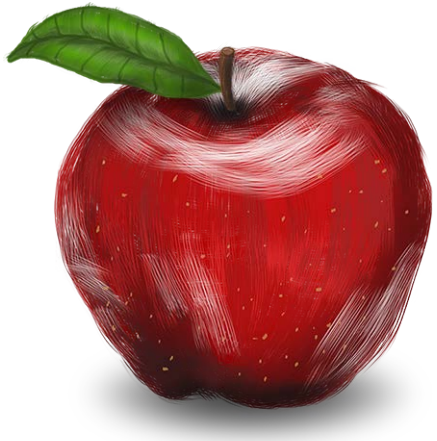
エクサスケールへ向けた多国間二十面体格子モデルプロジェクト

理化学研究所 計算科学研究機構
複合系気候科学研究チーム
吉田龍二

はじめに

2020年頃…エクサスケール計算機が出てくるだろう.

まだまだ計算機のアーキテクチャは不明…



出てきてすぐに使えないと旬を逃してしまう

何事も事前対策が重要です.
でも何から手をつければいいの？

➡ まずは現状把握でしょう！

コンテンツ

- ICOMEXプロジェクトの紹介
  - ✓ Working Programsの紹介
- 日本チーム：物理的なモデル比較
  - ✓ モデル比較実験の概要
  - ✓ 傾圧波実験
  - ✓ 気候値実験
- 計算科学的なモデル比較
  - ✓ 演算性能
- エクサに向けて
  - ✓ 足回り(I/OとPost処理)
  - ✓ その他

ICOMEXプロジェクトの紹介

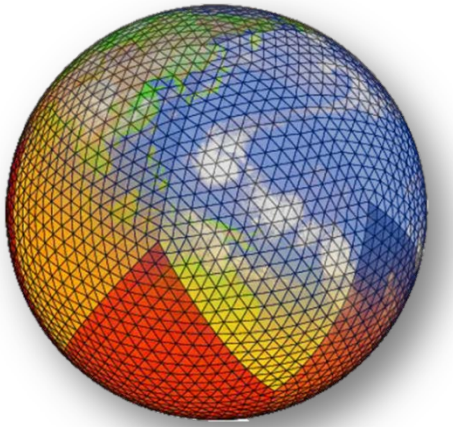
The Icosahedral-grid Models for Exascale Earth System Simulations

- 多国間国際研究協力事業のひとつである.
- エクサスケールマシンで十分な成果が見込めるような気候モデルを開発することを目的とする.
- 超並列マシンで性能を発揮しやすいと考えられる格子法に基づいたモデル(特に二十面体格子)をターゲットにする.
- 1チームでは解決不可能な膨大な問題に協力関係を武器に挑む.
- プロジェクト期間は2011年秋～2014年秋である.

メインメンバー

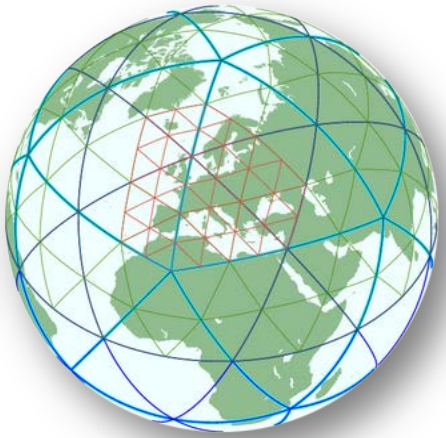
日本チーム（AICS，AORI）　　フランスチーム（IPSL）
ドイツチームA（MPI-M，DWD）　ドイツチームB（Univ. of Hamburg）
イギリスチーム（Univ. of Exeter）

ICOMEXに参加しているモデル



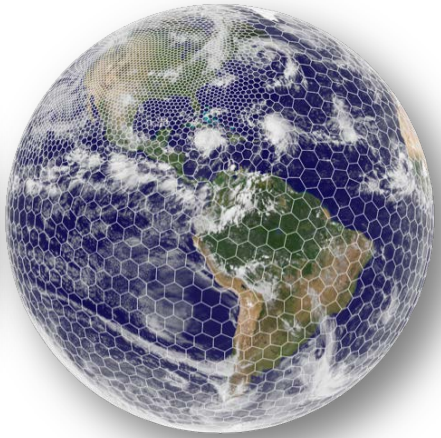
NICAM
日本

主な開発者:
Hirofumi Tomita (AICS),
Masaki Satoh(U.Tokyo)
開発期間：10年以上
格子系:
Structured A-grid



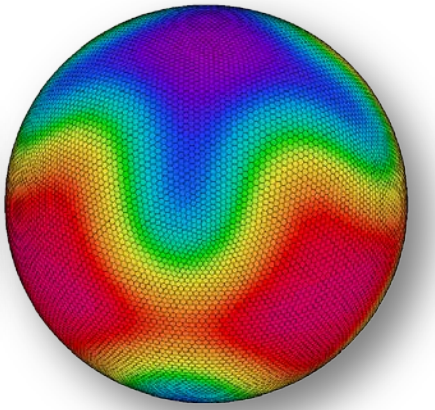
ICON
ドイツ

主な開発者:
Zängl Günther (DWD),
Marco Giorgetta (MPI-M)
開発期間：10年以上
格子系: Unstructured
triangular C-grid



MPAS
アメリカ / イギリス

主な開発者:
William C.
Skamarock(UCAR),
開発期間：3〜4年
格子系: Unstructured
Volonoiメッシュ C-grid
＊ICOMEXへはProf. John
Thuburnを通して参加



DYNAMICO
フランス

主な開発者:
Thomas. Dubos (IPSL)
開発期間：2〜3年
格子系: Structured
hexagonal C-grid
＊静力学モデル

Working Programs

- **現状把握するために**
  → WP1：参加モデル相互比較・評価（日本）
- **計算機への適応を図るために**
  → WP2：データとループストラクチャの最適化（ドイツA）
- **膨大なデータに対応するために**
  → WP3：並列I/Oと並列ポストプロセス（ドイツB）
- **新しいアーキテクチャに対応するために**
  → WP4：GPU（加速器）の可能性調査（フランス）
- **様々な可能性を探るために**
  → WP5：並列計算機における水平陰解法（イギリス）
- **計算機の情報を知るために**
  → WP6：ハードウェアベンダーとのコラボ（ドイツA）

日本チームについての紹介とWP1の結果概要について

気象・気候学的な性能調査

日本チームとWP1の紹介

メンバー

佐藤　正樹　(東京大学/AORI)　富田　浩文　(RIKEN/AICS)

吉田　龍二　(RIKEN/AICS)　八代　尚　(RIKEN/AICS)

その他にAICS, JAMSTEC, AORIに協力者がいます.

目的

Exaに向けた改善のために現状の問題点を洗い出す

- 気象・気候学的な観点からの問題点
- 計算機科学的な観点からの問題点

➡ **重要な基礎情報**

取り扱う実験の種類

1. 傾圧波実験（Jablonowski and Williamson 2006）　2. 気候値実験（Held and Suarez 1994）
3. 水惑星実験（Hayashi and Sumi 1986; Neale and Hoskins 2001）　4. CMIP5 AMIPラン
5. 演算性能の比較

水平格子間隔の定義

各モデルのスカラーグリッド (セル)の数に基づいて
水平格子間隔を定義するとして統一した.

$$dx = \sqrt{\frac{surface\ area}{num\ of\ cells}}$$

ICON	num cells	[km]
R2B4	20480	158
R2B5	81920	79
R2B6	327680	39
R2B7	1310720	20

DYNAMICO	num cells	[km]
GD40	15212	183
GD80	62412	90
GD160	252812	45
GD320	1017612	23

NICAM	num cells	[km]
GL05	10242	223
GL06	40962	112
GL07	163842	56
GL08	655362	28

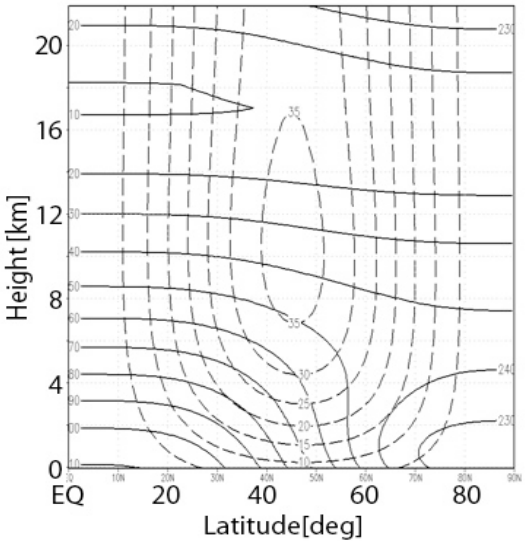
MPAS	num cells	[km]
x1.10242	10242	223
x1.40962	40962	112
x1.163842	163842	56
x1.655362	655362	28

