

## 第4章 開発における検証

### 4.1 開発における実験と検証<sup>1</sup>

#### 4.1.1 はじめに

検証は数値予報モデルの開発にとってきわめて重要である。開発では、実験の結果を検証値と比較して予測精度の変化を確認し、加えた変更内容と突き合わせて予測精度の変化をもたらした要因とメカニズムを分析し、その結果から次の方針を決めていく。検証結果は、数値予報を利用する観点だけではなく、数値予報モデルをどのように改良していくかを判断するためにも必要不可欠な情報と言える。本章では、GSMの開発で行われている検証について、その目的と方法を簡単に説明したあと、検証結果からみえる課題について、特に系統的な誤差に着目しつつ概観する。

目的ごとに適した実験の仕様と検証手法は異なるため、開発者は複数種類の実験を駆使して開発を進めている。例えば、大気現象の時空間スケールは多様であり、検証対象とする現象によって適した実験と検証手法は異なる。GSMが対象とするスケールは数10 kmからプラネタリースケール、数時間から年単位にわたる。現象の発生頻度の点でも、毎日起こっているものから数十年に一度の現象まで様々である。これらを、同一の実験により取り扱うのは困難である。

一方で、開発効率の観点からは実験の仕様の乱立は望ましくない。そのため、GSM開発者が共通して行う実験と検証は目的ごとに標準化されており、その仕様が定められている。実験の仕様を決めるにあたっては、開発を効率的に進めるため、開発作業のターンアラウンドタイム(TAT: Turnaround Time)の設定が重要になる。実験の実施と結果の検証に数か月を要することは現実的ではなく、実験と検証に要する計算時間で測ったTATは規模の小さい実験なら半日、最大でも2週間程度が望ましい。また、開発時からなるべく現業運用に近い状態で予測特性の変化を確認することが効率的であるため、現業運用と実験仕様の間なるべく差異が存在しないように考慮する必要がある。

#### 4.1.2 開発者が共通して行う実験

GSMの開発基盤として、以下の実験が整備されている。

- ・ 性能評価試験
- ・ 低解像度予測実験
- ・ 再予報型実験
- ・ 1年積分共通評価実験
- ・ 長期積分型実験

ただし、ここではSCM(Single Column Model)をはじめとする理想実験環境は除いている。これらの実験はそれぞれ対象とする検証内容が異なっており、全体

でなるべく多くの点を検証できるように相互補完的に設計されている。

本章では、第4.2節で性能評価試験における検証を、第4.3節で水平解像度を下げたモデルについて再予報(Hamill et al. 2006; 高谷 2012)を用いる検証を、第4.4節では1年積分共通評価実験や長期積分型実験による検証を対象とする。本節では、まずそれぞれの実験の役割分担を簡単に説明する。個別の詳しい仕様や狙いは各節をご覧いただきたい。

性能評価試験は、解析と予測から構成される実験であり、高解像度モデルを用いて11日先までを予測する。その時の最新の解析システムを利用するため、通常は概ね2年以内の夏及び冬を対象にする。予報モデルの変更においても解析から実験を行うが、これは、数日先までの予測においては、解析値の変化が予測精度に与える影響が大きく<sup>2</sup>、同一初期値を用いた実験による精度の比較では不十分であるためである。この実験のTATは概ね一週間である。

上記の性能評価試験の補助的な実験として、予測のみを低解像度のモデルで行うことも多い。この主な目的は、予測特性の解像度依存性を確認することや想定していない予測の差がないかを確認することであるため、比較対象について同一初期値を用いて実験する。また、アンサンブル予測システムを利用して、予測時間を2週先程度まで延長した実験を行い、バイアスの予測時間への依存性などを確認することも行われる。

再予報型実験は、気象庁長期再解析を初期値として過去30年程度を対象に行われる実験である。性能評価試験や、そこで作成された解析値を用いた低解像度予測実験では、その仕組み上実行できる期間に限られてしまうため、年々変動の大きい現象やその影響を強く受ける予測要素について検証を行うことが出来ない。再予報型実験はこの欠点を補う実験である。特に、数値予報モデルの変更内容が、いわゆる気候的な場<sup>3</sup>の影響を受けやすい場合には、複数年に渡る精度検証を行うことが望ましいため、再予報型実験が重要になる。また、この実験では、平年値に対する偏差の予測精度を確認することも可能である。ただし、再予報型実験では、高解像度GSMとモデルの仕様を揃えることを優先しており、海面水温と海水密接度に未来時刻の解析値を与えておらず、その時間発展を気候値の時間変化で代替している。予測時間が進むと海面水温と海水密接度の精度が悪化するため、その影響を強く受ける現象の確認には向いていない。この実験について、ア

