

A satellite image of a tropical cyclone, likely a typhoon, with a well-defined eye and spiral cloud bands, positioned over the Philippines. The landmasses are shown in green and brown, contrasting with the dark blue of the ocean and the white/grey of the clouds.

メテオインフォマティクス (meteoinformatics) への構想とその展開

国立情報学研究所

北本 朝展

<http://www.kitamoto.org/>

発表の概要

1. 「メテオインフォマティクス」とは
2. 「デジタル台風プロジェクト」の紹介
3. メテオインフォマティクスに関連する研究の紹介
4. メソスケール事象との関連
5. 今後の展開と課題

気象学での情報技術の活用

- 歴史的に見て、気象学者はコンピュータの最もヘビーなユーザであり続けた。
 - ーリチャードソンの数値天気予測構想
 - ー世界最初期のコンピュータENIACを用いた大気予測の先駆的実験
 - ー地球シミュレータによる高精度・高分解能な大気シミュレーション
- 情報技術は既に十分に活用されている？

2つのアプローチ

メテオインフォマ
ティクス

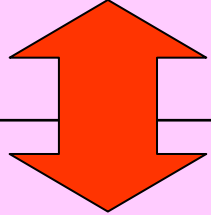
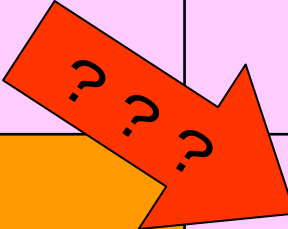
データの取得、
蓄積、体系化、
データベース化、
解析及び可視化
を含めた研究

コンピューター
シヨナル気象学

理論的手法, 数
学的モデリング,
およびコンピュー
タシミュレーショ
ンの研究

メテオインフォマティクスの狙い

	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



3つのアプローチ

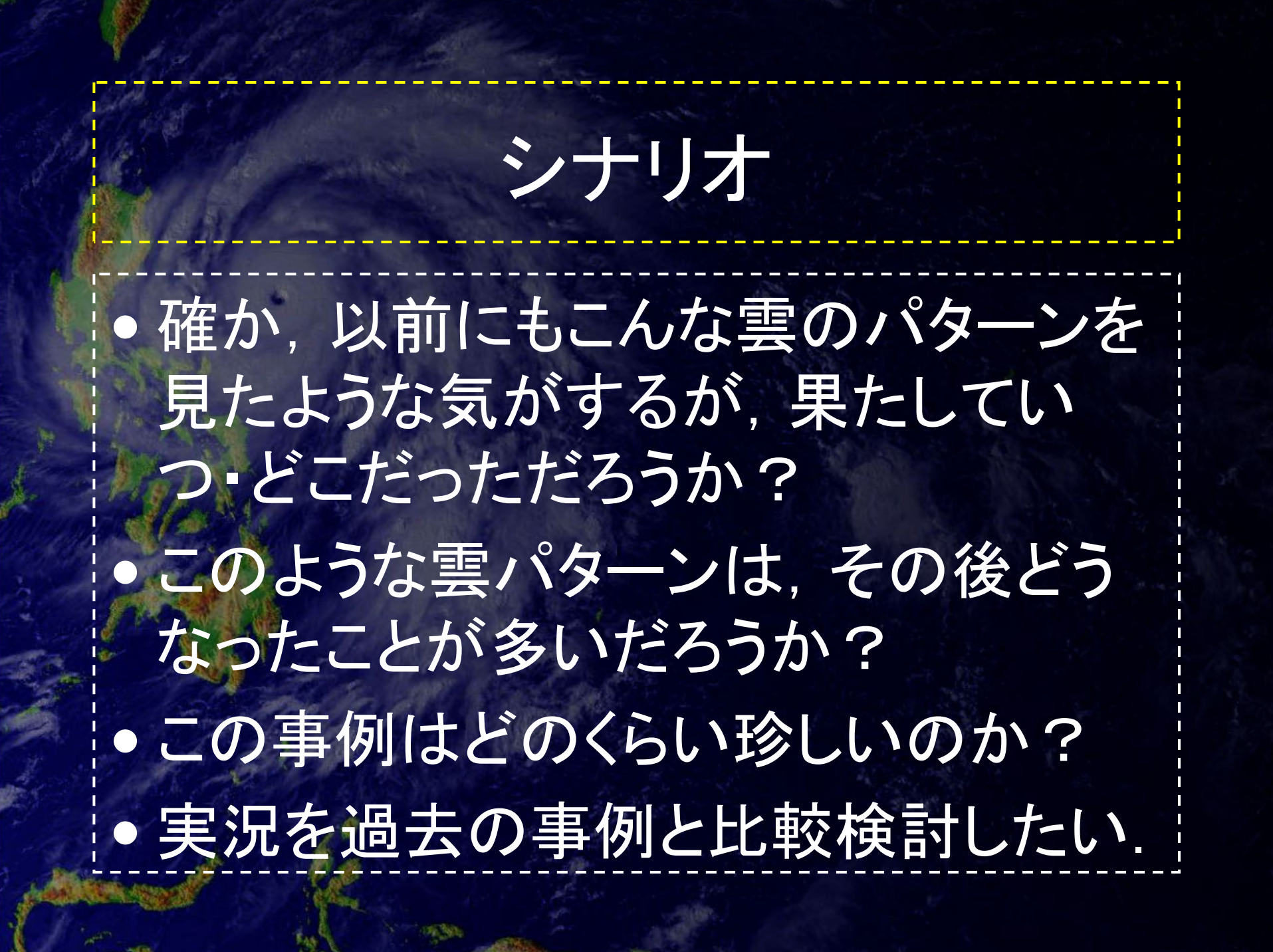
認知モデル	生成されたデータを人間はどう見るか？	メテオイン フォーマティクス
データモデル	生成されたデータはどのような性質をもつか？	
物理モデル	データを生成する物理過程はどんなものか？	大気力学理論

メテオインフォマティクスの特徴

1. データからの特徴抽出・要約・抽象化
2. データの比較
3. データの関連性の発見
4. データの分類・まとまりの発見
5. データの可視化
6. データの規則性や外れ値の特徴づけ
7. データから情報・推論・意思決定へ

メテオインフォマティクスの目標

1. 事例の集合(データベース)からモデル・法則を(自動・対話的に)導くこと.
2. 既存の気象学的モデルが有効でない場合に, 意思決定に有効な情報を提示し, 意思決定を支援すること.
3. 気象データをもっとわかりやすく, 使いやすく, 親しみやすくすること.



シナリオ

- 確か，以前にもこんな雲のパターンを見たような気がするが，果たしていつ・どこだっただろうか？
- このような雲パターンは，その後どうなったことが多いだろうか？
- この事例はどのくらい珍しいのか？
- 実況を過去の事例と比較検討したい。

メテオインフォマティクス風の方法

- 台風解析

1. 台風形状の分類・特徴づけ(ドボラック法)
2. 観測データからの中心勢力推定

- テレコネクション

1. 時空間的に離れた観測データに潜む, 意外な関連性の発見
2. 全球におよぶ大量のデータ処理が必須.

- 類推法(analogues)

台風解析

- 通常は実測できない，中心気圧や中心気圧などを，衛星画像や他のデータに基づいて推定する(ドボラック法).
- 画像上の雲のパターンが重要な役割を占める点が特徴的である.
- データからの意思決定問題であり，適用する人によってその結果は異なる.

テレコネクション

- ある場所の影響が遠くの天候に影響を与える現象をさす。エルニーニョや南方振動などがよく知られている。
- 時空間的に遠く離れた地点の共起関係を発見しなければならない。
- このような、データに潜む共起関係（アソシエーション）の発見は、情報学におけるホットな話題である。

類推法(analogues)

- 過去の類似事例を用いて将来を予測.
- 大気の非線形性に起因するカオス的性質によって、この手法の有効性は悲観的に考えられている.
- ただし、カオス的性質が支配的とならない範囲においては検討の余地がある.
 1. メソスケールの現象
 2. 短時間の現象

メテオインフォマティクスの根拠

1. 「実際に発生した事実」という重み.
2. 事例の効果的な収集により, 信頼性を向上させることが可能.
3. 十分な事例数はとてつもなく大きい
ため, 断定的な結論は出しにくい.
4. 客観的・主観的判断を系統的に取り
入れていく必要がある.

発表の概要

1. 「メテオインフォマティクス」とは
2. 「デジタル台風プロジェクト」の紹介
3. メテオインフォマティクスに関連する研究の紹介
4. メソスケール事象との関連
5. 今後の展開と課題

これまでの研究内容

- 「デジタル台風プロジェクト」

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

- 南北太平洋の気象衛星台風画像40,000件以上を網羅的に収集.
- さまざまなデータ検索・組織化などの機能を通して、台風画像コレクションという事例集合から何を発見できるのかを探る.

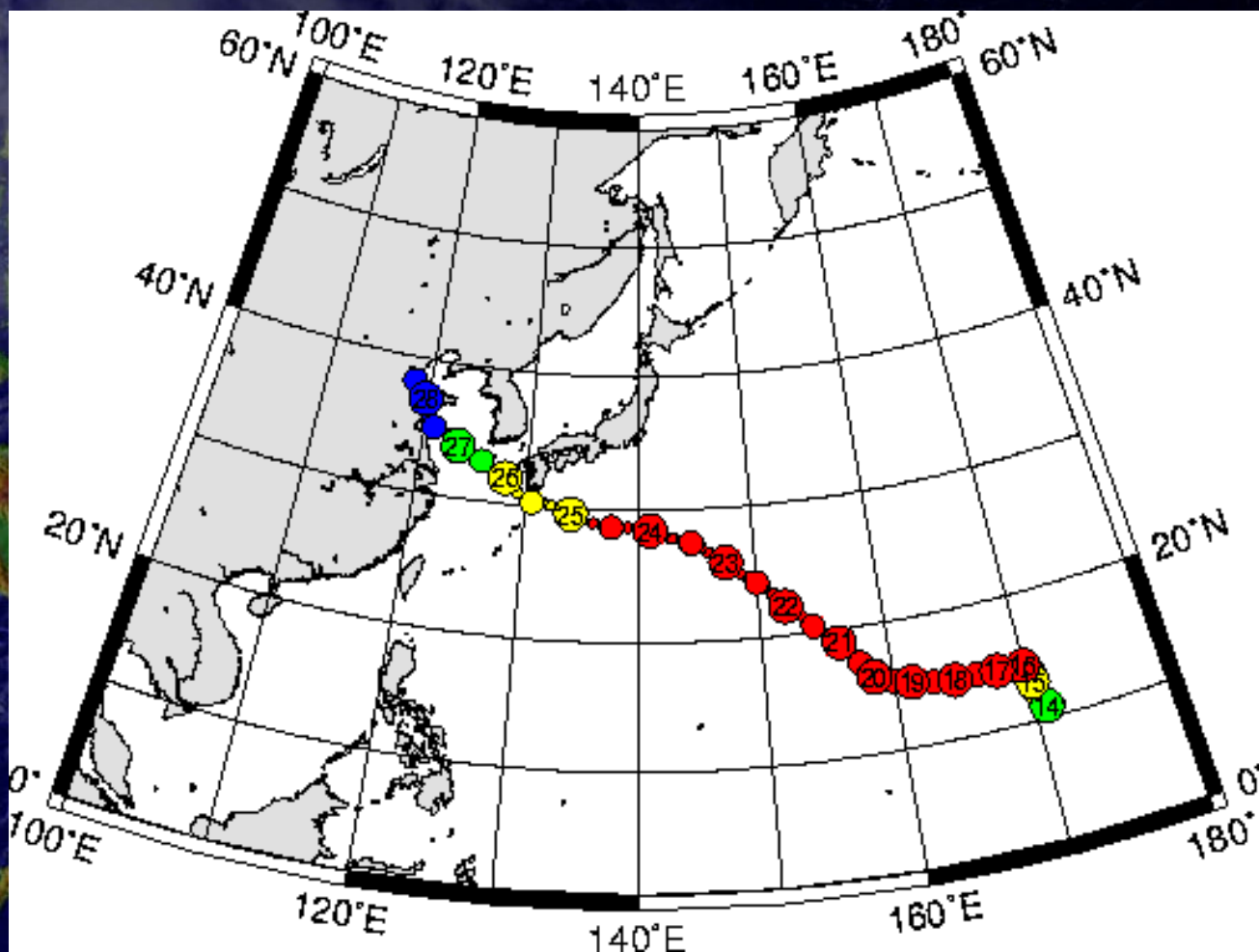
なぜ台風を対象としたか

- メテオインフォマティクス向きの研究対象.
 1. 良質のデータが公開されている.
 2. 挙動が未知の部分がまだ多い.
 3. 時系列データを追跡できる.
 4. 雲パターンが特徴的なパターンであり、適度の多様性を示す.
- 社会的に重要なことも理由の一つ. なぜなら、多くの人々が情報を欲しているから.