＊Surface area of the Earth = 510,072,000 [㎢]

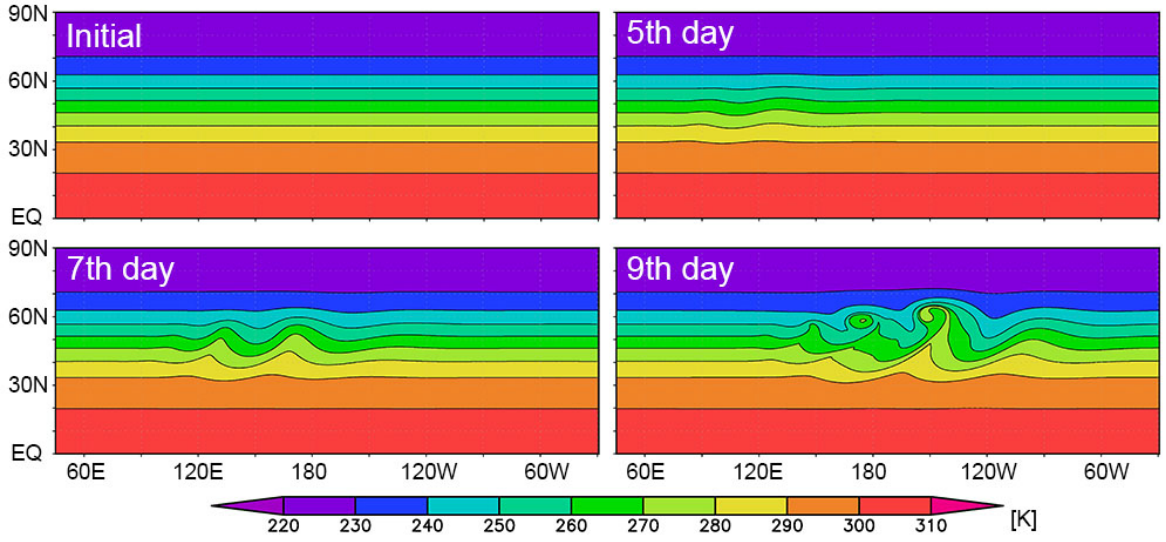
傾圧波実験 (Jablonowski and Williamson 2006)

目的：モデルの力学コアに対して決定論的な評価を行う．

手法：気候値に近い温度分布，帯状風分布を与えた傾圧場を用意し，初期値に小さな擾乱を与えることで，擾乱が傾圧波として発達して行く様子をシミュレーションする．
水平格子間隔，鉛直層の取り方を変えて結果を比較する．

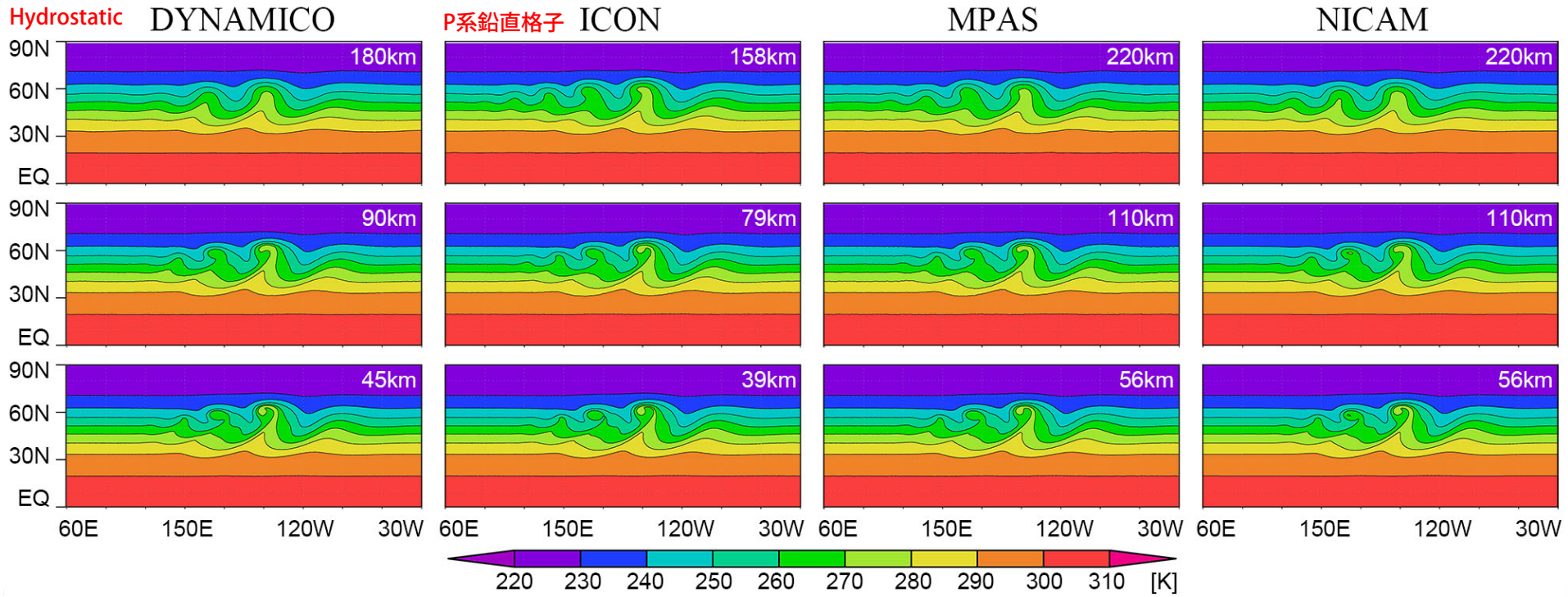


初期値のプロファイル
解析的に与えられている．



NICAM GL09(dx=14km)の計算結果：高度850hPaにおける温度

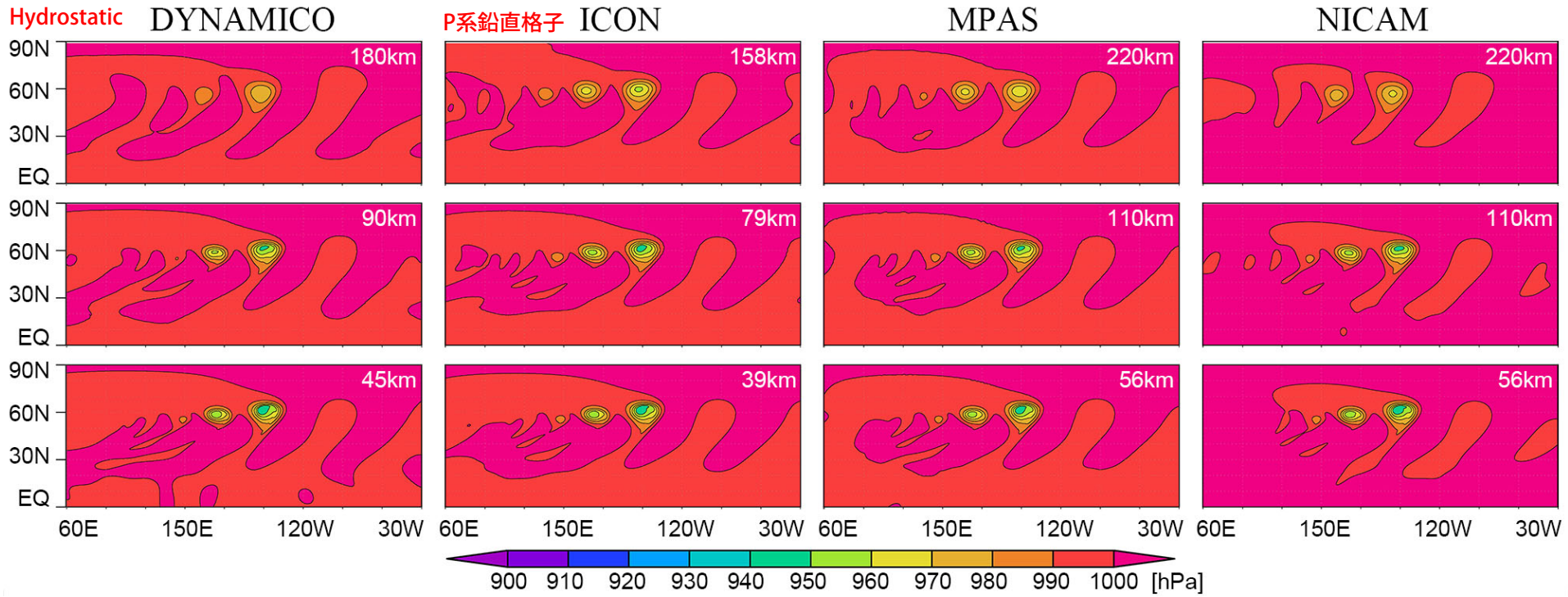
傾圧波実験：9日目（t850）



鉛直40層の結果，850hPa高度の温度分布：どのモデルも水平格子間隔を狭めるとより細かい構造まで表現するようになることがわかる.