<sup>2</sup> GSMを変更すると、解析において第一推定値が変わることや、品質管理をパスして解析で利用される観測データが変わることなどにより解析値が変化する。

<sup>3</sup> 中緯度の高低気圧よりも時空間スケールの大きな場や、エルニーニョ・ラニーニャなどの海洋状態。

<sup>1</sup> 米原 仁

ンサンブルメンバー数と初期日数の選択により TAT が大きく変わるが、開発では概ね 1 週間程度となるように仕様を定めている。

1 年積分共通評価実験と長期積分型実験<sup>4</sup>は、いわゆる境界値問題としての性能やモデル気候値を検証したい場合に実施する。長時間の積分結果でのみ評価できる現象やモデルの問題点も多いため、モデル気候値の検証は全球モデルの基本的な評価として不可欠である。また、これらの実験では海面水温と海水密度の境界値に解析値を与えることが可能であり、その影響を強く受ける様々な現象を、より現実的な観点から評価可能である。1 年積分共通評価実験では、TAT が概ね 1 日となるように、実験対象とする年や使用するモデルの分解能、アンサンブルメンバー数を定めており、高い頻度での実行を可能にしている。また、複数の全球モデル間の相互比較を可能するため、実験構成を必要最小限なものにして、実験を行いやすくしている。長期積分型実験では、積分年数やアンサンブルメンバー数は、必要性や利用可能な資源に応じて幾つかの選択肢から選ばれる。

ここで紹介した実験システムごとに検証環境が整備されている(米原 2017a)。これら検証環境の間では、検証項目が異なるため完全な一体化はできないものの、設計思想を共有しつつ、共通化可能な部品は同じものを用いている。

他にも、業務化試験と呼ばれる実験も存在するが、それらを含めて GSM を用いる各数値予報システムの実験と検証については金浜・山下(2013)、経田ほか(2013)、佐藤(2013)にも解説があるので参考にさせていただきたい。

## 参考文献

- Gates, W. L., J. S. Boyle, C. Covey, C. G. Dease, C. M. Doutriaux, R. S. Drach, M. Fiorino, P. J. Gleckler, J. J. Hnilo, S. M. Marlais, T. J. Phillips, G. L. Potter, B. D. Santer, K. R. Sperber, K. E. Taylor, and D. N. Williams, 1999: An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55.
- Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46.
- 金浜貴史, 山下浩史, 2013: 全球決定論予報検証. 数値予報課報告・別冊第 59 号, 気象庁予報部, 25–33.
- 経田正幸, 山口春季, 檜垣将和, 2013: 週間・台風アンサンブル予報の検証. 数値予報課報告・別冊第 59 号, 気象庁予報部, 34–44.
- 佐藤均, 2013: 1 か月アンサンブル予報システムの検証. 数値予報課報告・別冊第 59 号, 気象庁予報部, 45–50.
- 高谷祐平, 2012: 再予報・ハインドキャスト. 天気, **59(6)**, 493–495.
- 米原仁, 2017: 活用例 (1)–全球モデル. 数値予報課報告・別冊第 63 号, 気象庁予報部, 29–49.

<sup>4</sup> 例えば大気モデルのみであれば Gates et al. (1999) など。

## 4.2 性能評価試験<sup>1</sup>

### 4.2.1 性能評価試験と検証の役割

性能評価試験は、GSMの開発における標準実験と呼べるものである。この実験は、現業仕様に準ずる構成を用いており、改良項目についての総合的な技術的検討を行うとともに、現業運用への導入を最終的に判断する試験（業務化試験）に進むかどうかの検討に用いる（石田 2017）。性能評価試験では、現業仕様の分解能を用いてサイクル解析<sup>2</sup>とその解析値を初期値とする予測を行う。通常、検証の対象期間は合計2か月間で、夏と冬から1か月ずつ選ばれる。また、融雪期の変化を確認する必要がある場合には春が追加されるなど、他の期間が追加されることもある。予測実験は、12UTC初期値のみについて予測時間264時間まで実施する。