台風画像コレクション

<i>As of Jun. 2003</i>	<i>W.N. Pacific</i>	<i>W.S. Pacific</i>
<i>Best Track Dataset</i>		
<i>Organization</i>	<i>Japan (JMA)</i>	<i>Australia (BOM)</i>
<i>Lat. range</i>	<i>N. of equator</i>	<i>S of equator</i>
<i>Lon. range (E)</i>	<i>100~180</i>	<i>90~170</i>
<i>Typhoon Image Collection (Total 44,832)</i>		
<i>Num of seasons</i>	<i>9 seasons</i>	<i>6 seasons</i>
<i>Num of seq.</i>	<i>193</i>	<i>70</i>
<i>Num of images</i>	<i>34,125</i>	<i>10,758</i>
<i>Images per seq.</i>	<i>53~433</i>	<i>25~513</i>

ベストトラック

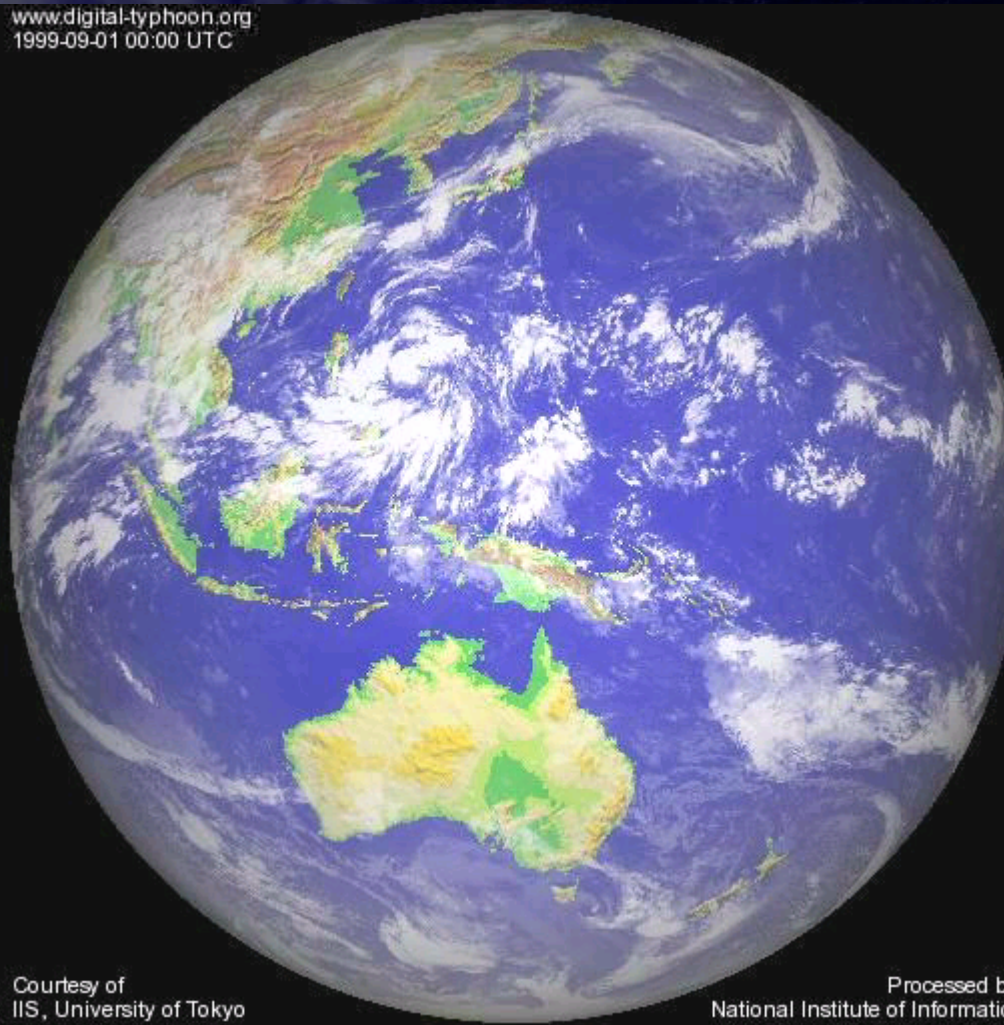


台風画像コレクションの特徴

1. 台風中心を追跡し、画像中心と位置合わせした**枠取り画像**を生成。
2. 自然界に存在する**複雑系パターン**の一例として、**カオス的な時間発展**を示す。
3. **注釈つき画像データコレクション**として種々の**テストベッド**にも利用できる。