＊MPASとNICAMは下層ほど細かくしたストレッチ鉛直層を使用している.

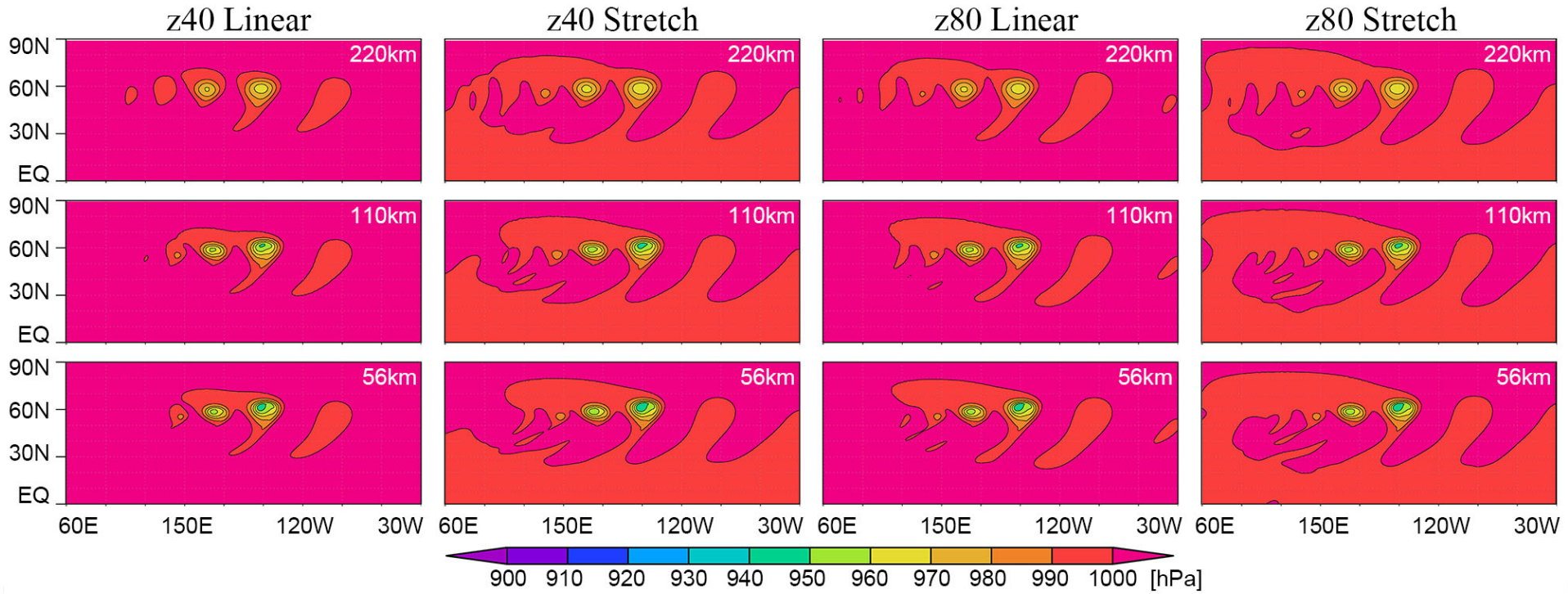
傾圧波実験：9日目（PS）

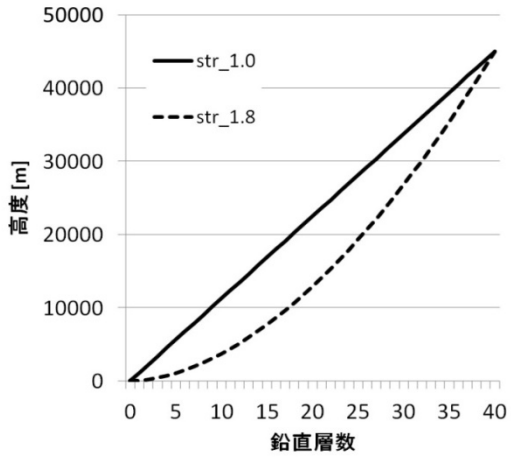


鉛直40層の結果，地上気圧分布：この図からもそれぞれのモデルにおいて水平格子間隔を狭めたことによる再現結果の向上が見られる.

＊MPASとNICAMは下層ほど細かくしたストレッチ鉛直層を使用している.