この実験では、ある程度の事例数について統計を取ることによって予測精度の詳細な比較が可能である。基礎開発の段階で見られていた変更のインパクトが現業仕様でも同様にみられるか、精度の改善が達成されているか、想定していない問題点が発生していないかなどを総合的に検証する。各過程の改良内容が観測事実や数理的正当性に基づくものであっても、数値予報モデル全体の予測精度を改善するとは限らない。その原因としては、基礎確認が不十分である場合や、compensating errors(堀田・原 2012)が存在する場合などが考えられる。また、解析予報サイクルにおいては数値予報モデルの変更が第一推定値を通じて解析値自体を変えるため、同一の初期値を用いた予測結果の比較では見えなかった違いが現れることがある。このため、性能評価試験による予測精度の確認は不可欠である。

同時に、検証結果を分析して今後の開発に有用な情報を得ることも性能評価試験の重要な目的である。本節では、第4.2.2項に説明する標準検証システムを用いた検証結果を中心に、開発上の課題に着目しながらGSMの誤差特性を解説する。なお、本節で解説する結果は概ねGSM1705についてのものである。

### 4.2.2 標準検証システム

GSMの開発プロセスは科学的方法に基づいており、実験と検証を繰り返しながら進んでいく。その中で性能評価試験は繰り返し大量に実施される。具体例を挙げると、GSM1603の開発では50以上のモデル構成に対して実験が行われている。大量に行う作業を効率化するため、検証を行うツール群（以下、検証システム）は共通ツールとして整備されており、DPSIVS (Deterministic Prediction System Integrated Verification System) と呼ばれている。Deterministic Prediction Systemは決

定論的予測システムの意味で、高解像度全球数値予報システムのアンサンブル予報システムに対比した呼び方である。DPSIVSについては米原 (2017a) に紹介がある。

DPSIVSを利用することで、全ての開発者が共通の手法による検証を行うことになり、開発者間のスムーズな相互理解を可能にしている。特に、近年のGSM開発の重要なテーマであったcompensating errorsの解消には、複数の過程にまたがる開発者間の相互理解が不可欠であったが、その連携推進にDPSIVSは中核的な役割を果たした。相互認識の形成を促進するため、DPSIVSに含まれている20種類程度の検証パッケージを用いて、実験結果について様々な観点から漏れの無い議論を行っている。

### 4.2.3 解析値の確認

予測精度の変化を確認するにあたり、まずは検証値のひとつとして用いる解析値の変化をしっかりと把握することが必要になる。しかし、解析値の精度検証は一般に難しいため、様々な方法でその変化が妥当であるかを確認している。具体的には、データ同化で利用される観測データと解析値や第一推定値の整合性が向上しているか、他の複数の現業数値予報センターの解析値と比較してクラスターから外れる方向に変化していないか、現象の表現が自然なものかなどを確認している。

図4.2.1に、850 hPa面気温の平均解析値について他の現業数値予報センターとの差を示す。図のタイトルにあるG002CNTLはGSM1705についての結果を意味する（以降の図も同様）。南半球の海上において、他3センターと比較してGSMは低温傾向でありやや乖離している。また、ラジオゾンデ観測と比較しても島や沿岸部を中心に低温傾向である（図略）。GSMの予測値は低温バイアス傾向を持つことが観測データとの比較から知られており、データ同化において観測による修正が十分に期待できない南半球の海上において、解析値でもそのバイアスが修正されずに残った結果と考えられる。多くの現業全球モデルにおいて、対流圏中下層の低温バイアスは共通した課題である（宮本ほか 2009）。GSMでは近年の改良により改善が進んでいるものの、依然として他センターよりもバイアスが大きい傾向にあり、今後の改善が期待される。

また、北極海でも同様に低温バイアスがみられる。この領域では下層雲が過剰に表現された結果、反射された短波放射量が気象上端で過剰であることが知られており、海水に入射する短波放射の不足が低温バイアスの一因と考えられる。

### 4.2.4 基本的な予測要素のバイアス分布

GSMの予測変数である気温や比湿、風について、ME (Mean Error) やRMSE (Root Mean Square Error)、

<sup>1</sup> 米原 仁、佐藤 均、下河邊 明

<sup>2</sup> 解析には利用する観測データ量の異なる速報解析とサイクル解析が存在する。サイクル解析は、観測データを可能な限り集めて正確な実況把握のために行うものである。速報解析による試験は業務化試験で行う。