全球画像アニメーション

www.digital-typhoon.org
1999-09-01 00:00 UTC

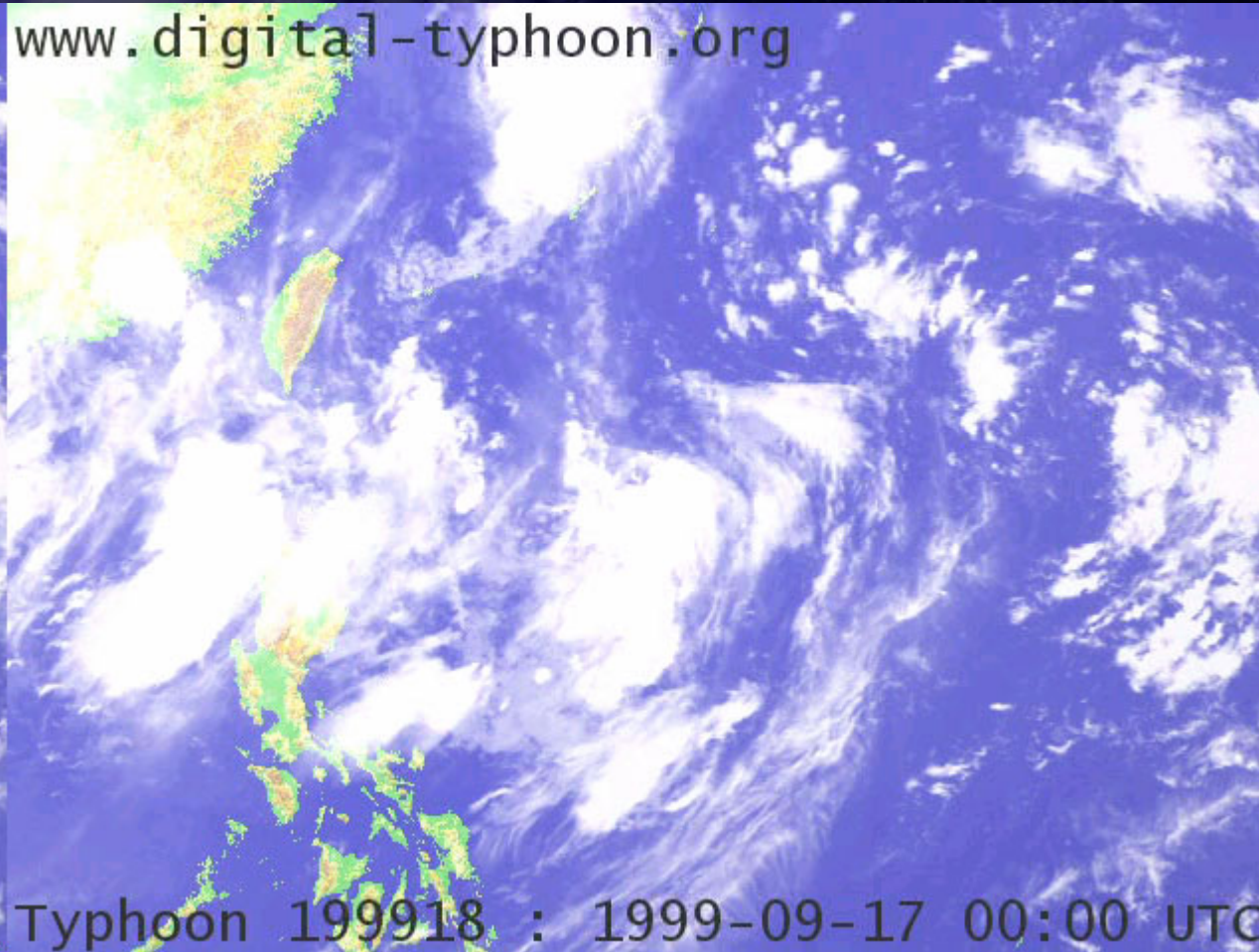


Courtesy of
IIS, University of Tokyo

Processed by
National Institute of Informatics

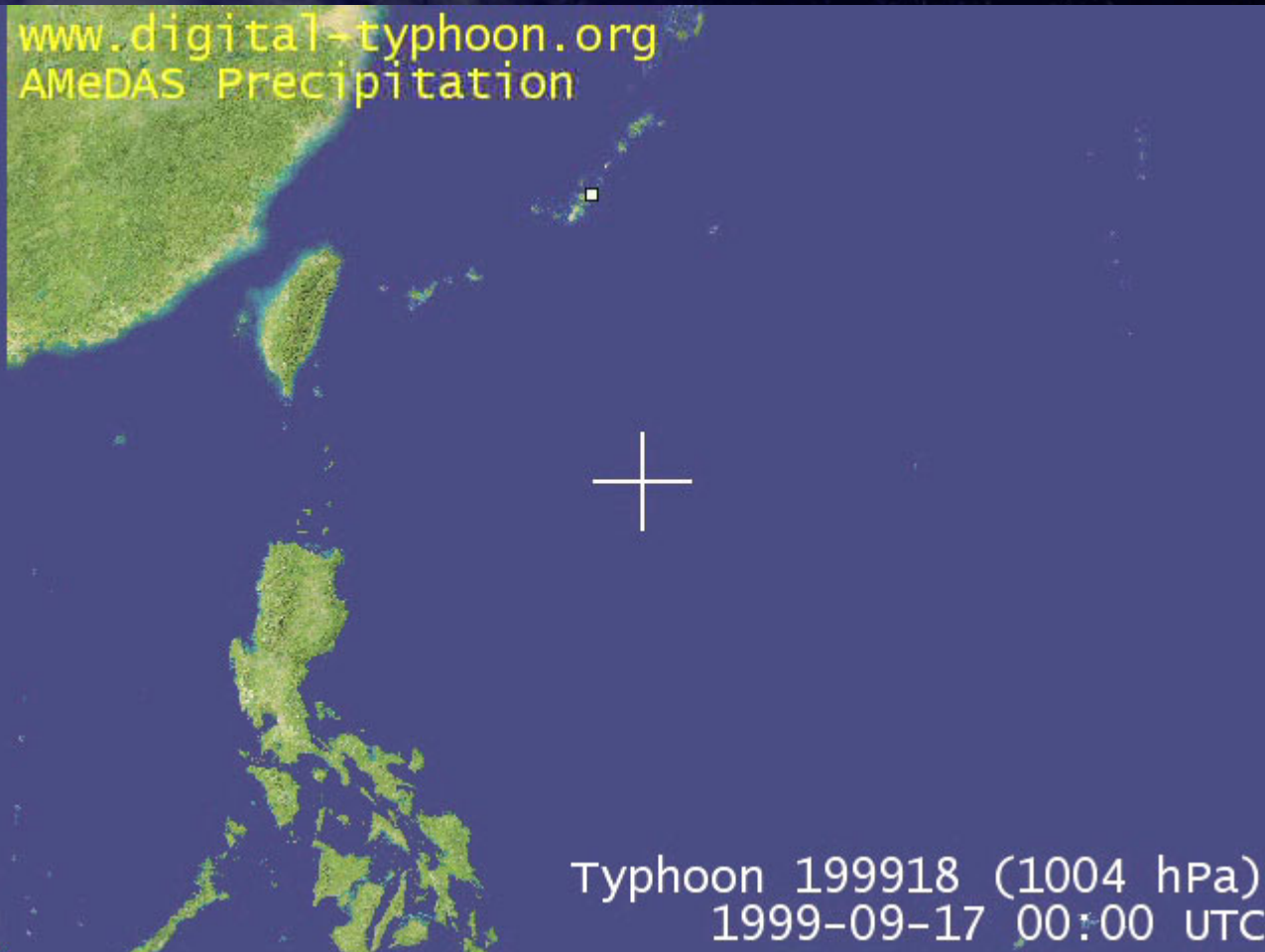
台風199918号の一生

www.digital-typhoon.org

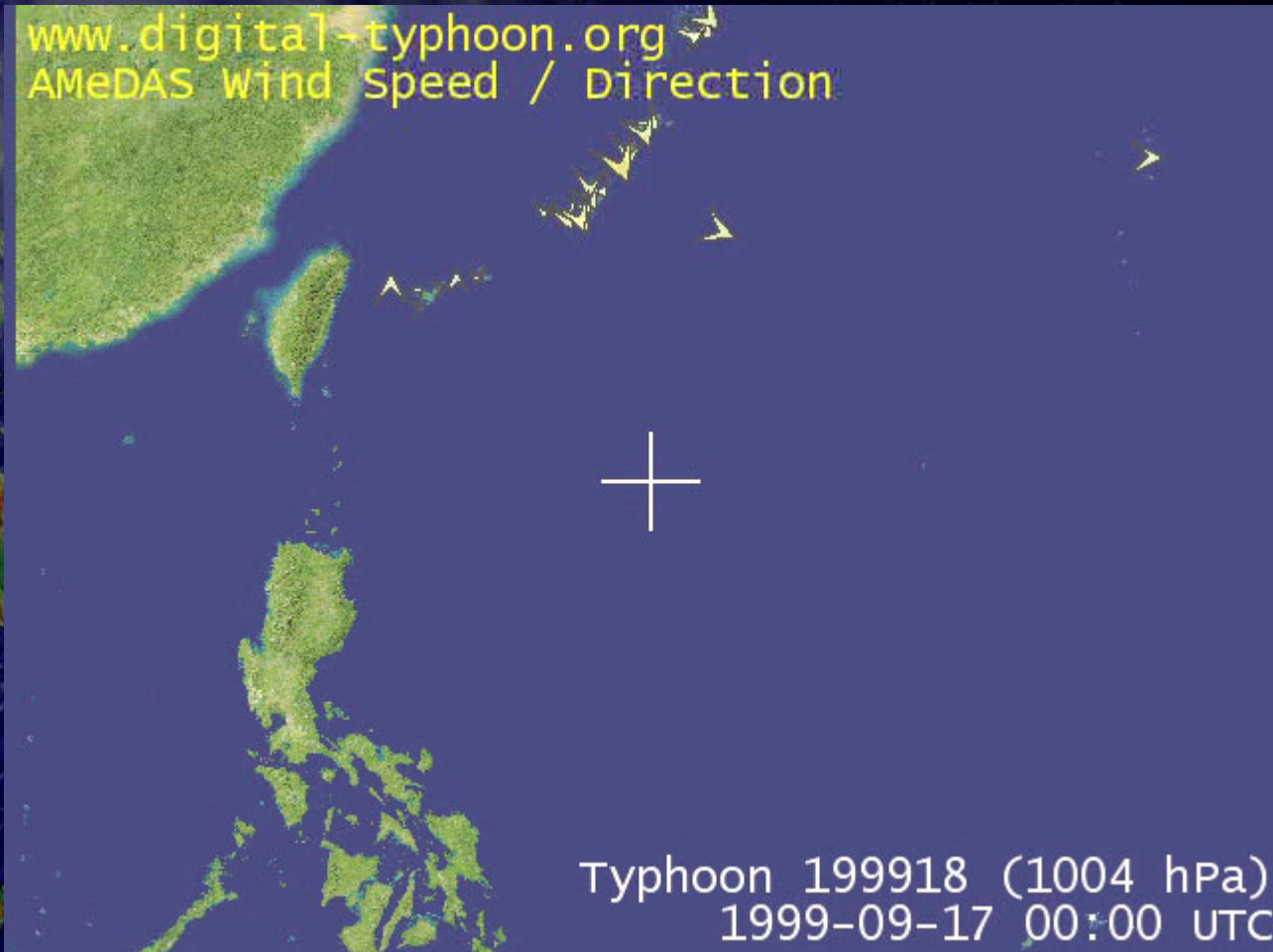


Typhoon 199918 : 1999-09-17 00:00 UTC

台風199918号に伴う降雨



台風19918号に伴う風



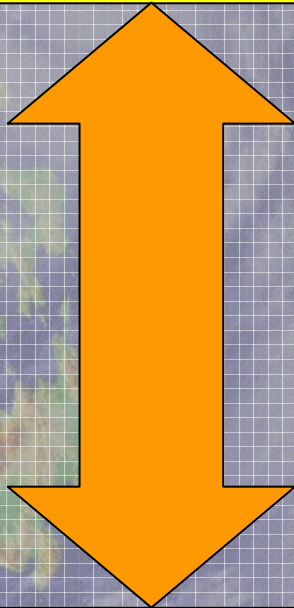
データベース基盤に基づく科学研究

- 生物情報学・バイオインフォマティクス
 - ーヒューマンゲノム計画
- 天文学・宇宙科学
 - ースカイサーベイ計画

科学の様々な分野において、**大規模データベース基盤の整備と活用**が、研究における必須の要素となりつつある。

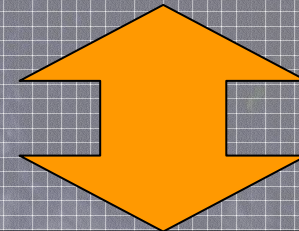
階層モデル

データ解析

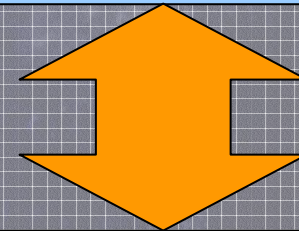


データ測定

データ解析



データベース



データ測定

科学目的の場合、データベースは公共的。

データ基盤構築の背景

- センシング技術およびデータ蓄積技術の発達により、従来よりも格段に大量のデータが生成できるようになった。
- 網羅的な測定に基づき、測定対象をセンサ上ではなくコンピュータ上で選択する考え方が一般的となった。

1. 大量データ管理のためのデータベース

2. 大量データ分析のためのマイニング

公共データベースの要件

1. 網羅性

- 解析に必要なデータを事前に測定。

2. 一貫性

- 全データを同等の品質・方法で測定。

3. 汎用性

- データ解析法独立のデータ定義方法。

4. 多様性

- データ検索法独立のデータ操作方法。

発表の概要

1. 「メテオインフォマティクス」とは
2. 「デジタル台風プロジェクト」の紹介
3. メテオインフォマティクスに関連する研究の紹介
4. メソスケール事象との関連
5. 今後の展開と課題

メテオインフォマティクスの課題

1. 高速なデータベース検索
2. 高度なデータベースモデル
3. データ記述・検索のための標準言語
4. 大量データからの情報抽出
5. 大量データの分類・要約・組織化
6. 大量データの可視化
7. 多様なデータの関連付け

データベースに関する課題

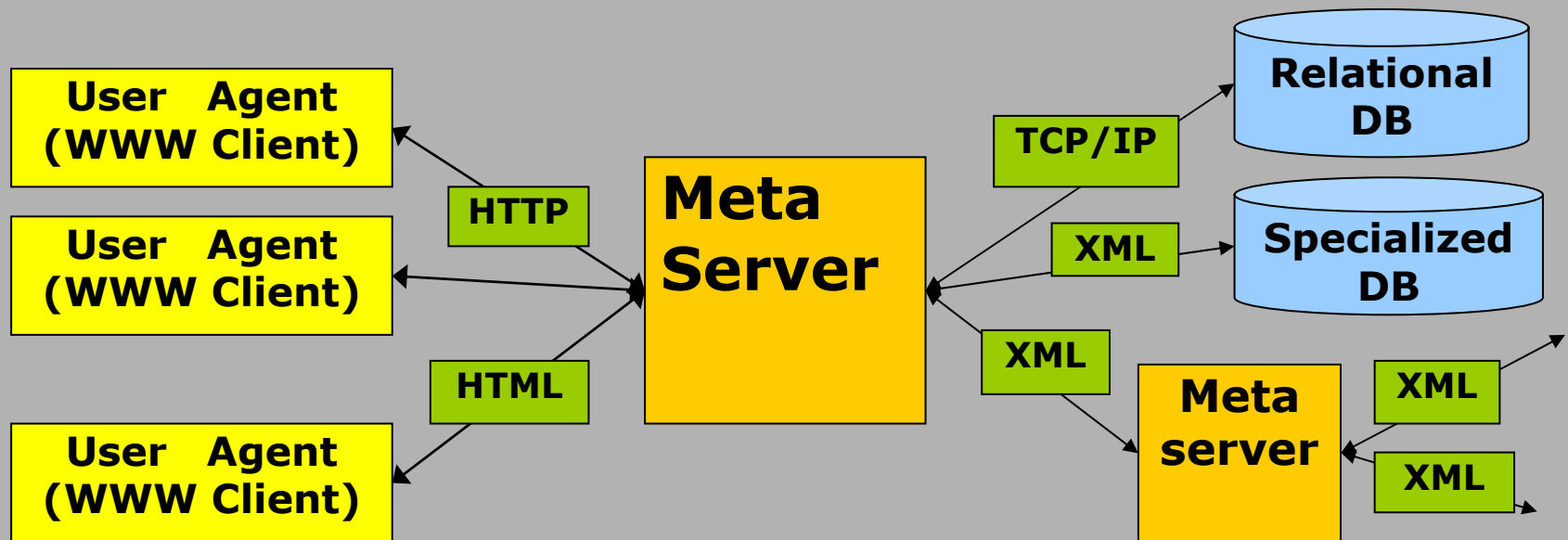
- 必要なデータをすばやく検索するための**インデックス構造**. 既に多くの研究がある.
- 多様な検索要求に応えうるような,
柔軟で強力なデータベースモデル.
本発表では扱わない.
- データ記述・検索のための**標準言語**.
XML構文を用いた方法を採用する.

現状のシステム構成(1)

- 関係データベース (PostgreSQL)
 - ー オープンソースのソフトウェア
- 画像特徴空間探索エンジン (FSE)
 - ー 独自に開発中の画像検索エンジン
- ウェブサーバ
- CGIプログラム群.

上記のコンポーネントを3層構造で結合.

現状のシステム構成(2)



すべてのリソースにウェブブラウザを通してアクセスできる分散データベースシステムを目指す.

画像特徴空間探索エンジン

- 画像の内容を指定することにより画像検索が可能な検索エンジン.
- 単に類似画像を検索するだけでなく、データの組織化や時系列画像もサポート(一部はまだ計画段階).
- 専用の問い合わせ言語を用いて、XML形式で検索要求を記述する.

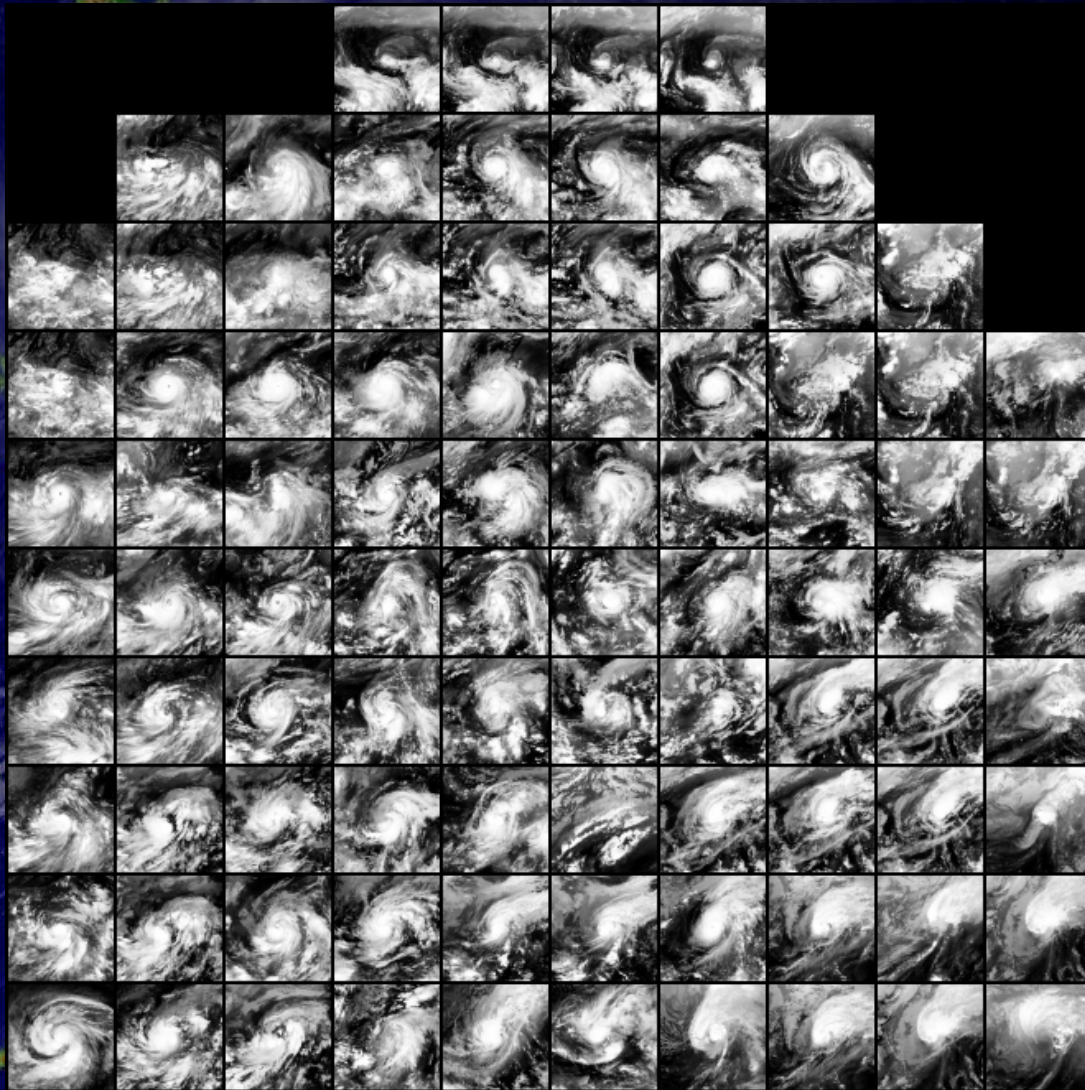
データ検索言語の例

```
<?xml version="1.0" encoding="UTF-8" ?> _ <envelope> _  
<header> <server port="59300">localhost</server> _ <session user="kitamoto"  
id="1"> <transaction>1</transaction> </session> </header>  
-<body> _ <query> _ <task> _ <example type="single">  
-<constant select="folder">9903</constant> <dynamic  
select="name">@random</dynamic> </example>  
  </task> _ <where> <filter select="folder" type="equals"  
not="1">@example</filter> </where> <sort-by order="ascending">value</sort-  
by> _ <fetch> <from>0</from> <size>5</size> </fetch> _  
<return> <select>folder</select> <select>name</select>  
<select>value</select> <select>example</select> </return> _ <group-  
by> <select>folder</select> </group-by> _ <for-each> _ <task> _ <let  
variable="value"> <function target="example">distance</function> _ <metric  
type="euclid"  
option="squared"> <min>0</min> <max>30</max> </metric> </let>  
  </task> <sort-by order="ascending">value</sort-by> _  
<fetch> <from>0</from> <size>2</size> </fetch> </for-each> </query>  
</body>  
</envelope>
```

大量データからの情報抽出

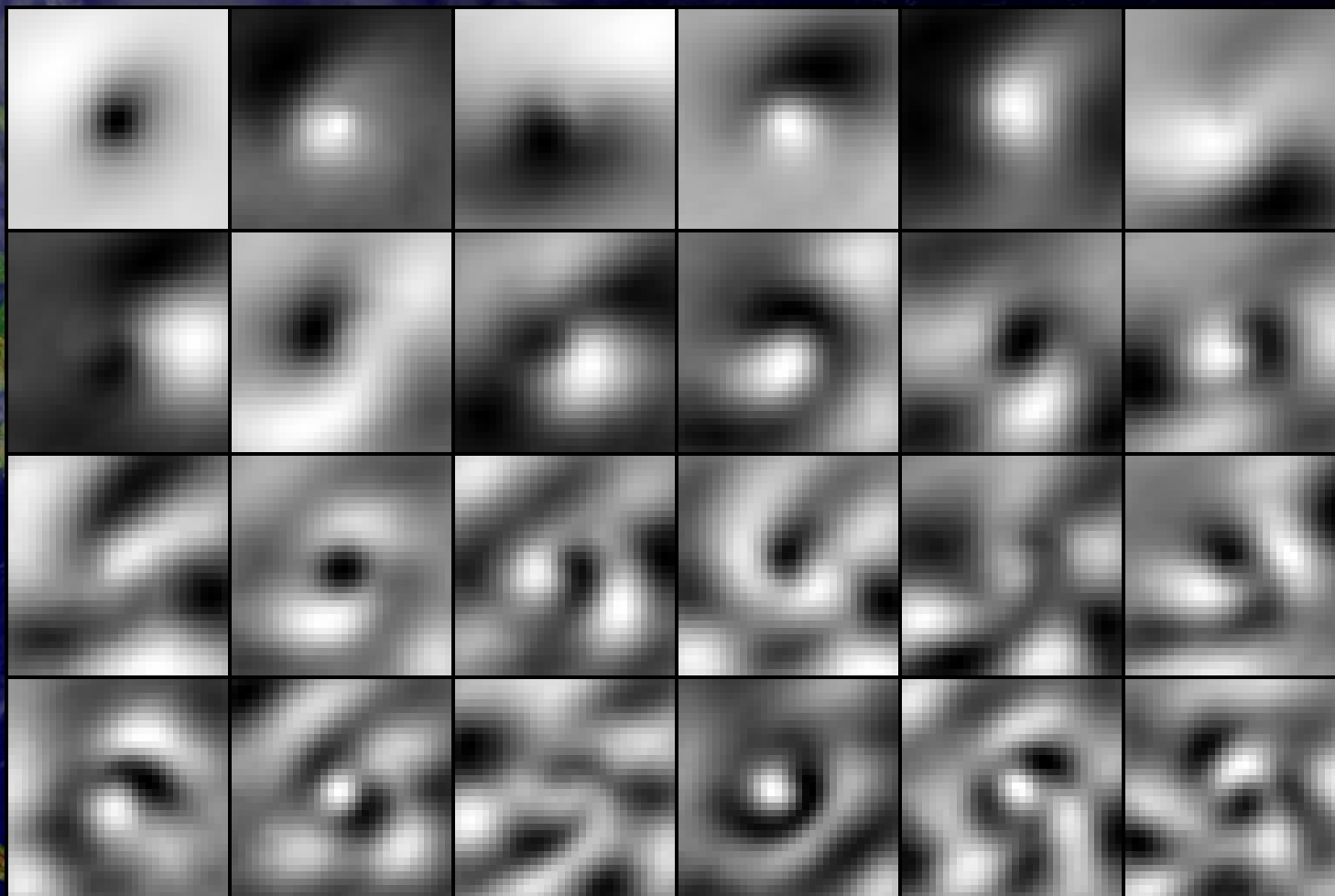
- データから重要な信号(パターン)を抽出し、以後の解析・検索に用いる.
- 信号処理的手法(信号の数学的性質のみを考慮)、画像認識的手法(視覚の特性を考慮)など、多数の手法が提案されている.
- ここでは信号処理的手法として、主成分分析を、画像認識的手法として、スパイラルバンドの検出について述べる.