傾圧波実験：鉛直層の変化



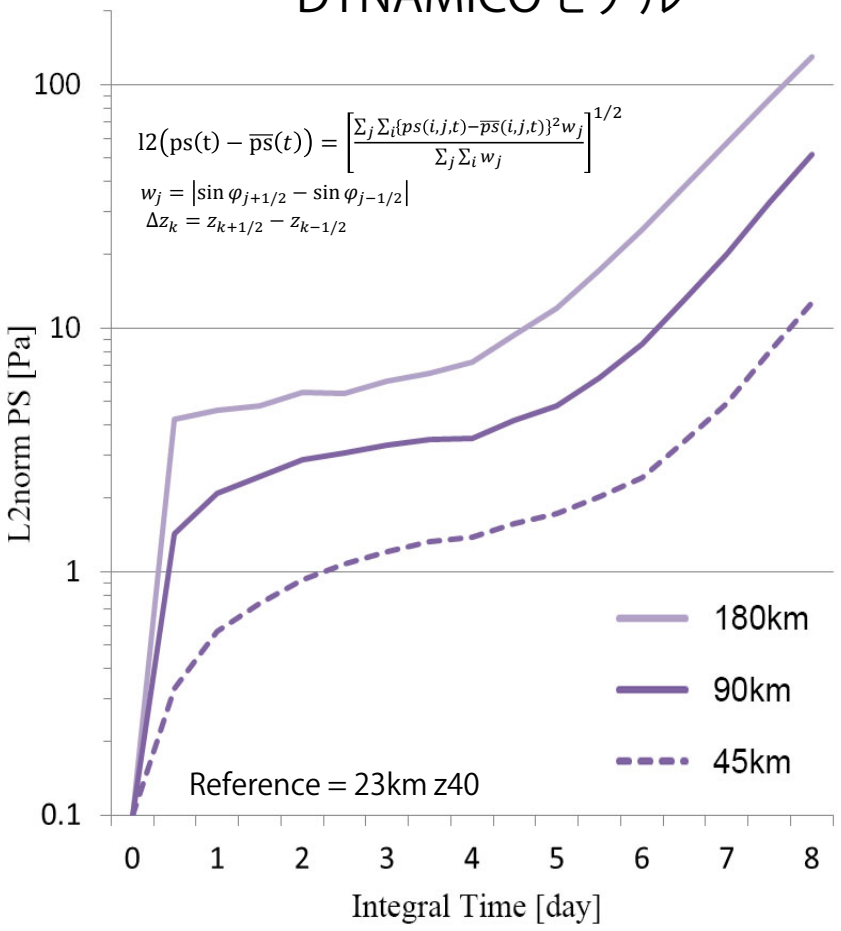


MPASの場合に，鉛直層の取り方を変えると結果が大きく変化する．
NICAMでも同様の結果が見られたが，ICONではわずかな差しか認められなかった．

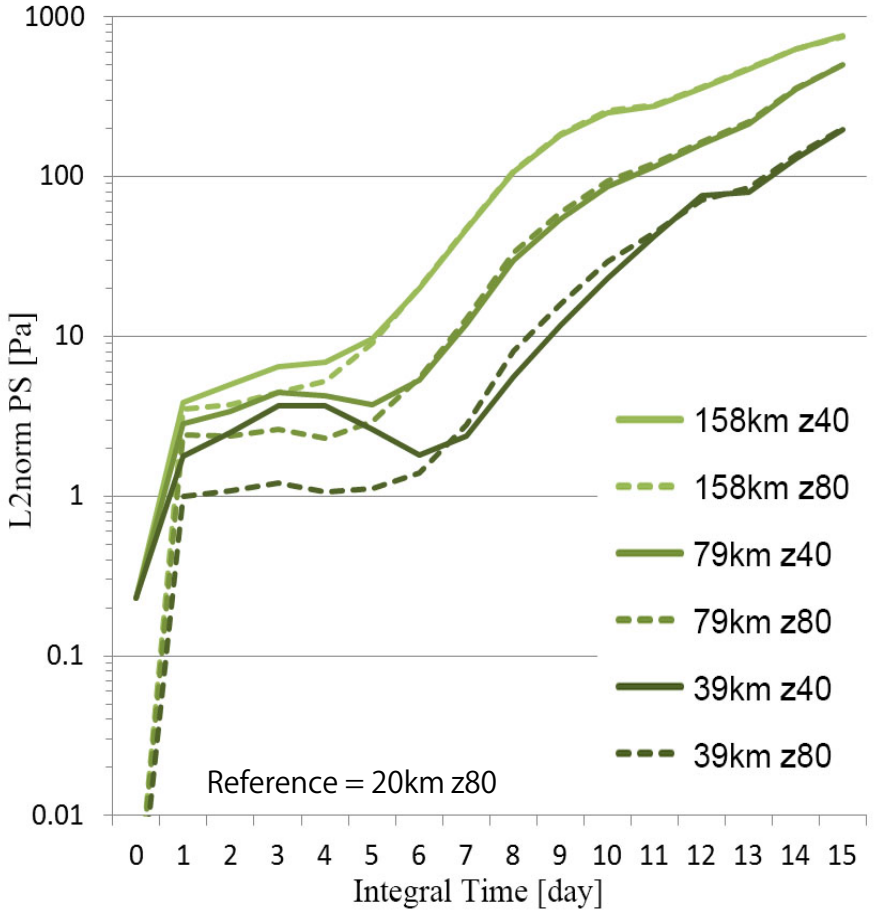
Linear: z(1) = 1125m　Stretch: z(1) = 59m

傾圧波実験：L2norm (dynamico and icon)

DYNAMICOモデル

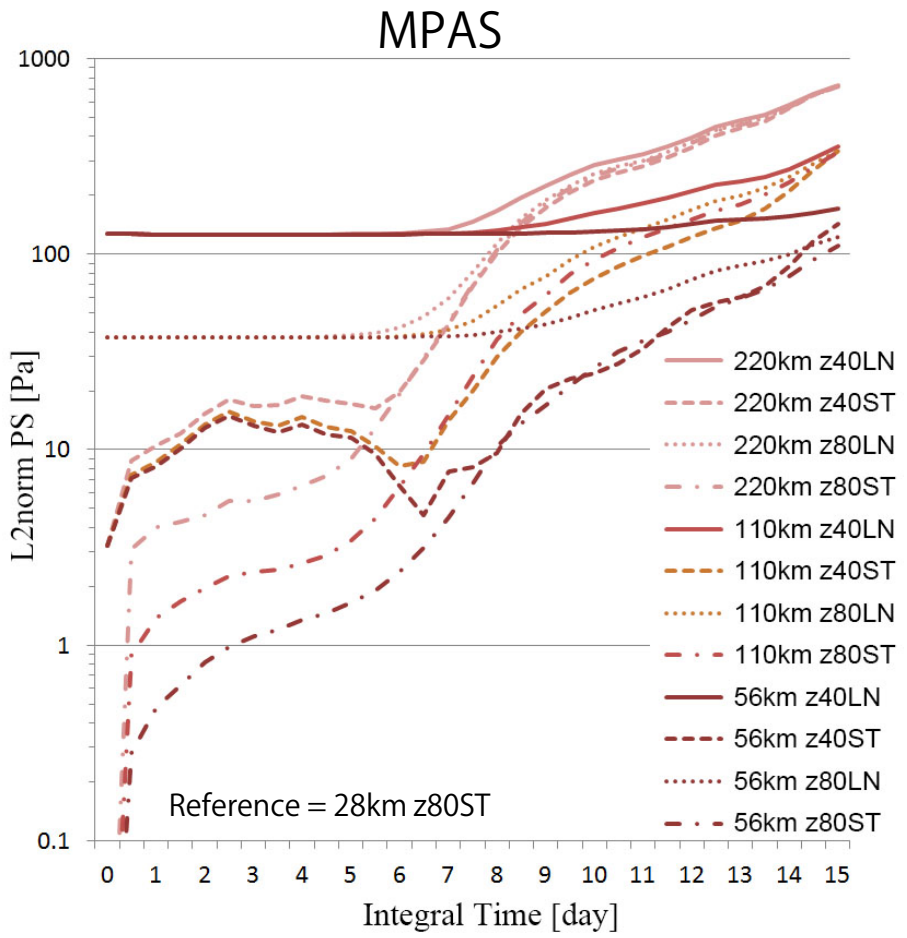


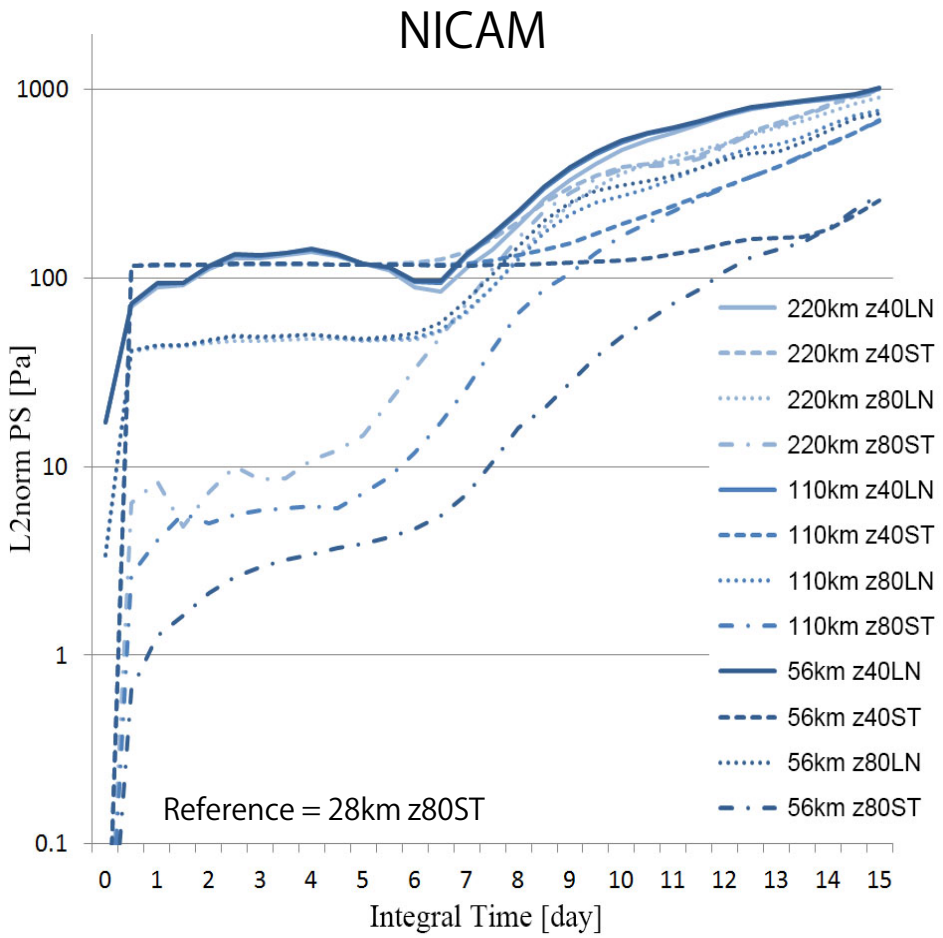
ICONモデル



- 各モデルの最高解像度の結果をReferenceにしてL2normを算出した.
- どちらのモデルも水平格子間隔の向上に従ってエラーの大きさが小さくなっていることがわかる．ICONは鉛直層数による変化は小さい.

傾圧波実験：L2norm (mpas and nicam)





MPAS，NICAMともに鉛直層の取り方による影響が大きく，ストレッチ80層の場合にのみ水平格子間隔の向上に伴うエラーの改善が良く現れている．

傾圧波実験のまとめ

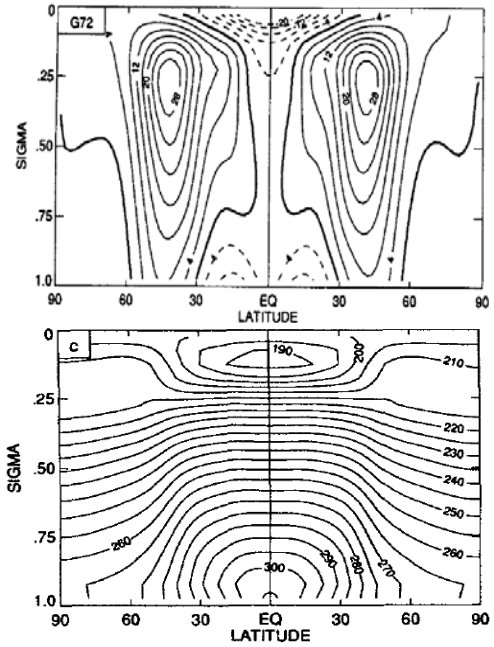
- どのモデルも基本的な傾圧波の構造の再現性に問題はない.