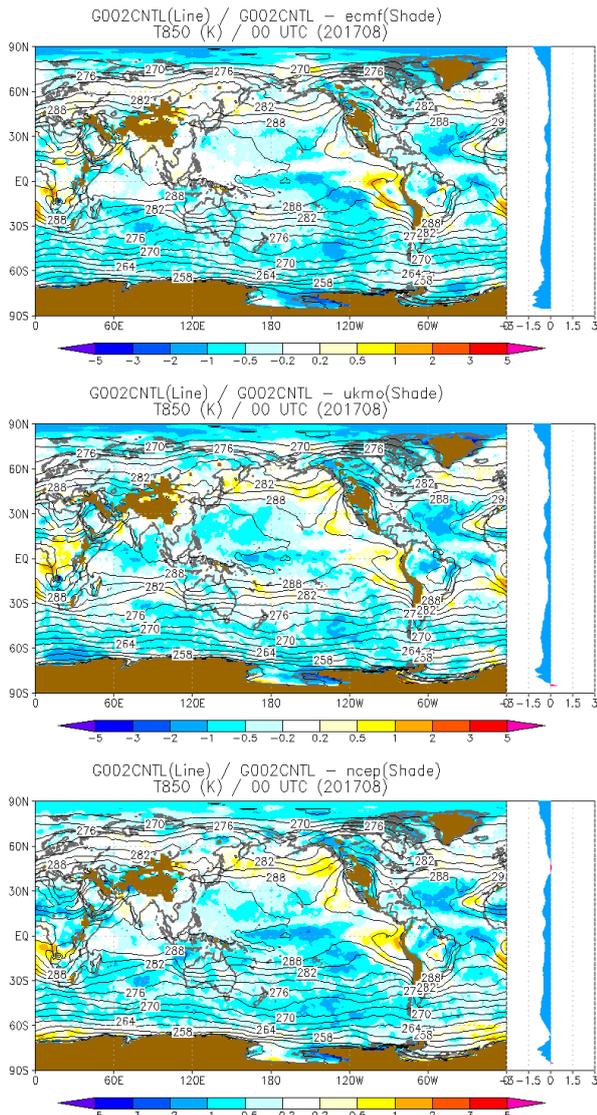


図 4.2.1 2017 年 8 月についての 00UTC における 850 hPa 面気温 [K] の平均解析値について、他の現業数値予報センターとの差を取ったもの（塗りつぶし）。コンターは GSM の平均解析値である。図の右側には東西平均したものを緯度ごとに表示している。図は上から ECMWF, UKMO, NCEP との比較結果である。

ACC (Anomaly Correlation Coefficient) を確認することは検証の基本である。DPSIVS では、検証値として GSM 解析値、他の現業数値予報センターの解析値、ラジオゾンデ観測を用いて、様々な時間や空間、要素について統計検証を行っている。

図 4.2.2 は、ラジオゾンデ観測を検証値にした 700 hPa 面および 925 hPa 面の比湿についての月平均 ME である。700 hPa 面をみると、海洋大陸付近からその北側や北米大陸の東海岸で大きな乾燥バイアスとなっている。このバイアスが大きい領域は対流活動の活発な領域に対応しており、バイアスの大きな領域の季節変化は対流活動の活発な領域の変化に対応して変化する（図略）。一方、925 hPa 面では全球的に湿潤