主成分分析

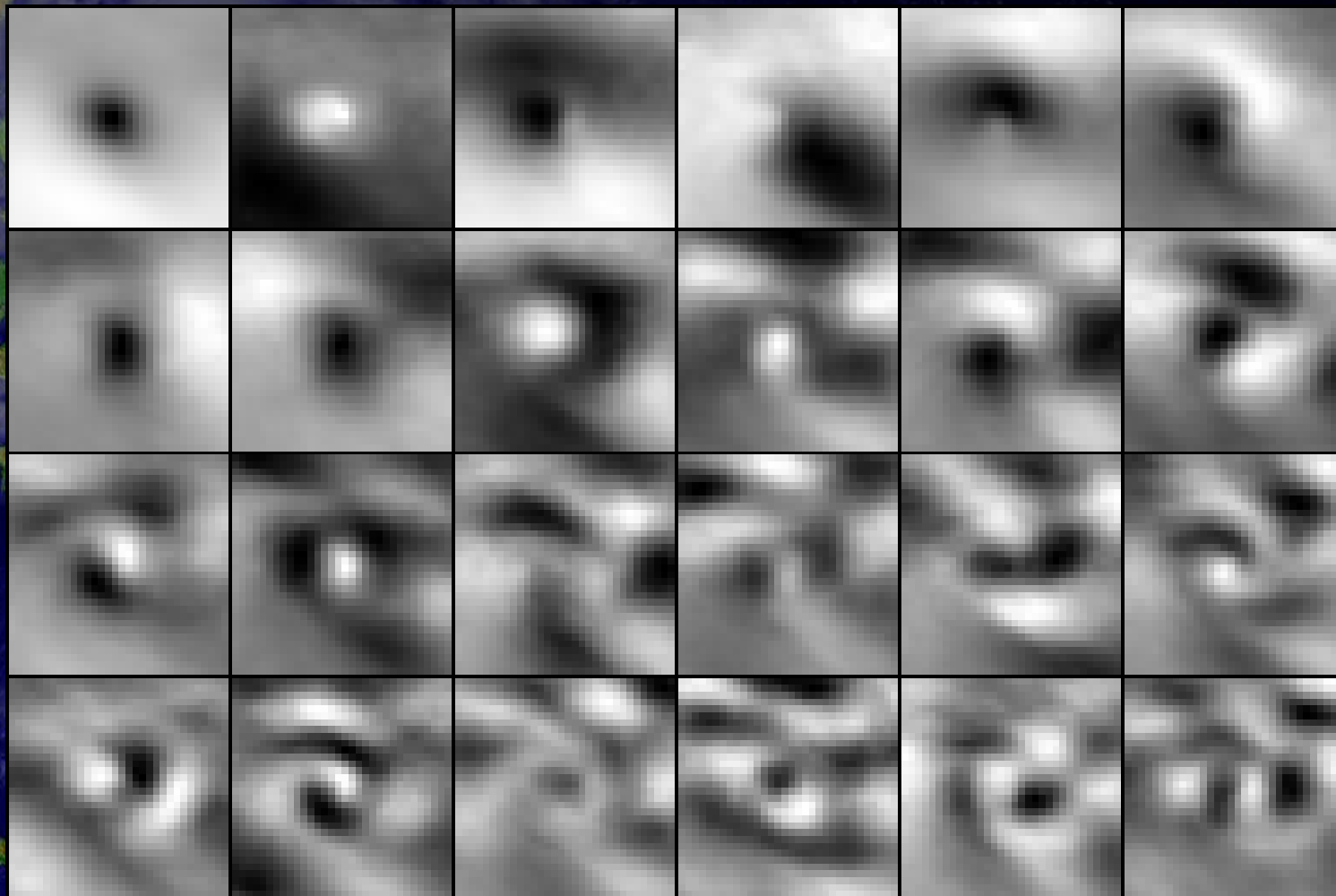


- データ集合の分散を最大に保存するような方向を見出す数学的手法.
- 2個の主成分が張る空間上に画像を配置することで, 主成分の「特徴」を把握できる.

固有台風画像(北太平洋)



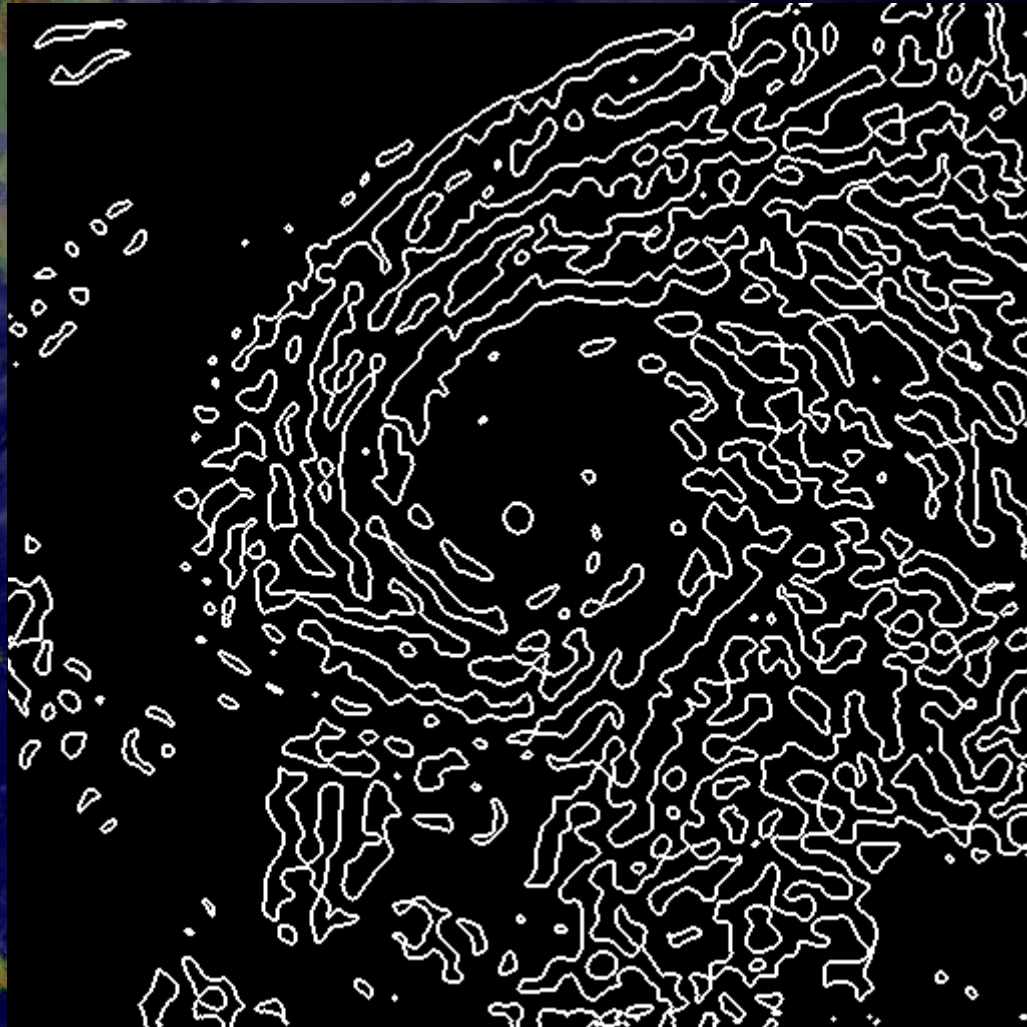
固有台風画像(南半球)



スパイラルバンドの検出

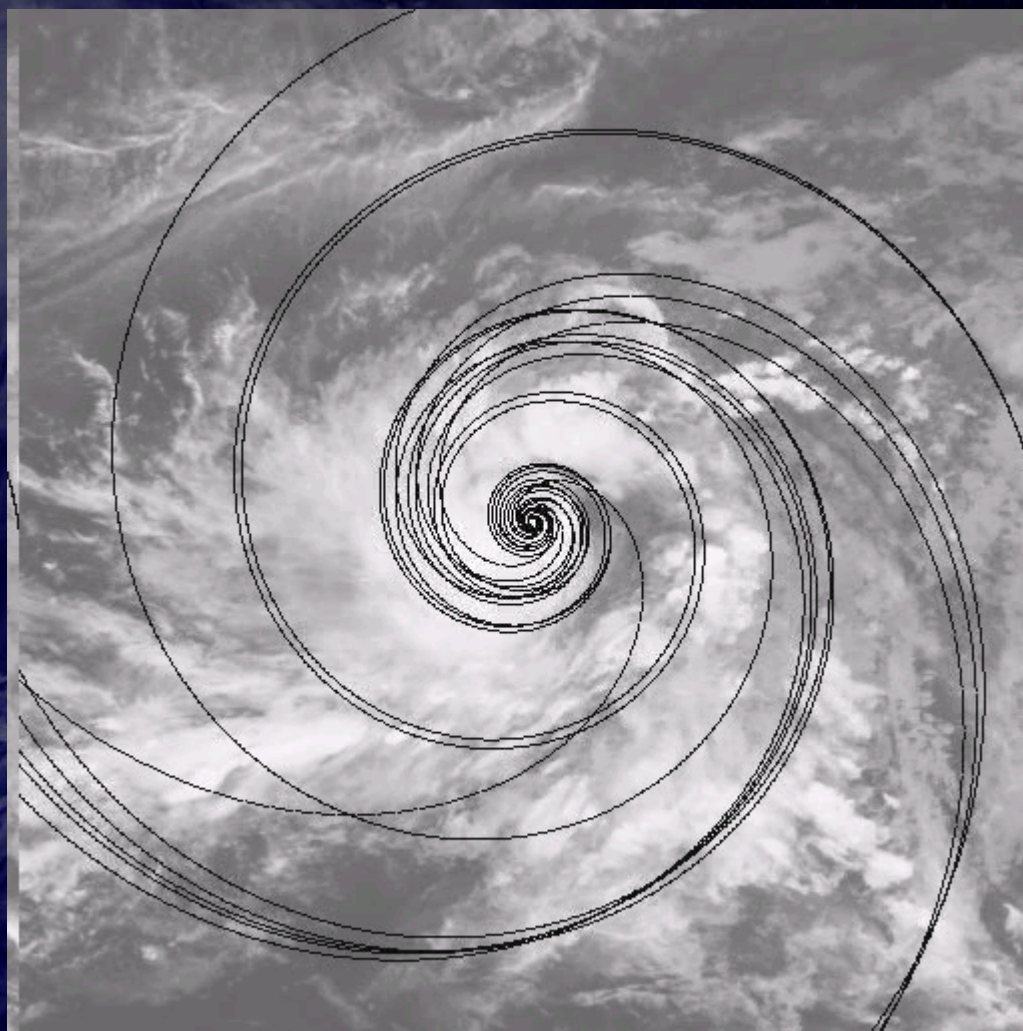
1. 直線検出のためのアルゴリズムとして有名な「ハフ変換」を改良.
2. 異なる角度や位相をもつ複数個のらせんを同時に検出する.
3. らせんは画像のエッジを追跡する.
4. カルマンフィルタ等の時系列モデルとの統合が有効であると予想される.

エッジ画像

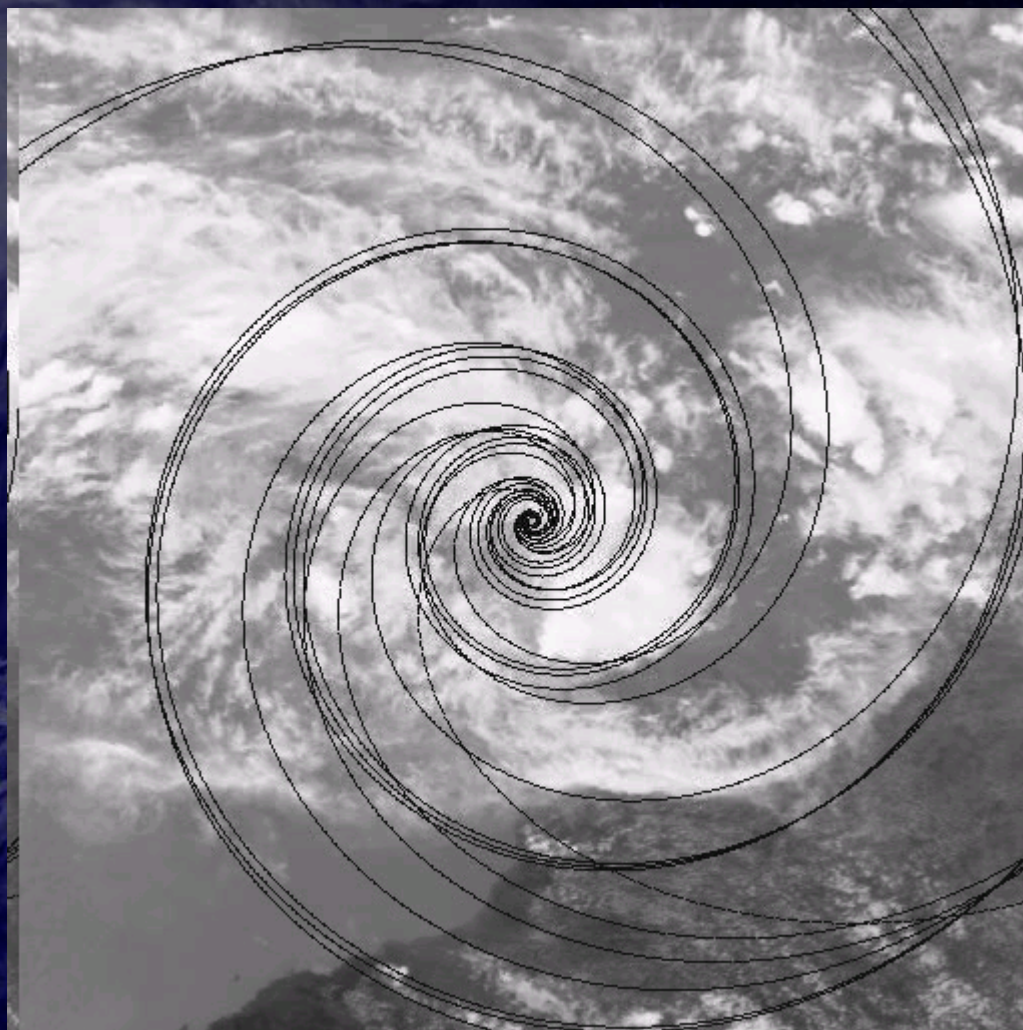


- 雲とそれ以外の大まかな境界線に相当する.
- この画像から, 投票原理によりらせんを検出.