- 離散化手法によっては鉛直解像度がエラーの大部分を占めることがあり，鉛直層の取り方が重要になる.

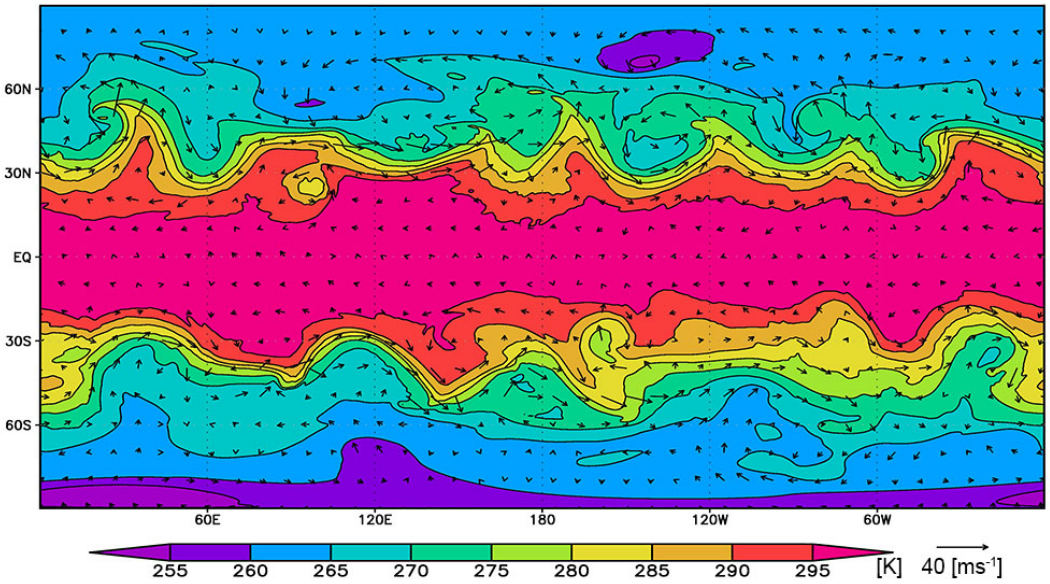
気候値実験 (Held and Suarez 1994)

目的：モデルが再現する気候値の統計的な特徴を調べる.

手法： 複雑な物理スキームを導入する代わりにある時定数で
規定プロファイルに引っ張る人為的強制を加えて実験を行う.
全部で1300日の積分を行い，後半1000日を対象に解析をする.
＊この実験はNICAMとICONモデルだけに限定して行った.