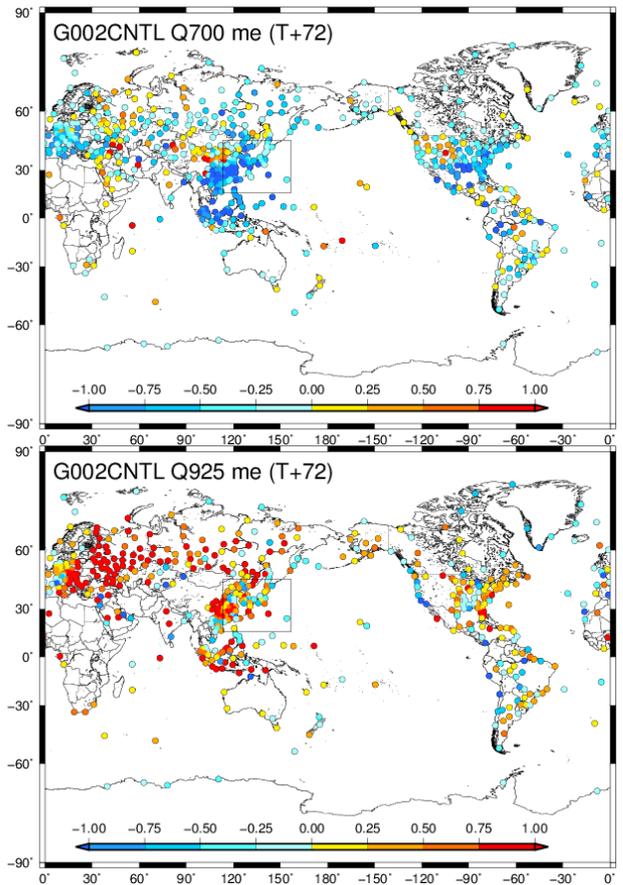


図 4.2.2 ラジオゾンデ観測を検証値にした 700 hPa 面（上図）及び 925 hPa 面（下図）の比湿  $[g\ kg^{-1}]$  についての月平均 ME。検証期間は 2017 年 8 月で、12UTC 初期値の予測時間 FT=72 についてのもの。

バイアスとなっており、特にユーラシア大陸の中緯度で誤差が目立つ。また、海洋大陸付近でも同様に湿潤バイアスとなっている。境界層内で湿潤バイアスかつ対流圏中下層の自由大気で乾燥バイアスを持つ誤差プロファイルは、程度の差はあるが多くの地点と季節においてみられる。これは、境界層内から自由大気に水蒸気を輸送する浅い対流や境界層上端における混合などのプロセスの表現に課題があることを示唆している。

図 4.2.3 に、アメリカ航空宇宙局によるサポートのもと RSS<sup>3</sup> により作成された可降水量 (TPW) プロダクトを検証値とする海上での月平均 ME を示す。海上には可降水量が過少バイアスを持つ領域が広く分布しており、特に海洋大陸付近からその北側で誤差が大きく、ラジオゾンデ観測による検証で見られた結果と整合的である。近年の改良により乾燥バイアスの大きさは半分程度に縮小しているものの（第 3.1 節）、このバイアスの改善は依然として大きな課題である。

<sup>3</sup> リモートセンシングシステム (Remote Sensing Systems) 社。http://www.remss.com/

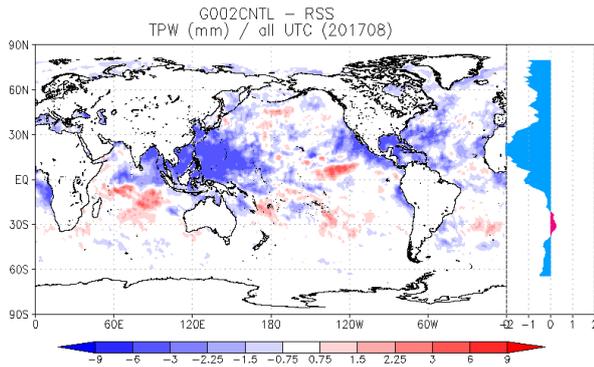


図 4.2.3 RSS による可降水量 [mm] を検証値とする月平均 ME、結果は海上のみである。検証期間は 2017 年 8 月。予測結果は 00, 06, 12, 18UTC 初期値の予測時間 FT=6 についてのもの。図の右側には東西平均したものを緯度ごとに表示している。

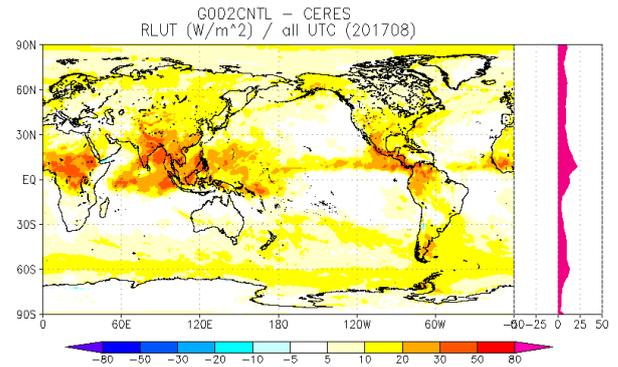


図 4.2.4 CERES プロダクトを検証値とした大気上端での上向き長波放射 [ $W m^{-2}$ ] についての月平均 ME。検証期間は 2017 年 8 月、00, 06, 12, 18UTC 初期値の予測時間 FT=6 についてのもの。図の右側には東西平均したものを緯度ごとに表示している。

#### 4.2.5 放射フラックスの検証

放射過程は大気大循環の駆動源の一つであり、全球モデルにとってその精度は基本的な検証対象である。DPSIVS では、そのバイアスについて面的な分布を中心に精度を確認しており、CERES Level 3B プロダクト<sup>4</sup>を検証値とし、月平均の放射フラックスについて評価を行っている。この検証では、プロダクトに含まれる大気上端及び地表面での長波及び短波放射フラックスについての  $1^{\circ} \times 1^{\circ}$  格子の月平均値を利用している。

