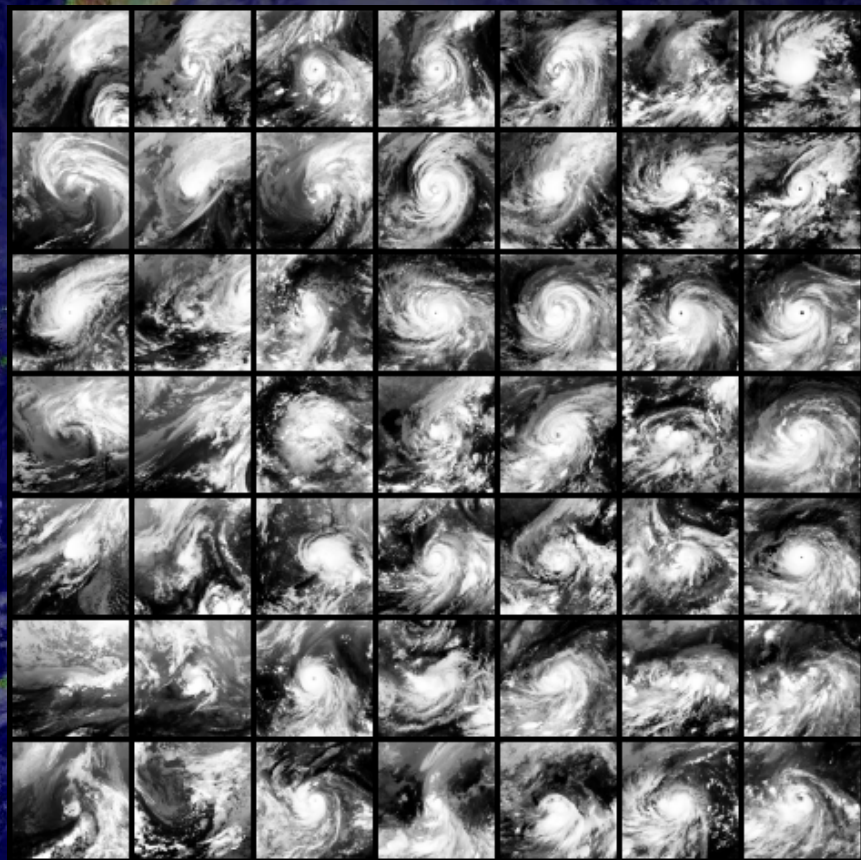
1. シミュレーション結果に類似した現象を過去のデータから検索する。
2. シミュレーション結果と過去のデータとは、どのように一致し、どのように異なるのかを分析する。
3. シミュレーションに使えそうな観測結果や法則を、論文データベースから検索する。

今後の研究課題

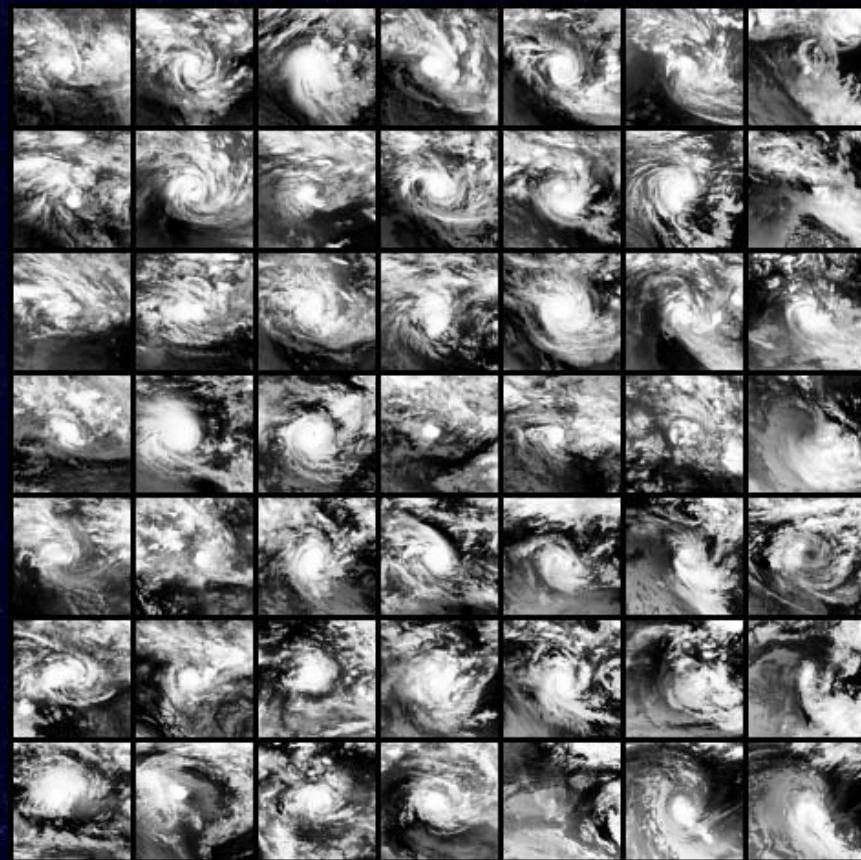
1. ダイナミックに変動する画像(空間)パターンの時系列モデリング
2. パターン発見に適した問合せ言語を備えた高速画像検索エンジンの開発
3. 大規模な分散データ収集システムの構築と集約した情報の即時的組織化

比較台風学

画像データコレクションのK-平均クラスタリング



北半球



南半球

全世界台風情報ネットワーク

- 全世界に発生する**熱帯低気圧の網羅的なデータベース**を構築し、科学的な研究に役立てる。
- 全世界で台風を見つめる人々の**出来事などを即時的に集約し組織化し配信するようなネットワーク**を形成し、災害情報などに役立てる。

おわりに

- 「デジタル台風」プロジェクトの概要とウェブサイトを紹介した。
- このプロジェクトは、メテオ・インフォマティクスモデルプロジェクトに位置づけることができる。
- 事例に基づく方法を現在は追究しているが、将来的にはシミュレーションに基づく方法と融合できるとよい。

謝辞

- GMS画像を研究目的にご提供くださる、東京大学生産技術研究所の安岡善文教授、喜連川優教授に感謝いたします。なおGOES画像は(財)気象業務支援センターによるものです。
- 台風中心位置等のデータは気象庁の観測に基づくものです。
- 本発表に関連する情報や画像などに関しましては、以下のウェブサイトをご覧ください。

– <http://www.digital-typhoon.org/>