台風200302号(北太平洋)



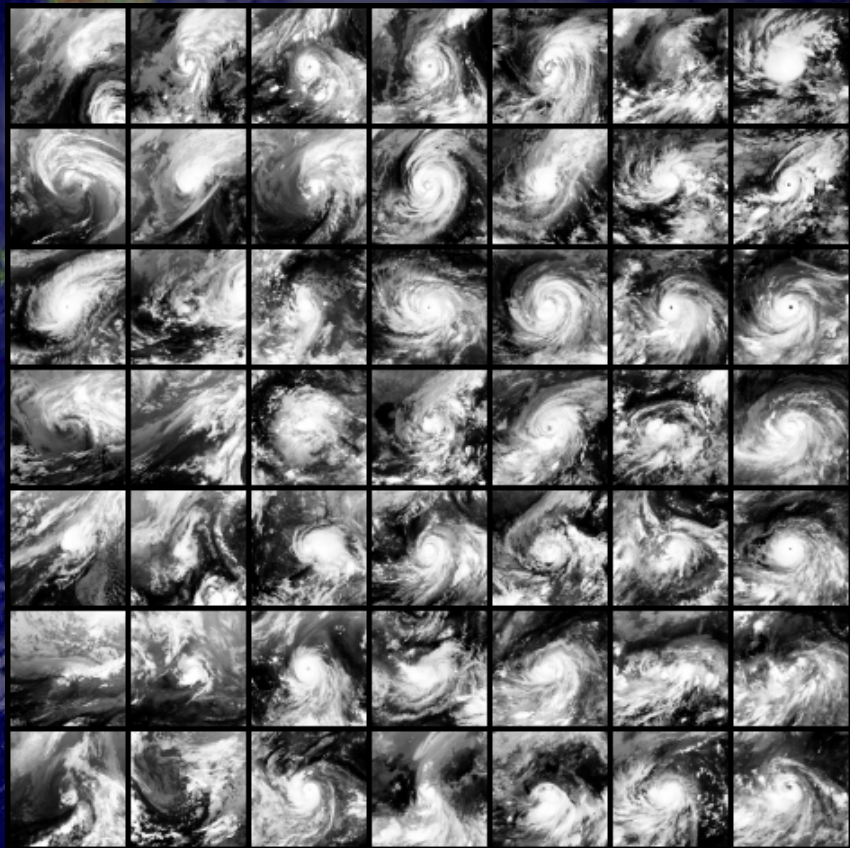
台風199902号(南半球)



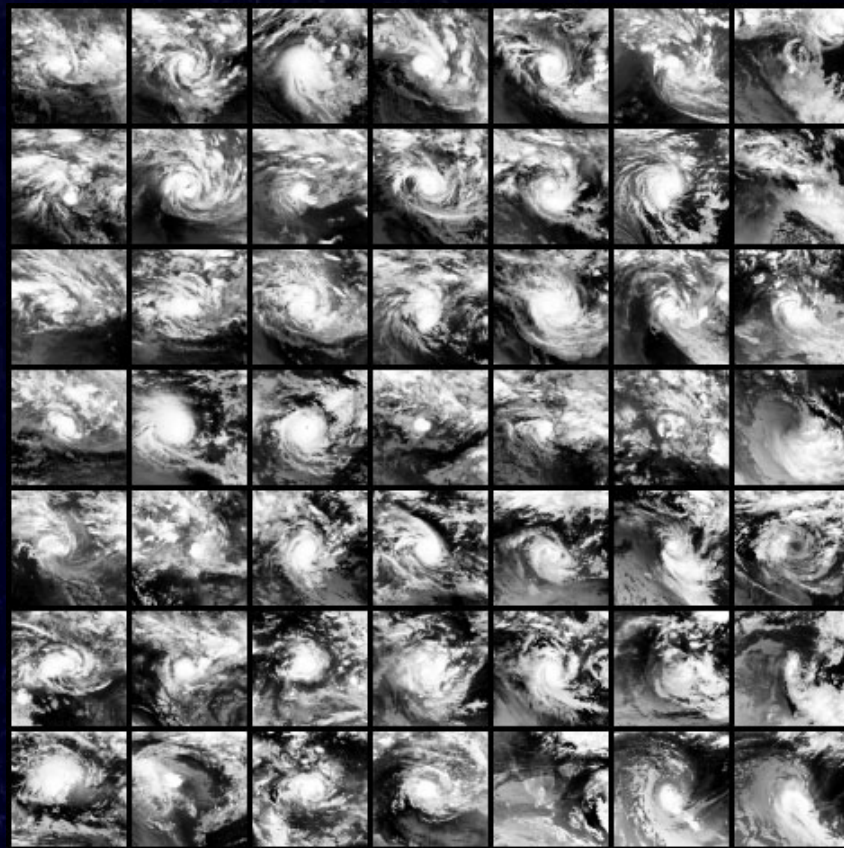
大量データの分類・要約・組織化

- 似たデータをまとめる(クラスタリング)
- 冗長なデータを簡潔にする(要約)
- 似たデータを近くに並べる(組織化)
- 本来は連続した空間を離散的に見ることにより、人間は理解しやすく、数学的にも扱いやすくなることもある。
- 災害時の情報集約に有用かもしれない。

クラスタリング (K-Means)

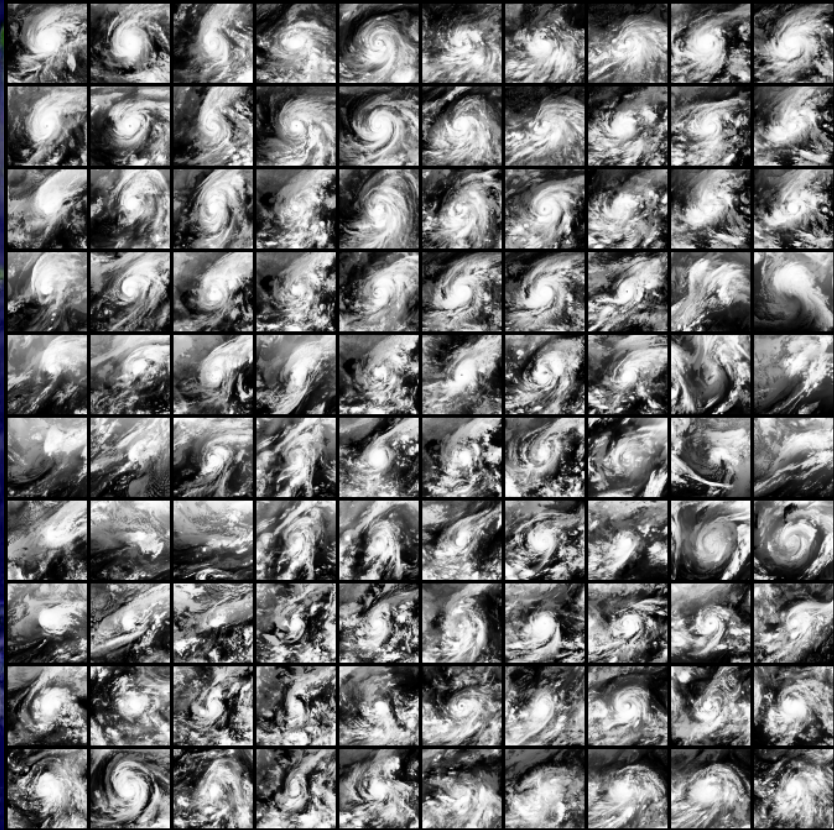


Northern Hemisphere

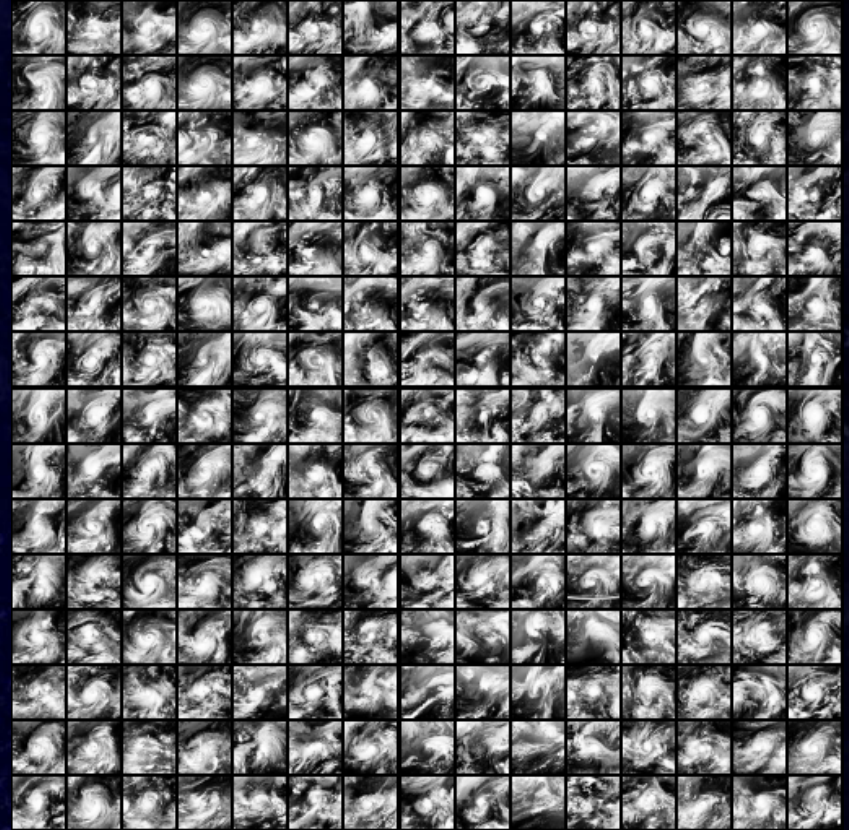


Southern Hemisphere

その他のクラスタリング手法

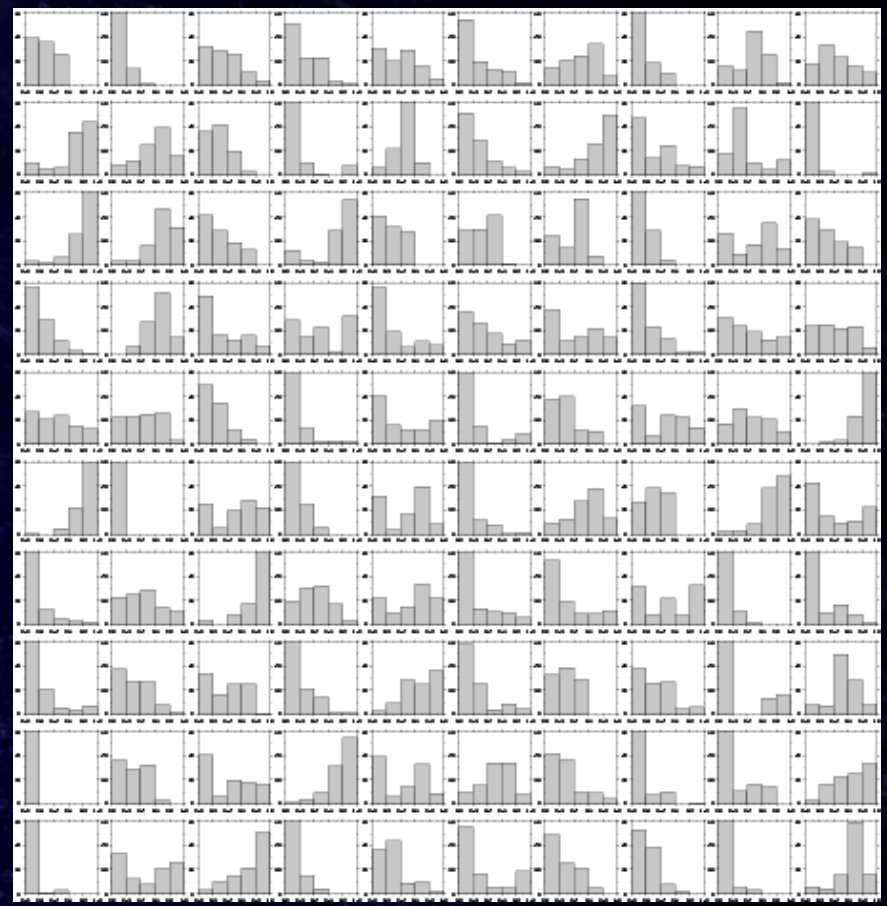
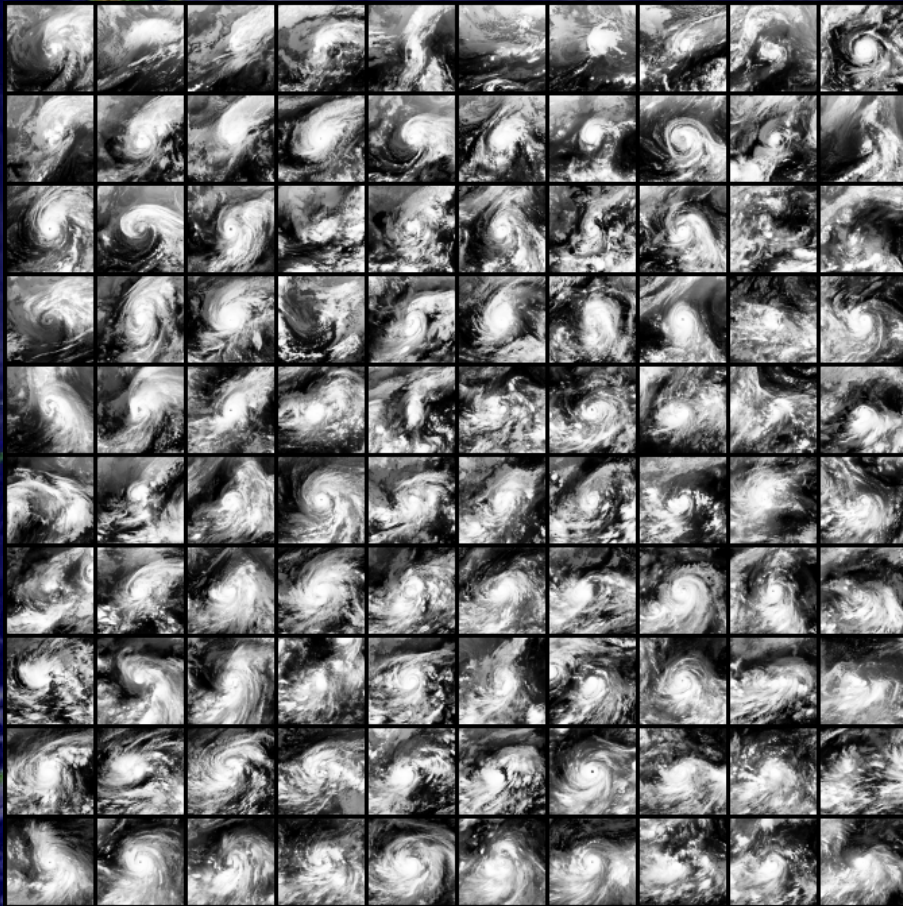