図 4.2.4 に、大気上端での上向き長波放射の ME を示す。対流活動が活発な領域を中心に、地球から放出される長波放射が過剰である。図は 8 月についての結果であるが、全球的に過剰である点は季節には依存しない。ただし、誤差のピークの位置は対流活動の活発な領域に対応しており、その季節変化に応じて変化する。このバイアスの原因は、明確に特定されているわけではないが、主に雲氷に関連する複数のパラメタリゼーションに様々な課題があると考えている。

図 4.2.5 に大気上端での短波放射の ME を示す。南極大陸を除く陸上では、季節によらず概ね短波放射の反射が過少である。これは、地表面に対しては短波放射の入射が過剰ということであり、地上気温予測の大きな誤差原因になっている。海上では、多くの全球モデルと同様に<sup>5</sup> 南半球の高緯度で反射が過少になっており、南半球が夏にあたる 1 月では月平均 ME を東西平均した値でピーク値が  $-50 [W m^{-2}]$  を超える。また、海洋大陸付近の過少バイアスや、インド洋と東太平洋の南半球側での過剰バイアスが目立つ。現状、GSM では海面水温や海水量が大气からの影響を受けない仕様であるため、海上における短波放射の誤差が大气に与える影響は抑えられているが、将来関連過程を精緻化

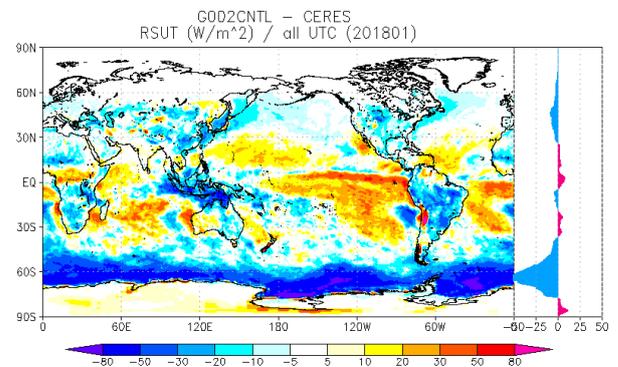
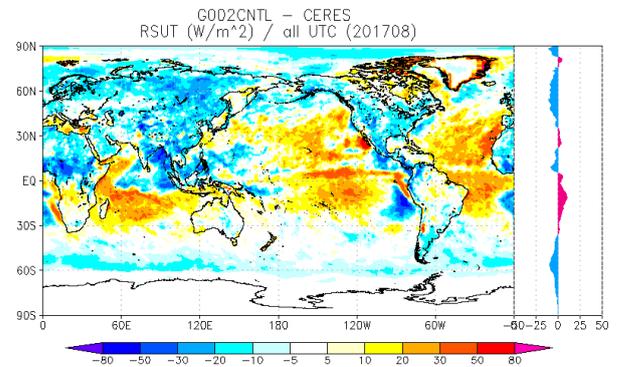


図 4.2.5 CERES プロダクトを検証値とした大気上端での上向き短波放射 [ $W m^{-2}$ ] についての月平均 ME。00, 06, 12, 18UTC 初期値の予測時間 FT=6 についてのもの。検証期間は上図が 2017 年 8 月について、下図が 2018 年 1 月についてのもの。図の右側には東西平均したものを緯度ごとに表示している。

するに当たり障害になる可能性がある。

#### 4.2.6 地表面要素の検証

地上実況気象通報 (SYNOP: Surface Synoptic Observations) を用いて、地上気温や比湿、前 24 時間降水量などの検証を行っている。地上気温や比湿は地表面状態の影響を強く受け、誤差特性は日変化が顕著であるため、各時刻の検証結果を確認している。ただし、