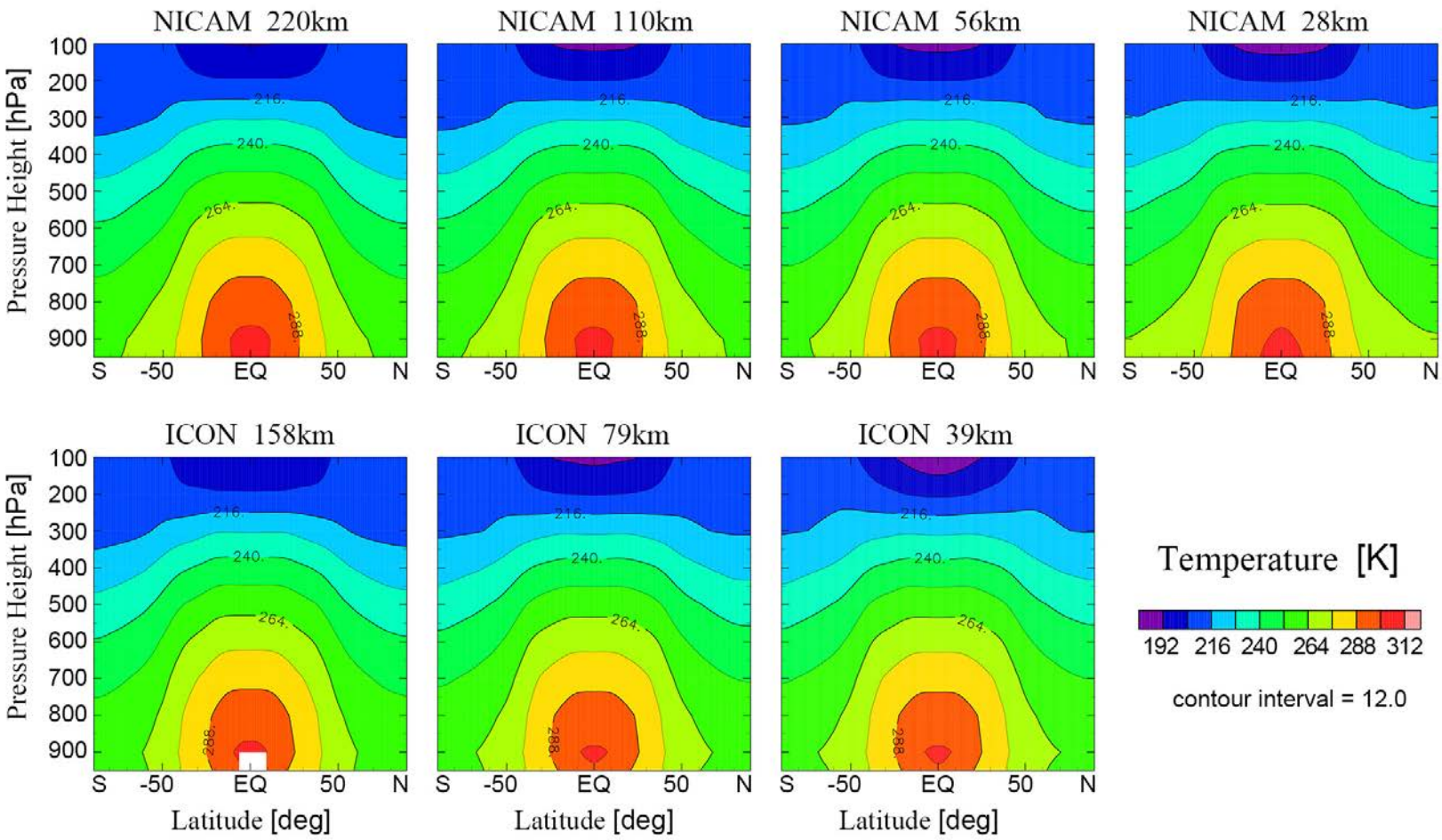


Held and Suarez (1994) 原論文で
示されている再現された気候値



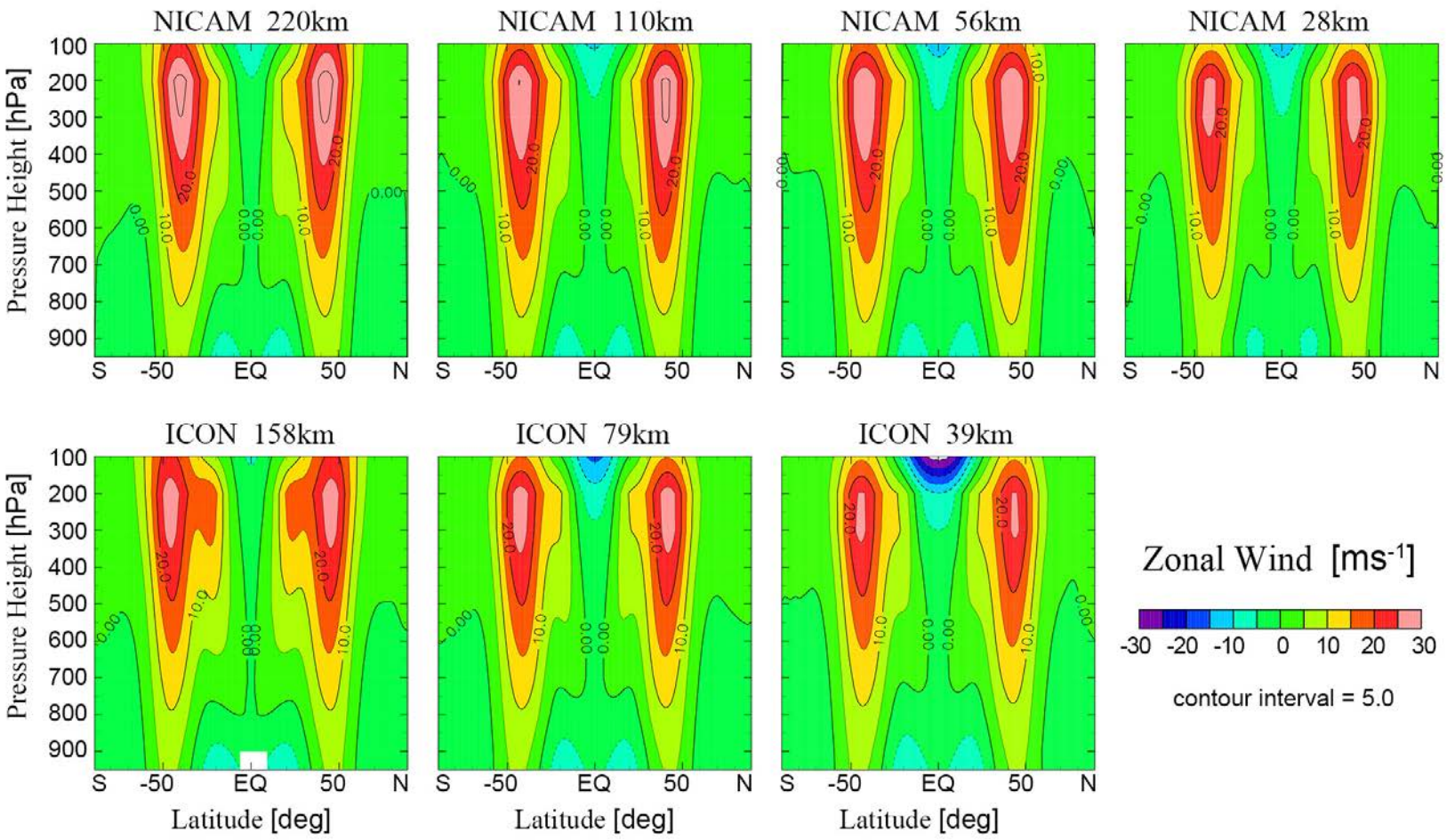
GL08(28km)の400日目のスナップショット：
シェードは850hPa 高度の温度，ベクトルは850hPa高度の風

気候値実験：温度(帯状平均)



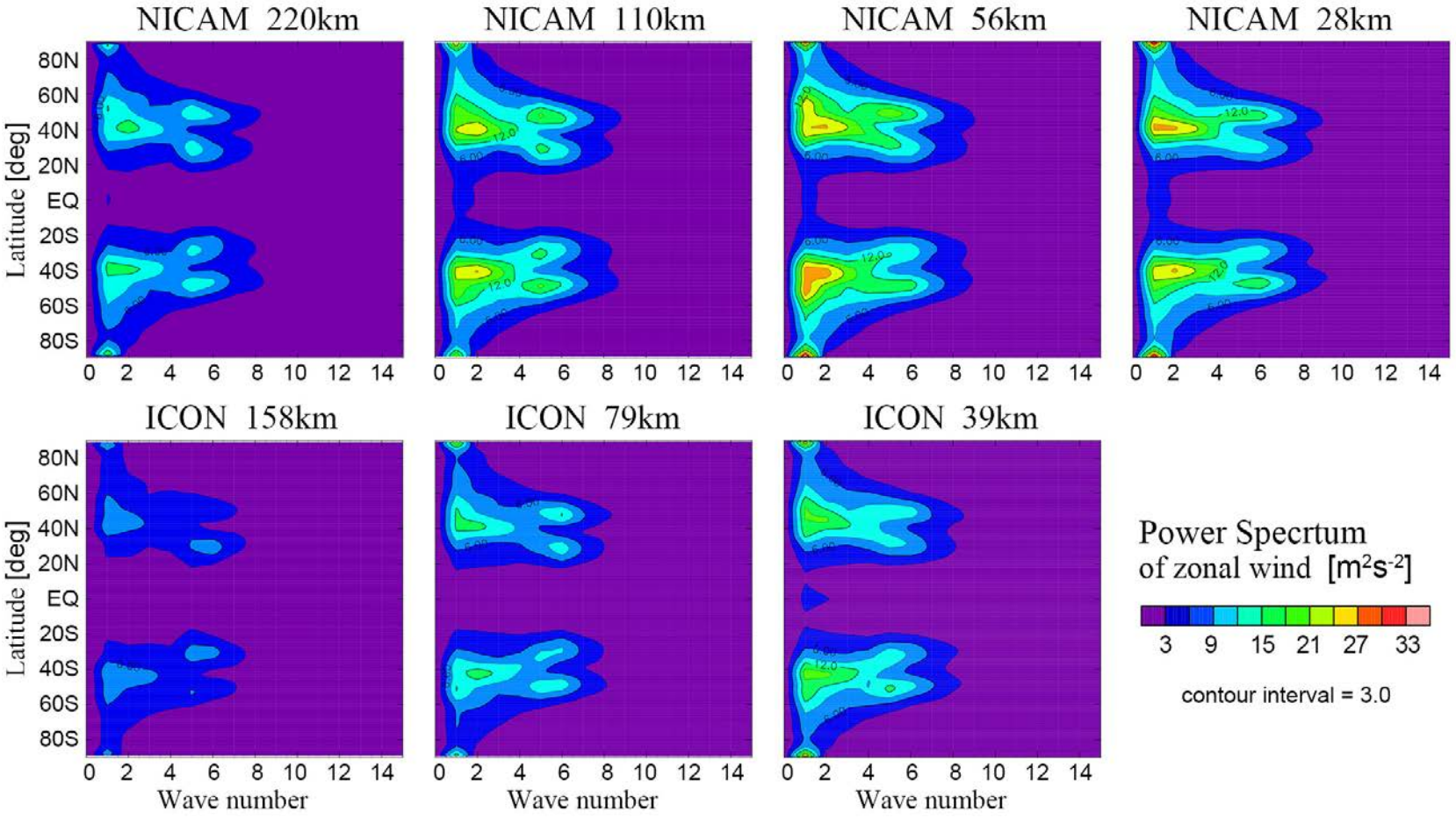
どちらのモデルも水平格子間隔が狭くなると熱帯域対流圏上層の低温域が顕著なる傾向が見られる.

気候値実験：帯状風(帯状平均)



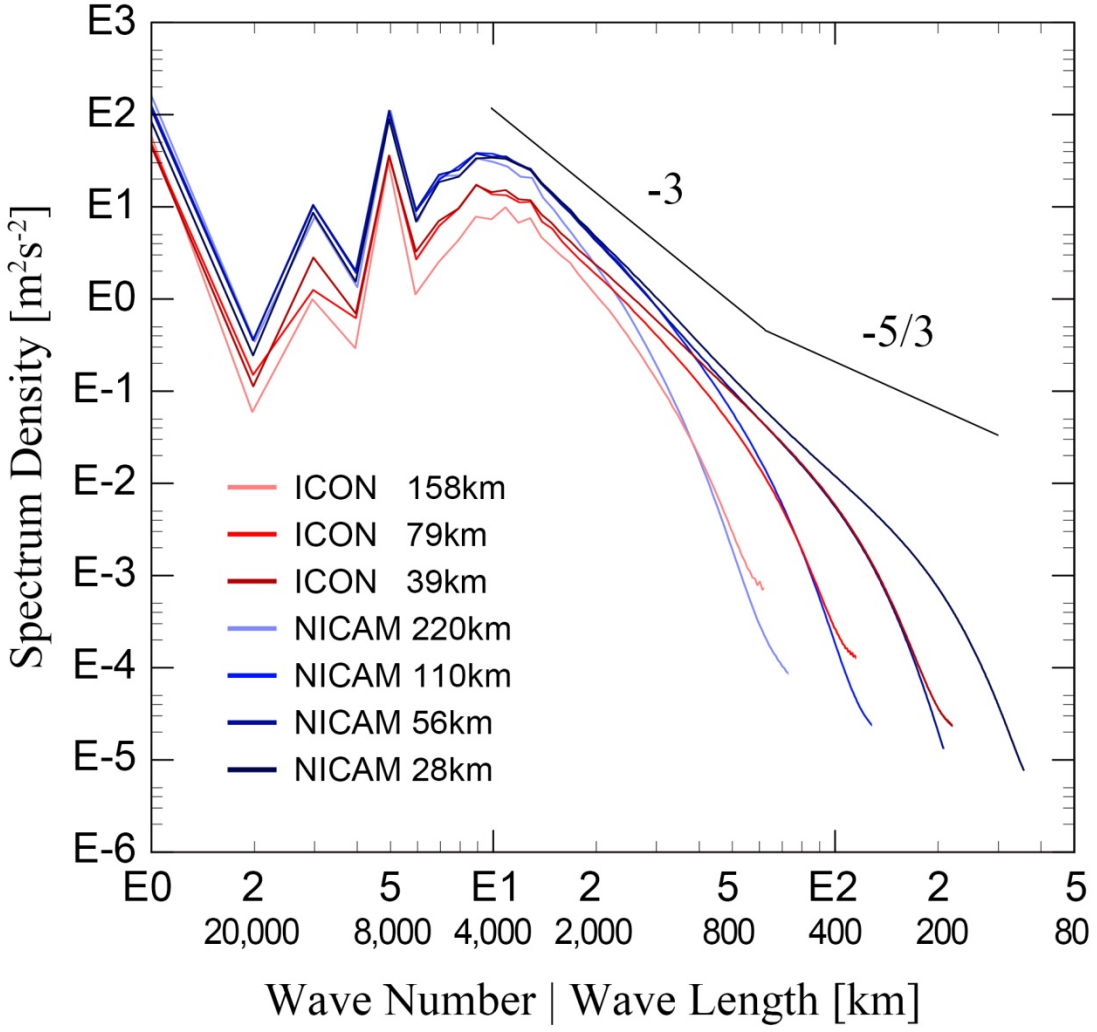
水平格子間隔が狭くなると西風ジェットが弱まる傾向が見られる.
2つのうちNICAMの方が顕著に見られ，ICONでは変化量が小さい.