Self Organizing Map



Generative
Topographic Mapping

クラスタリングと台風齢の関係



Clustering by K-means

Age Distribution

データの比較

- 類似事例をデータベースから検索し、意思決定の補助資料として活用。
- 「類似性」をどのように数学的に定義するかが大問題。
- 適切な定義には、専門家の知識の直接的・間接的なモデル化が必須であるが、実際にはこのモデル化は困難。

類似した画像時系列の発見

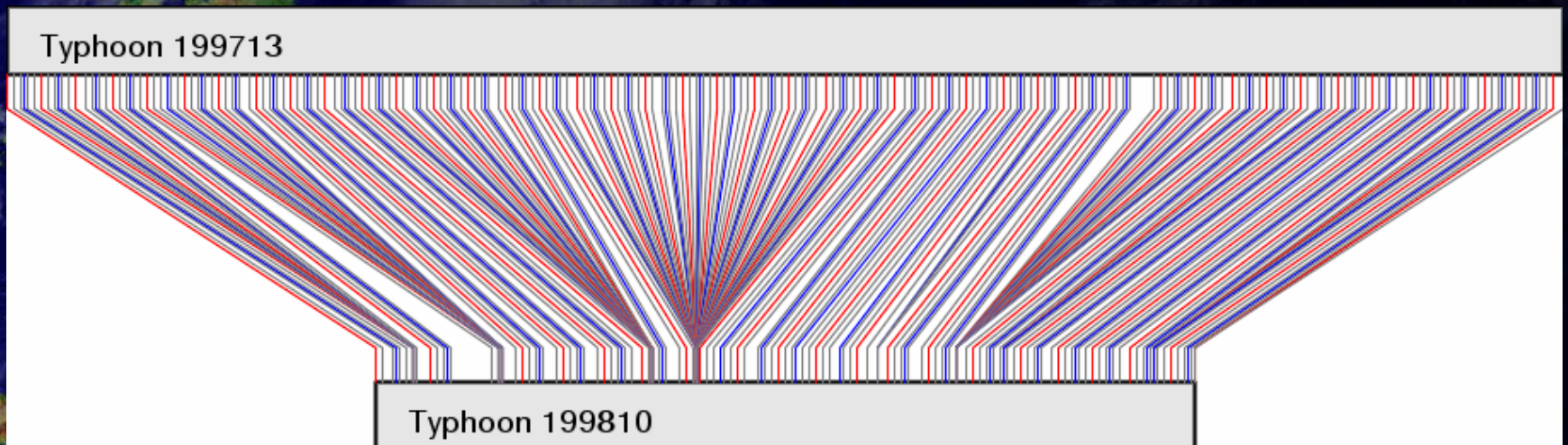
複数の台風時系列が似たような時間変化パターンを示すことがある。

そのような断片をつなぐことで台風のライフサイクルをモデル化したい

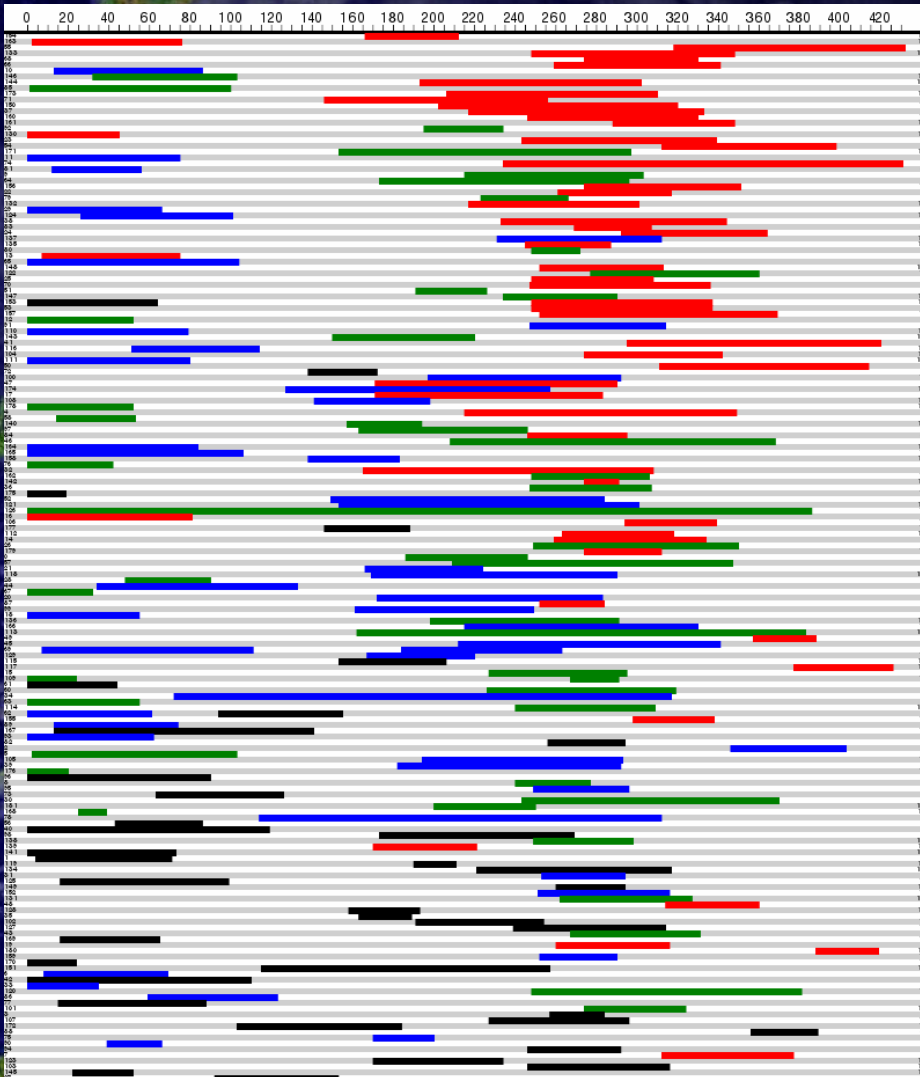
動的計画法と階層的クラスタリングによるマルチプルアラインメント

ペアワイズ・アライメント

- 二つの台風時系列画像をグローバルに整列し、類似した時間変化を抽出.
- 動的計画法が典型的な解法.
- マッチングスコア最小のパスを計算.

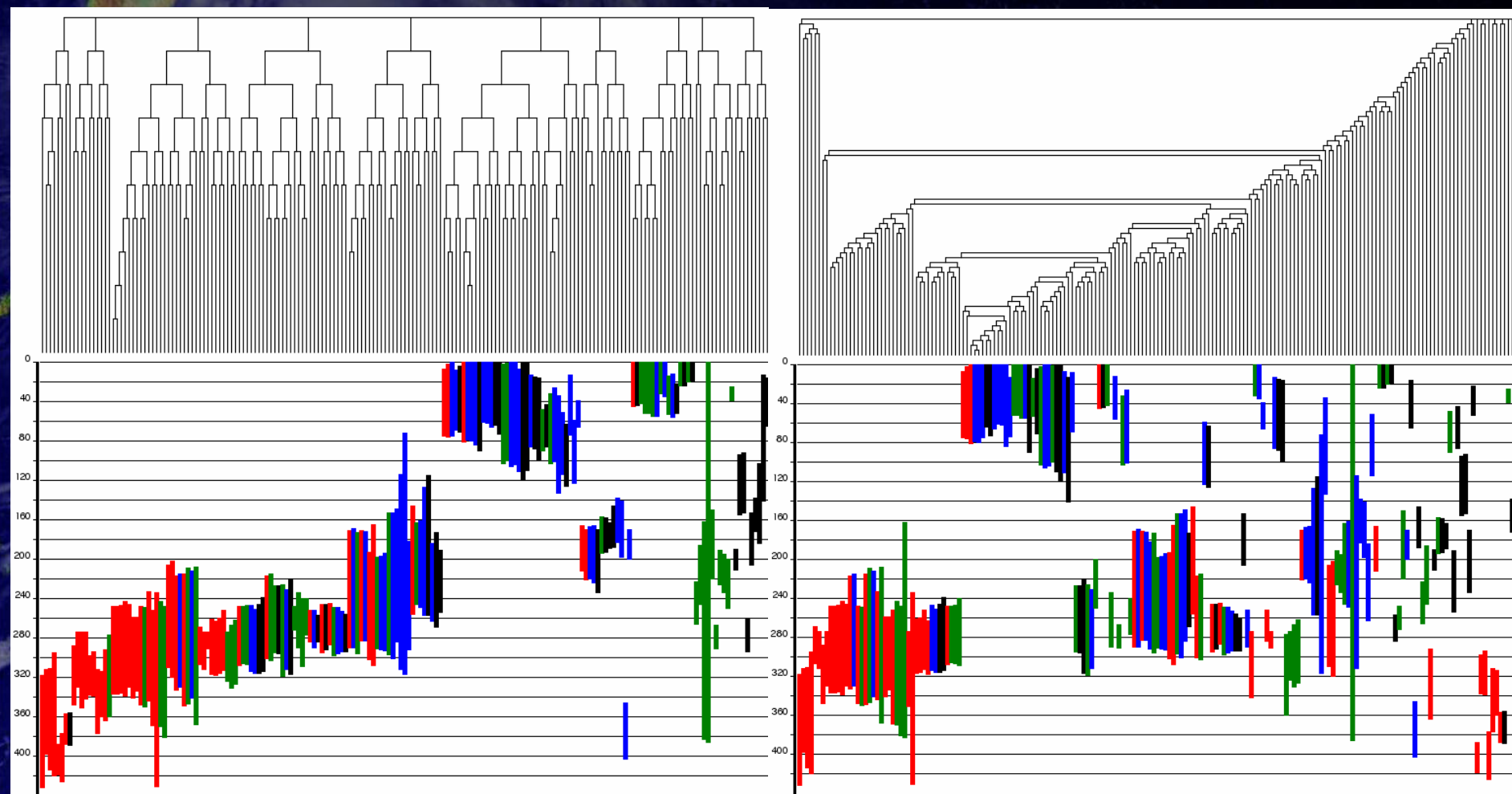


アラインメント第一ステップ



- ピボット系列を基準にして、他の系列の断片を整列.
- 両系列で平行にマッチングする部分のみを抽出.
- グローバルに類似した系列を上に表示している.

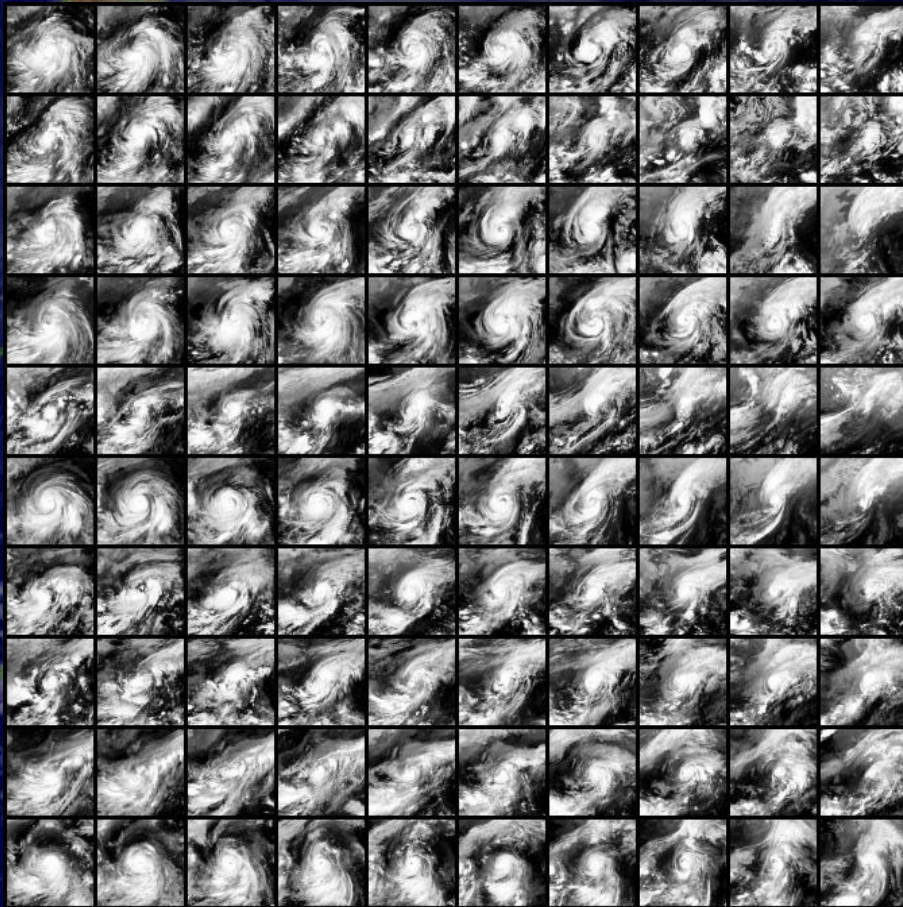
アラインメント第二ステップ



階層的クラスタリング

- 類似するフラグメント同士をグルーピングし、類似部分系列を発見したい.
- 距離の近いものからグルーピングする階層的クラスタリング手法を活用.
- 漸化式の違いにより、いくつかの異なる結果が得られる.
- Ward法を用いてクラスタリング.