<sup>4</sup> 全球放射収支計によるプロダクト。 <http://ceres.larc.nasa.gov/>

<sup>5</sup> 例えば Trenberth and Fasullo (2010)。

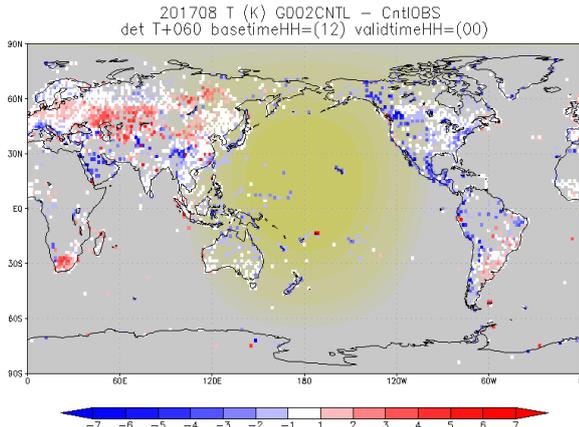


図 4.2.6 SYNOP 観測を検証値とした地上 2 m 気温 [K] の月平均 ME。検証期間は 2017 年 8 月、12UTC 初期値の予測時間 FT=60 についてのもの。図中、黄色で示している領域は大気上端に日射がある領域。

降水量は通報の都合上、前 24 時間降水量についてのみ検証可能となっている<sup>6</sup>。

図 4.2.6 に、SYNOP を検証値とした地上 2 m 気温 [K] の月平均 ME を示す。図に示した陸上における夜間のバイアスは、地域と季節に強く依存している。夜明け前の最も気温が低下する時間帯における、中央アジアに広がる高温バイアスが最も目立つ。図は省略するが、同様の誤差傾向は夏の北米大陸中央でも見られる。一方で、アラビア半島では低温バイアス傾向である。また、日射量が少ない、積雪があると考えられる領域では低温バイアスである。

一方で日中については、地域や季節によって違いはあるものの低温バイアスという共通点を持つ (図略)。図 4.2.7 に示すように、GSM の地上気温予測は日中の最高気温ピークを適切に表現できないことが多い。これは、地表面に対して短波入射が過剰であることから考えられるものとは逆の誤差であり、陸面過程及び境界層過程に課題があることを示唆している。実際、成層が不安定な場合の境界層過程の挙動は近年十分に改善されていない。また、陸面過程も GSM の短波放射の誤差特性に過剰に対応している可能性がある。

#### 4.2.7 海面フラックスの検証

OAFlux<sup>7</sup> (Objectively analyzed air-sea fluxes) プロダクトを用いて、期間平均した全球海面フラックス

<sup>6</sup> 日本域ではアメダスや解析雨量を用いてより時間分解能の高い検証を行っている。しかし、全球的な雨量検証は SYNOP もしくは衛星観測を利用した各種降水プロダクトのみであり、今後より時間分解能の高い観測データの利用が望まれる。

<sup>7</sup> 全球の大気と海洋間のフラックスについての研究開発プロジェクトによる解析プロダクト。NOAA Climate Observations and Monitoring (COM) program による支援のもと、WHOI OAFlux プロジェクトにより提供されている。  
<http://oaf Flux.who i.edu/>

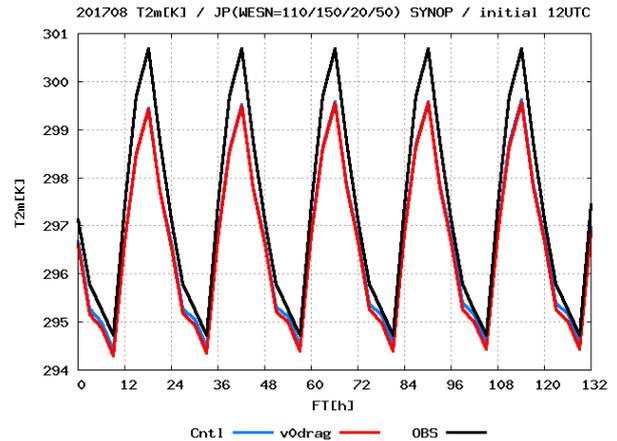


図 4.2.7 縦軸が地上 2 m 気温 [K] の月平均値、横軸が予測時間である。黒線は SYNOP 観測値、青線は GSM1705、赤線はテスト中の GSM の結果 (両者は概ね重なっている)。検証期間は 2017 年 8 月、12UTC 初期値の予測についてのもの。検証対象とした領域は 110°-150°E, 20°-50°N である。