気候値実験：緯度-波数分布



中緯度帯に波数5～7の大きな振幅が見られ，原論文と整合的な結果を得ている．
水平格子間隔が狭くなるほど振幅が大きくなる傾向が見られる．
ICONはNICAMに比べて振幅が小さい．

気候値実験：運動量スペクトル



先と同様に波数5付近に大きなピークが見られる．これは両モデルに共通である．波数5に特にピークが出やすいかどうかをチェックする必要がある．

先の定義に従った水平格子間隔が異なっていても，二十面体の分割数が同じだと似たスペクトル分布を得ている．

偶然？

数値拡散による解像度のdegradeがどの程度効いて調べる必要がある．

気候値実験のまとめ

- 原論文と比較して，NICAMもICONも妥当な気候値を示す.
- 格子間隔が狭いからといって，必ずしも高解像度だという訳ではない.
- 安定に長期積分出来る数値拡散の大きさと，解像度を評価できるような指標を考えてみたい.

ところで…

力学コアだけとはいえ，比較的低解像度の実験とはいえ，長期積分になると結構つらくなってくる.

だから少しでも演算性能を高くしたい.

計算科学的な性能調査

着目する評価変数

- **FLOPS：演算器がどれだけ働いたか？**
  浮動小数点を単位時間あたりに処理できた回数
- **Memory Throughput：作業対象がスムーズに運ばれたか？**
  メモリの通信帯域の実効性能
- **SIMD：演算器がどれだけ効率よく働いたか？**
  ある単一の命令を複数のデータに適用して処理速度を向上させる技術
- **並列化効率：演算器を増やした効果を示す指標**

京コンピュータにおける性能

京コンピュータの性能プロファイラで測定した結果　＊1プロセス，1スレッドにおける値を示す．

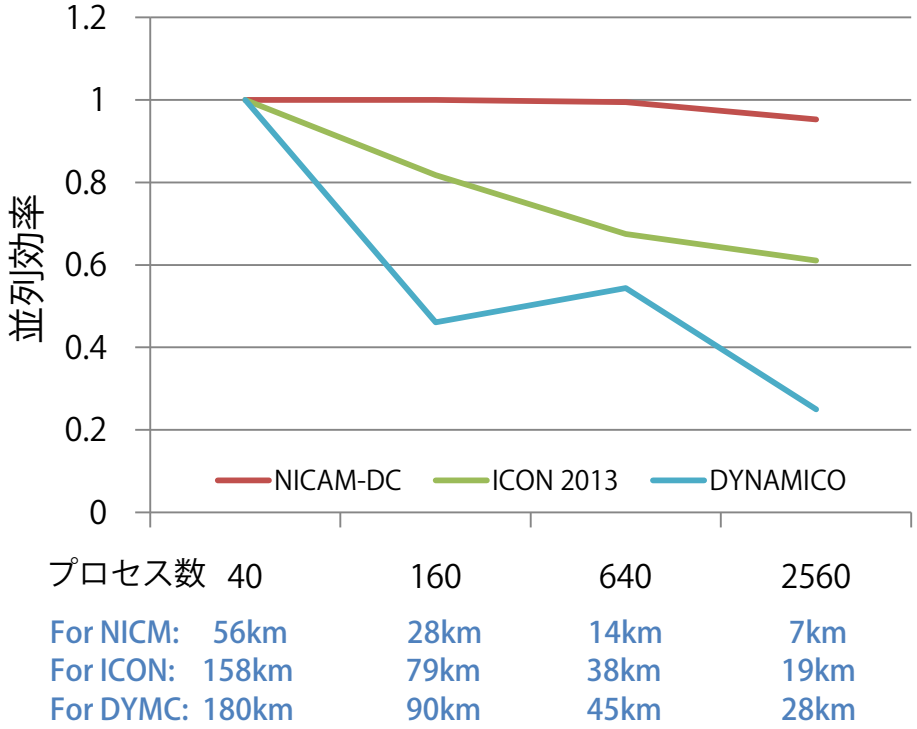
Model	プロセス数	MFLOPS	MFLOPS /Peak(%)	Mem-TP(GB/S)	Mem-TP/Pk(%)	SIMD(%)
ICON	48	229	1.4	2.3	3.5	5.3
DYNAMICO	10	535	3.3	3.4	5.3	2.2
NICAM	40	7,625	6.0	34	53.7	63.8

＊NICAMは京向けにチューニング済みだが，ICONとDYNAMICOはAS-ISコードを利用している．

- NICAMに比べて，ICONとDYNAMICOは計算機の性能を活かし切れていないことがわかる．
- 十分な性能を得るためにはチューニングが必要である．

並列性能比較

1プロセス当たりの演算量を
一定にした場合のMPI並列の性能



Hybrid並列におけるスレッドごとの
実行時間のバラツキ

Model	MFLOPS	SIMD
ICON	4.7	1.3
DYANAMI	4.6	3.8
NICAM	1.1	1.2

NICAMは，プロセス数が増えてもMPI並列の性能を保っている.
プロセス内におけるスレッドごとの実行時間や性能のバラツキも小さい.

もちろんNICAMも元からこのように性能が高かった訳ではない.

AS-ISコードの問題点

NICAMのチューニング時に問題になった点の例

- 変数を宣言したあとのゼロやUNDEFの代入
- 大量な一次変数(特に配列)の使用
  → 絶対に必要で無い場合はやめる
- 文関数の使用
  → 普通の関数にする
- 最内Loop内のif文
  → ifを外側へくくり出すことを考える
- 配列式の使用（京では今は基本的に大丈夫）
  → DO Loopで記述する

I/O性能・手法の重要性

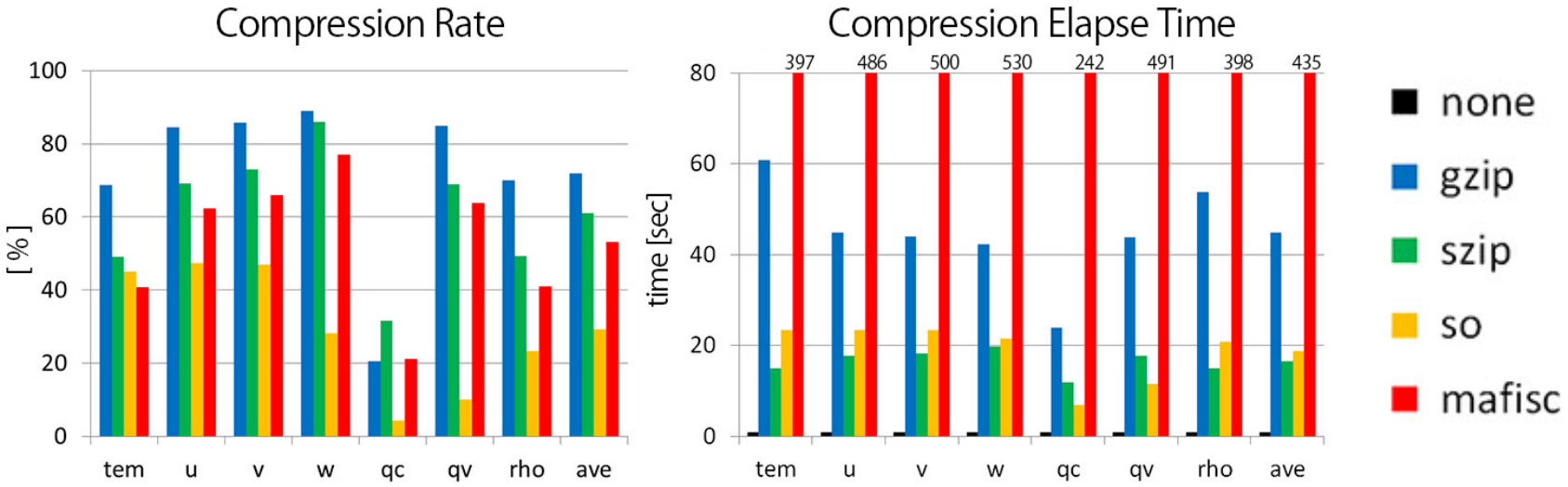
- エクサスケールの計算では出力データサイズがペタクラス？
- いまでも1回のRUNで数十TB，実験1セットで数百TBのデータを吐く．