アラインメント結果

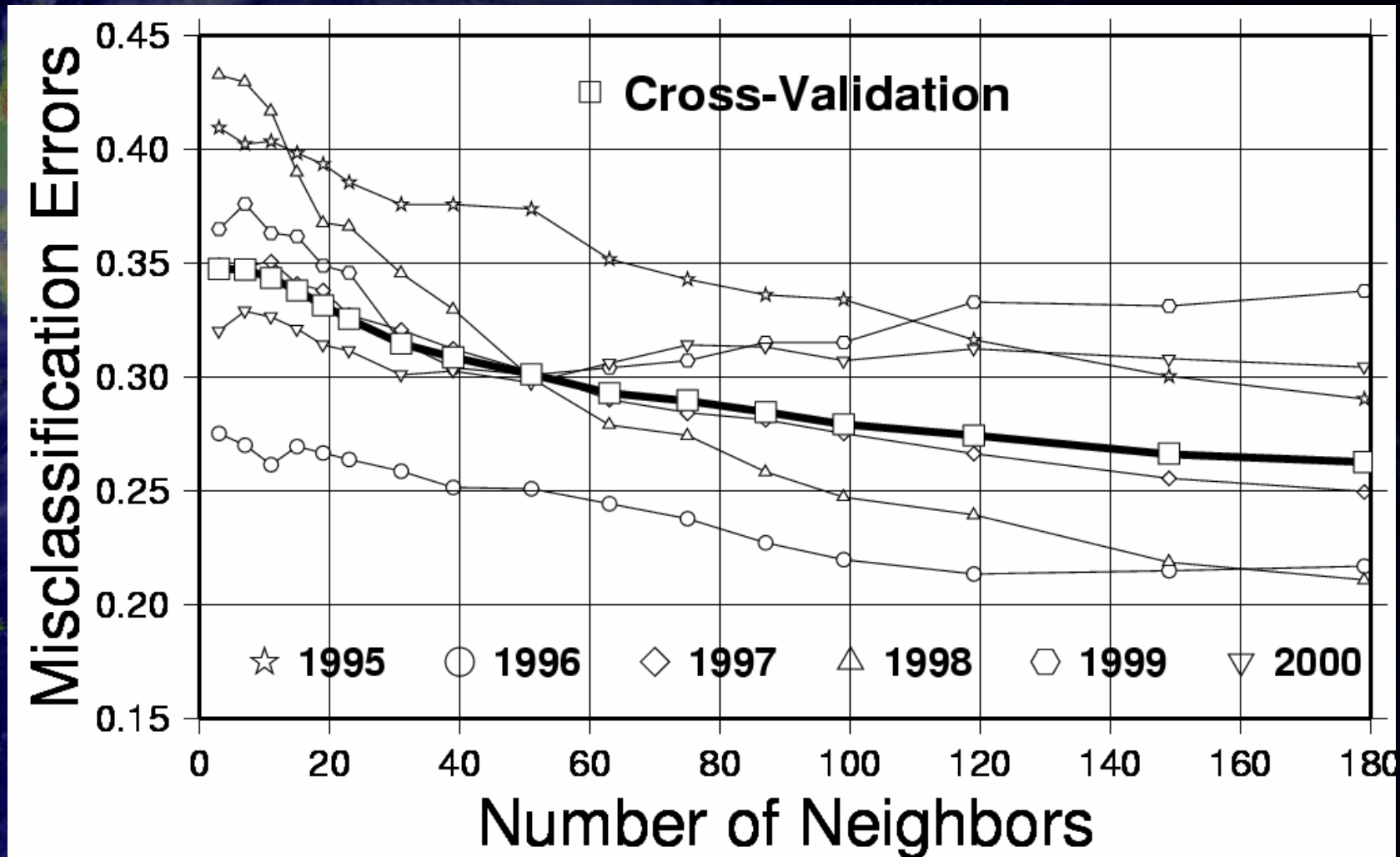


- 発生期から衰弱期までの典型的な台風ライフサイクルを表現する断片集合。
- 個々の画像の意味づけや時空間パターンの変化をより明確に把握可能。

発達する台風と衰弱する台風の分類

- 急速発達現象 – 24時間以内に中心気圧が30-40 hPa以上低下する現象.
- それが起こる前に、台風雲パターンからその「予兆」を発見することは可能だろうか？
- もうすこし簡単な問題： ある程度は発達・衰弱する台風を雲パターンのみで分類する.
- しきい値は $|10\text{hPa}| / 24\text{hour}$.

K-NN法による分類結果



nu-SVMを用いた分類結果(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

nu-SVMを用いた分類結果(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

分類結果

- 雲パターン情報のみを用いて、発達する台風と衰弱する台風を分類すると、その精度は75%程度となった.
- 画像特徴量としてより有効なもの（例えば時間変化情報）を用いれば、さらに精度は向上する可能性がある.
- 特に急速に発達しそうな台風の予兆を、事前に発見する方法に適用したい.

大量データの可視化

- 可視化（ビジュアライゼーション）とは、データをわかりやすく表示し、データに対する人間の理解を深めること。
- 複数視点から眺める、重要部分を抜き出し表示する、などの機能が重要。
- 画像解析技術とコンピュータグラフィクス技術との融合。

気象衛星画像の3次元表示

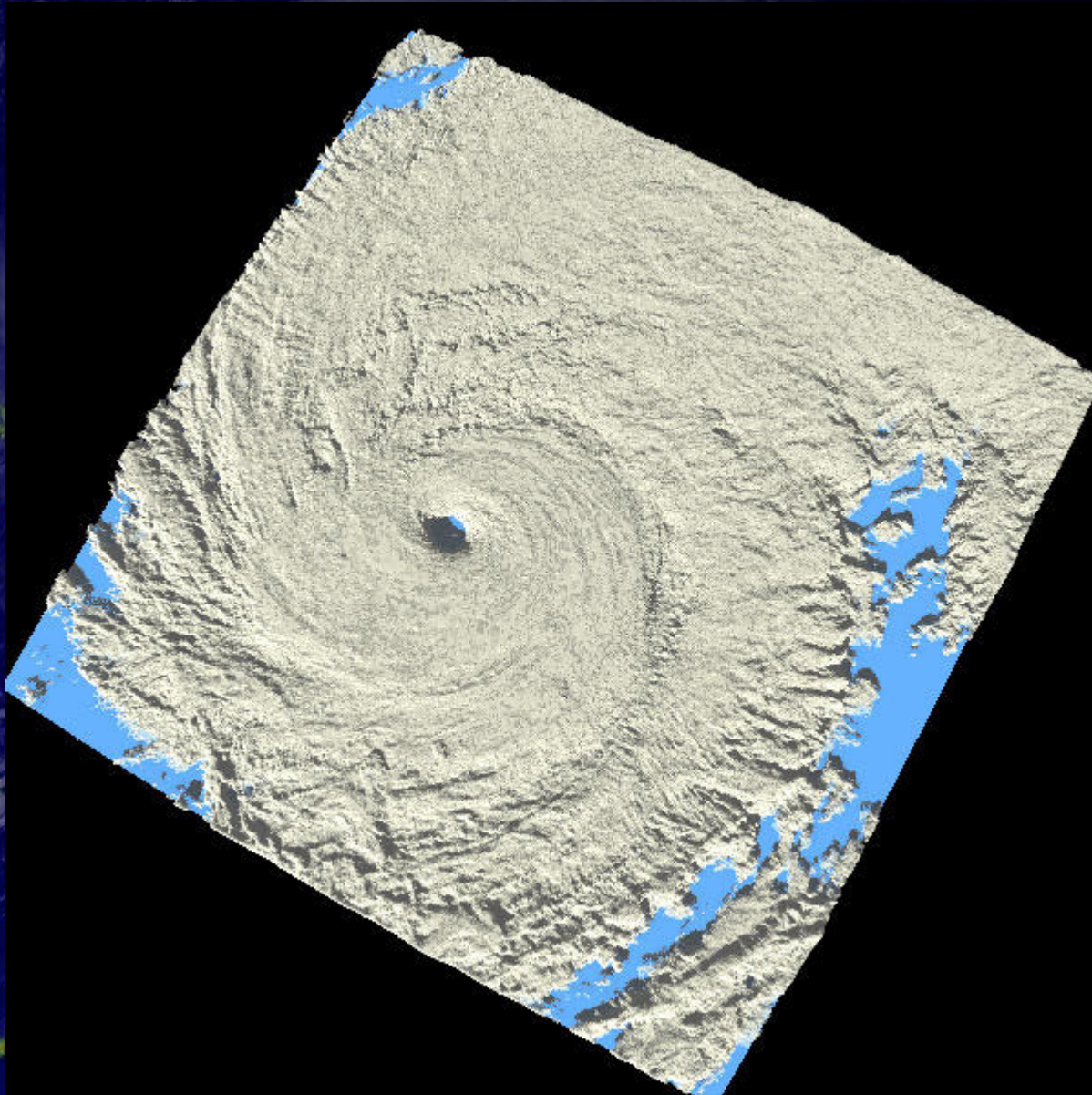
1. 第一世代

- ボリュームレンダリングを用いる方法.
- 雲の分類がされておらず, 光輸送の扱いも正しくない.

2. 第二世代(開発中)

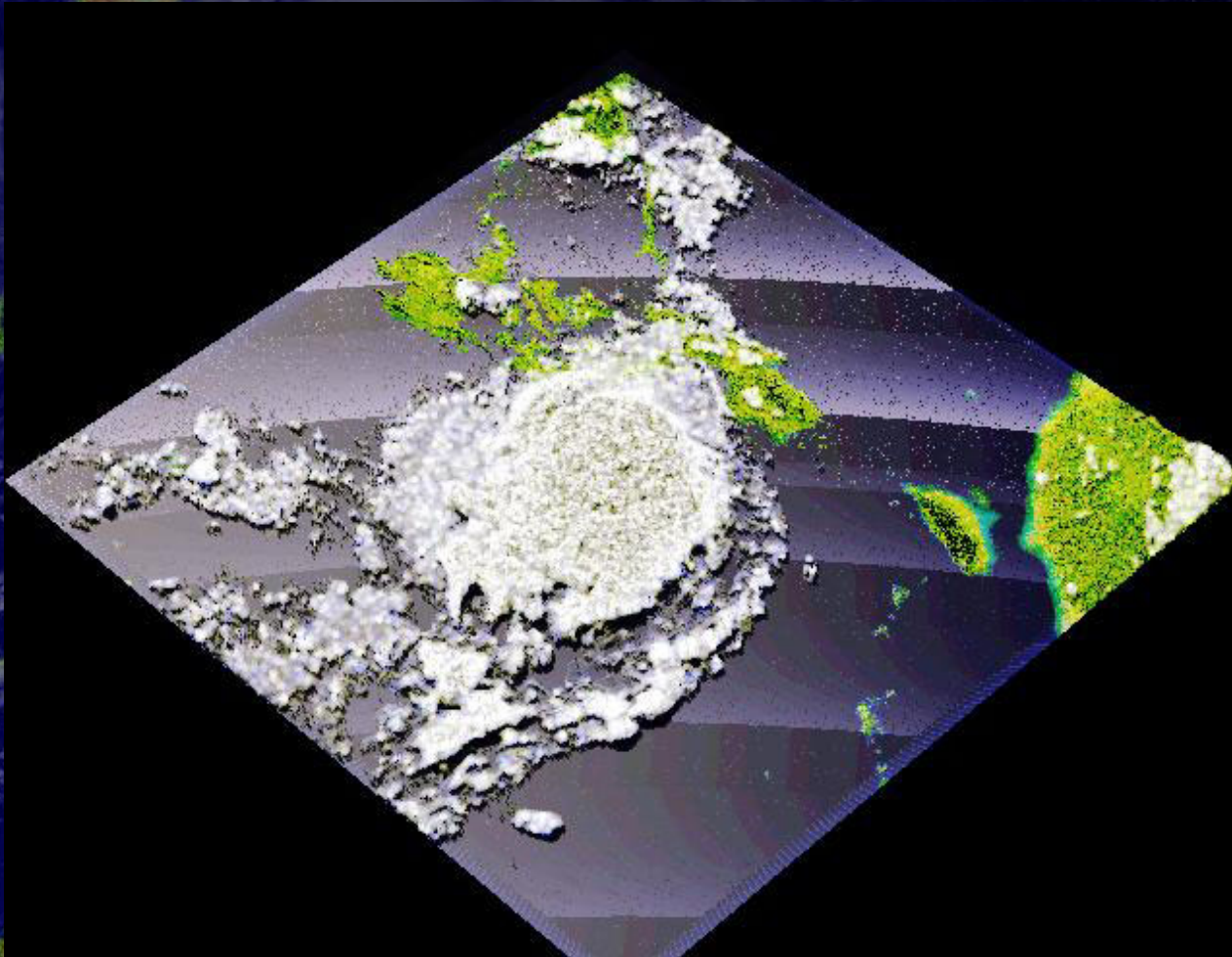
- フォトンマッピングを用いる方法.
- 雲の分類により, 雲の幾何学的厚さ・光学的厚さをより適切な値に設定.
- 雲に関する初歩的なmicrophysicsを考慮した光輸送の適切な推定.

第一世代



- ボリュームレンダリングを用いる方法.
- 雲の立体構造はわかるが、実物よりも硬い感じがする.

第二世代



- 雲の立体構造が直感的にわかれば、雲の湧き出しなどを実感でき、意思決定に有効な情報になる

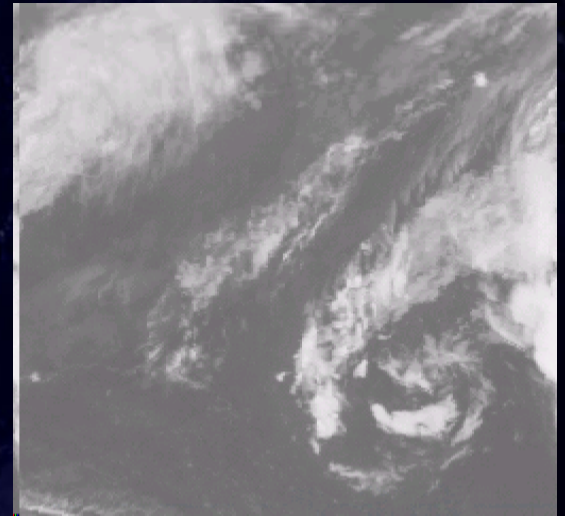
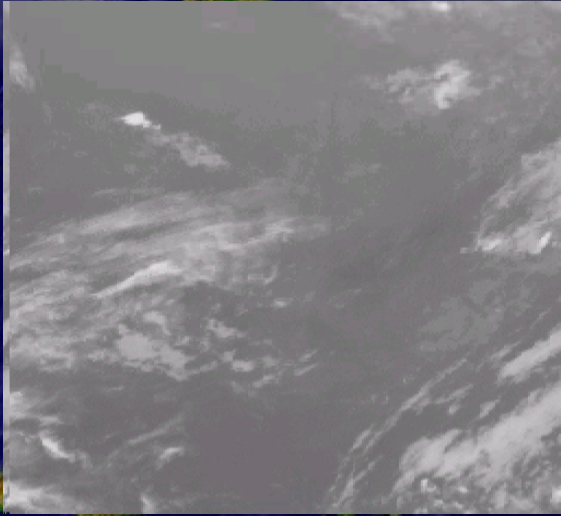
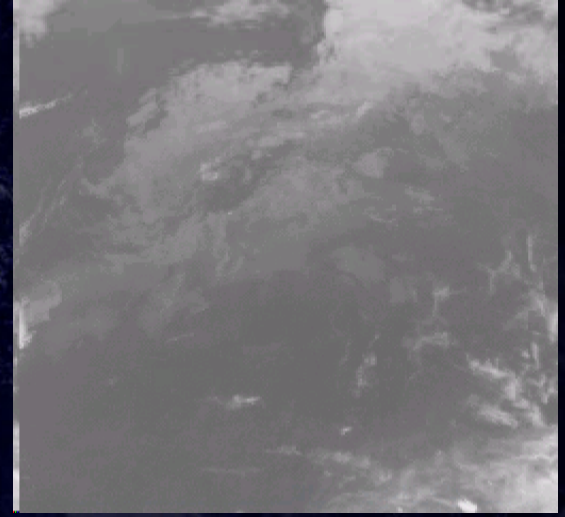
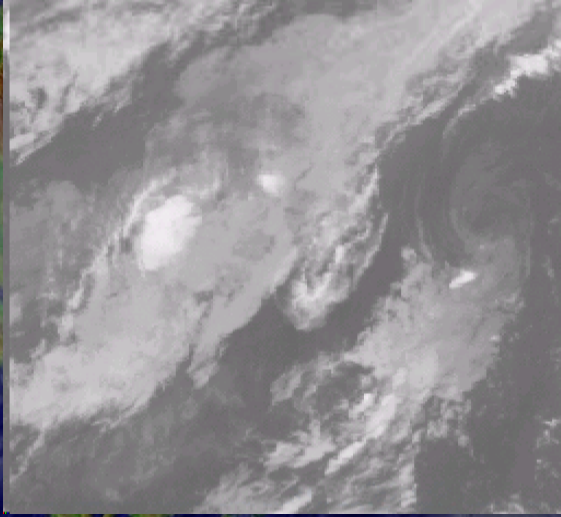
発表の概要

1. 「メテオインフォマティクス」とは
2. 「デジタル台風プロジェクト」の紹介
3. メテオインフォマティクスに関連する研究の紹介
4. メソスケール事象との関連
5. 今後の展開と課題

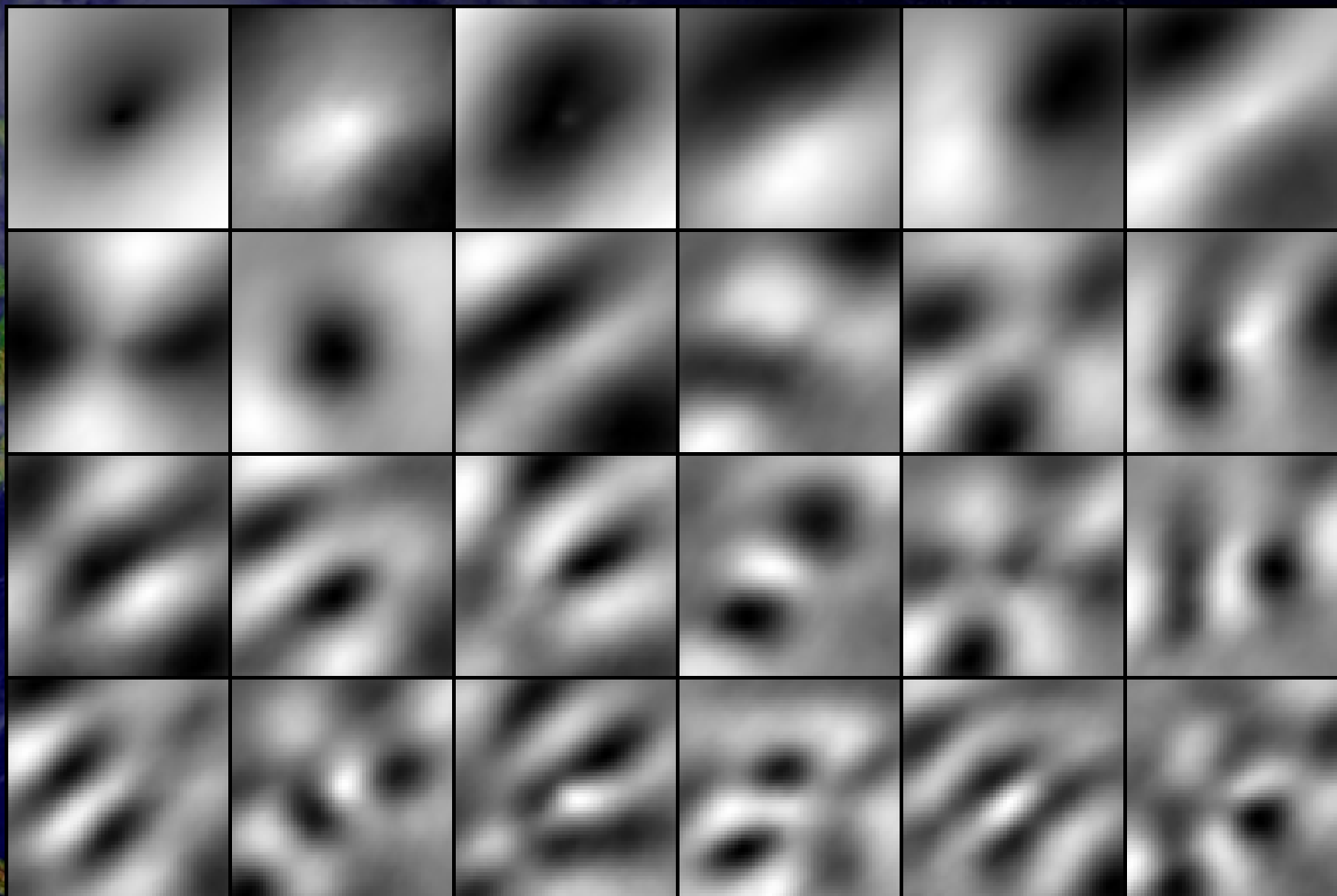
メソスケール現象への適用

- メソスケールの現象は、数値モデルで再現するのが難しい場合がある。
- いわゆる「集中豪雨」を対象として、画像上で何が観察できるかを探る。
- アメダス観測点において50mm/h以上を記録した事例を網羅的に収集する。
- 1995年から2002年に1779事例(24時間に複数回の場合は1回に統合する)

アメダス観測点44131の事例



主成分分析

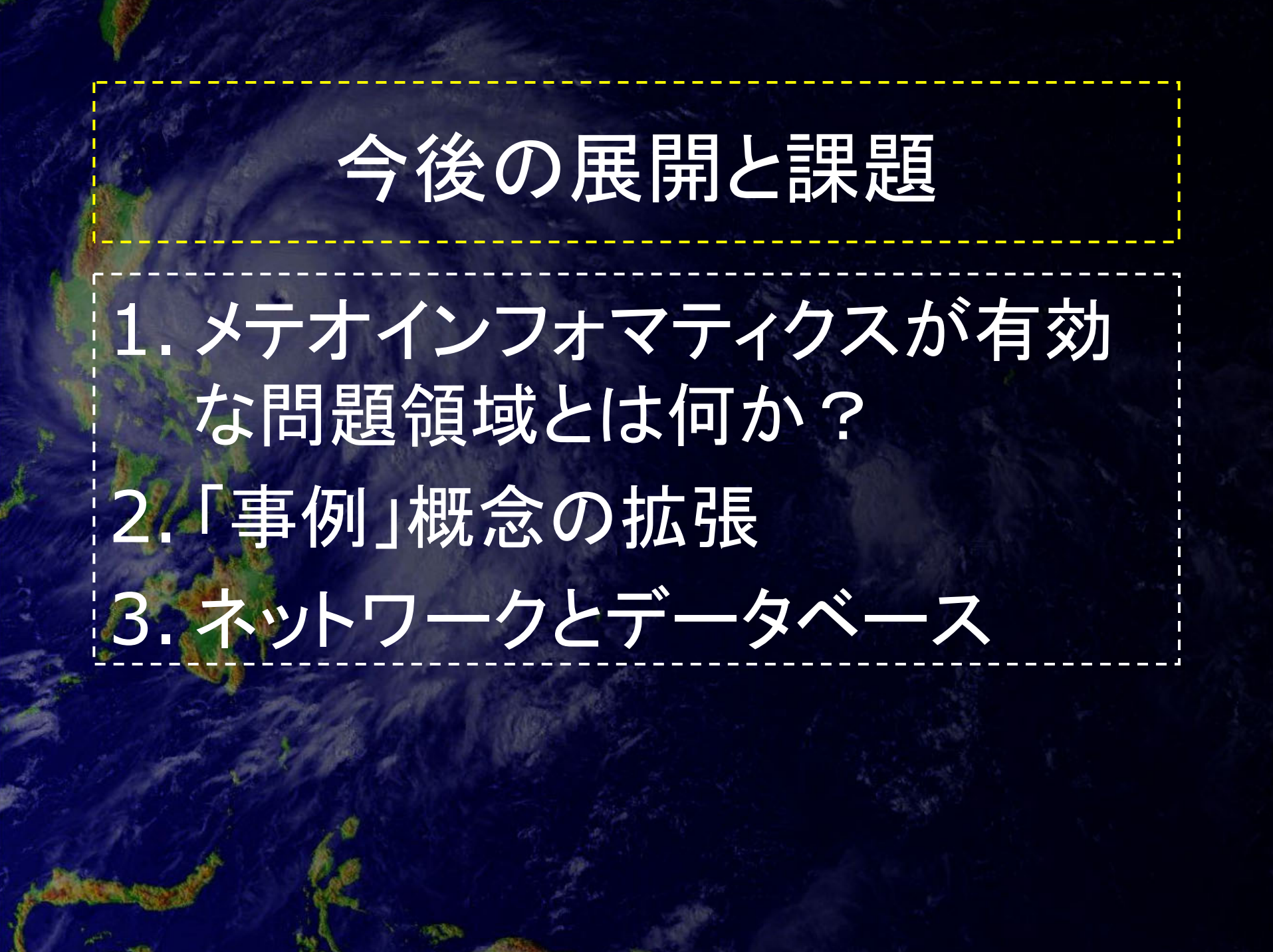


今後の計画

1. 集中豪雨の前後の現況を、画像時系列を用いていくつかのパターンに分類する.
2. 各分類クラスごとに、その特徴を学習させる.
3. その学習結果を用いて、実際のパターンでうまく動作するかを見る.
4. なぜうまくいくのかを、できるだけ可視化しながら検討する.

発表の概要

1. 「メテオインフォマティクス」とは
2. 「デジタル台風プロジェクト」の紹介
3. メテオインフォマティクスに関連する研究の紹介
4. メソスケール事象との関連
5. 今後の展開と課題



今後の展開と課題

1. メテオインフォマティクスが有効な問題領域とは何か？
2. 「事例」概念の拡張
3. ネットワークとデータベース

メテオインフォマティクスが 有効な問題領域とは何か？

- 大規模データベースと高速データベース検索を基本要素とするメテオインフォマティクスが有効に働く問題領域とは何か？
- 空間的分布(台風解析)やデータの関連(テレコネクション), メソスケール現象(集中豪雨)などが重要な場合には、有効ではないだろうか？

「事例」概念の拡張

- 今後は大量のシミュレーションデータが発生してくる.
- シミュレーション結果は厳密には「事例」ではないが、事例不足という根本的な障害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とを比較できるような、巨大科学データベースが必要になる.

ネットワークとデータベース

- 観測データやシミュレーションデータを交換共有するための、超高速ネットワークとミドルウェア(グリッド).
- 巨大公共データベースとそれを利用する多数の専門家からなる、観測と解析とを分業した科学研究スタイル.
- このような研究スタイルは、気象学においても有効か？

まとめ

- 「メテオインフォマティクス」という言葉の内容とこれまでの関連研究成果についてまとめた.
- これは気象学に対する、情報学側からのアプローチである.
- このような新しい分野では、両分野の専門家が共同して研究し、知見を深める必要がある.

最後に

- 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/>