いまよりモデルの足回りが実行時の重要課題になる．

- NICAM：とにかくプロセス毎にファイルを吐く(単純バイナリ)
- MPAS：MPI-IOで並列I/O（netcdf）
- ICON / DYNAMICO：MPIでGatherして1ファイルに吐く（netcdf）

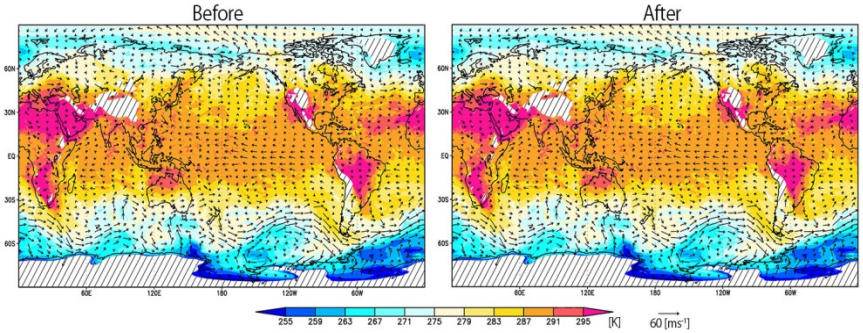
他に吐き出す量を小さくする方法（ファイル圧縮）が考えられる

ファイル圧縮の可能性



- ICOMEXのWP4において並列I/Oや圧縮手法の開発が行われており，そのうちの一つにHDF5用のフィルター”MAFISC”がある.
- HDF5のプロパティとして直接呼び出せるプラグイン型ロスレスフィルターである.
- NICAMの現実場再現データを用いてMAFISCの性能評価を行った.

- 圧縮性能は比較したgzipやszipよりも良い.
- 圧縮にかかる時間が大きいことが問題だが今後チューニングが行われれば有用なフィルターである.



計算実行の容易さ

エクサスケール計算機は実行方法にいろんな制約があるかもしれない？
吐き出すファイル数や実行時に使えるディスクサイズなど…

- どんな風にログを吐くか
  - 実験後に何を行ったかを確認できる唯一の重要なファイル
  - たとえば読み込んだ鉛直層設定をログに吐く

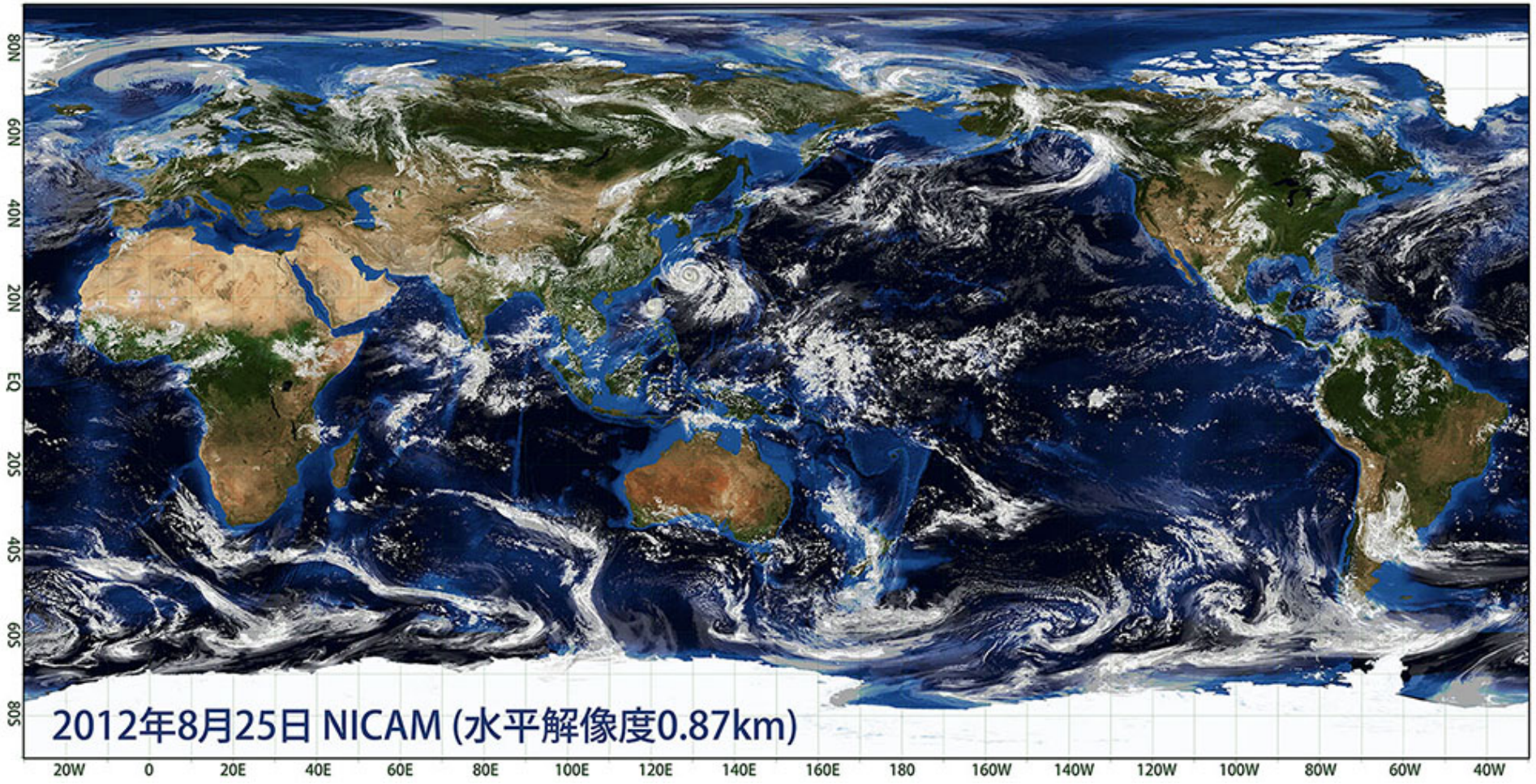
大規模計算は一度ミスると失うリソースも大きい…
万全の体制でJOB投入できる準備が必須になる.

- どんな風にRestartさせるか
  - 実験設定の一部は自己記述的なrestart fileにするなど
  - 前回のnamelistとrestart fileの整合性，前回のバイナリとの整合性などのチェック機能
  - 計算が落ちたら止まる機能（NANの検出など）

- 仕様をコロコロ変えない
  - 仕様変更を後から追いかける作業が大規模 or 長期ランの実行には本当に骨になる

大規模ランの例

NICAM (870m 96層) による現実場再現実験



＊海洋研究開発機構・東京大学大気海洋研究所（HPCI戦 略プログラム分野3） および理化学研究所計算科学研究機構の共同研究

大規模ランの後処理

870m格子（30分間隔，1日分，48コマ）の処理の例
元データサイズは180TB程度

- 二十面体格子から描画用のLatLon格子へ変換（1〜2ヶ月）
- LatLon格子からレンダリング用のデータ変換（1週間）
- 1回のレンダリング（30秒〜1分）
- 全コマのレンダリング（1〜2日）
- オーサリング（数時間）

- こんな状況なので出来るだけ解析は二十面体格子のままで，京をつかって解析をする.
- 解析プログラムの高並列化，並列I/Oの対応なども大変に重要な課題である.
- 計算と後処理の並列化もICOMEXのミーティングでもよく言われるアイデアになっている.

まとめ

- 自由に実験するためには，まず現行マシン上で効率よく走るようにチューニングが必須
- 大規模計算を行うにはI/Oの対策は必須
- 計算実行時の負担やミスを少なくするような足回りの準備が必須
- 出てくるデータサイズに対応した並列後処理システムが必須

ご静聴ありがとうございます