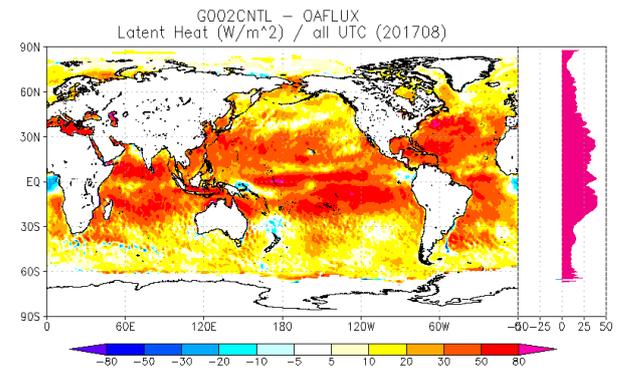


図 4.2.8 OAFlux プロダクトを検証値とした、海上での潜熱フラックス  $[W m^{-2}]$  についての月平均 ME。00, 06, 12, 18UTC 初期値の予測時間 FT=6 についてのもの。検証期間は 2017 年 8 月である。図の右側には東西平均したものを緯度ごとに表示している。

バイアスを検証している。図 4.2.8 は潜熱フラックスの月平均 ME であるが、GSM には海面から大気に潜熱が過剰に流れる大きなバイアスが存在する。他にも、海面では温度は低温バイアス、海面上の比湿は乾燥バイアス、風速は弱風バイアスを持ち、顕熱フラックスも過剰である (図略)。特に潜熱フラックスは誤差が大きく、 $30 [W m^{-2}]$  を超える ME を持つ領域が熱帯を中心に広く分布している。

第 4.2.4 項で、海上において 925hPa でみた GSM の境界層は湿潤バイアスであることを示したが、OAFlux と比較すると海面上の比湿は乾燥バイアス、潜熱フラックスは過剰となっている。これは、境界層内の水蒸気輸送プロファイルに課題があることを示唆している。

これまで示した誤差の傾向について全体のバランスをみると、地表面から潜熱が過剰に大気に供給されているが、可降水量では大きく乾いている。海上からの

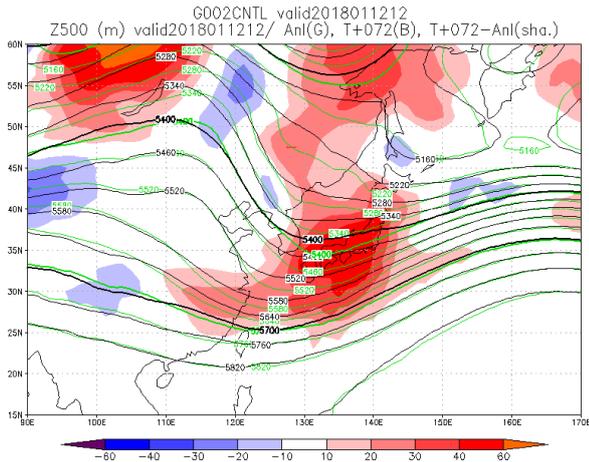


図 4.2.9 2018 年 1 月 9 日 12UTC 初期値の FT=72 の検証図。500 hPa 高度場 [m] についてのもので、黒線で予測値、緑線で解析値、塗りつぶしで誤差を表示している。

顕熱供給も過剰であるが、対流圏中下層は低温バイアスで、大気上端での上向き長波放射は大きく過剰である。おそらく、この誤差構造は放射により大気から過剰に奪われた熱を凝結熱で補償しているために生じていると考えられ、その解消には物理過程全体を通じた改良が必要となるだろう。

#### 4.2.8 日本付近の事例別検証

日本付近については、特に日々の事例ごとに予測精度を確認している。確認で用いている図の例として、図 4.2.9 にある日の 500 hPa 面の高度についての誤差図を示す。

図 4.2.9 に示した事例は、GSM のトラフ予測の典型的な誤差パターンを示している。本州の上空を通過しているトラフについてその前面が浅く予想されているが、この誤差パターンは冬季の日本付近の予測で繰り返し見られるものであり、図 4.2.10 に示すように月平均バイアスでも明瞭にあらわれている。この誤差特性は日々の予報作業に影響を与えているためその改善は長年の課題となっている。これまでの調査では、サブグリッド地形に由来する抵抗プロセスの影響が大きいことが分かっており、地形性重力波過程と境界層過程を変更したインパクト実験では、この誤差が改善される可能性があることが確認されている（第 3.4 節）。

図 4.2.9 と同様に、海面更正気圧や 850 hPa 面の気温、上層風、降水などの要素について、FT=120 程度まで性能評価試験の結果の全事例を確認している。特に、日本付近の領域で統計を取った様々な要素についての日々の予測スコアの変化と事例検証の結果を突き合わせて、変更内容から想定した変化が個別の事例検証でも現れているかを確認するのは重要な作業である。

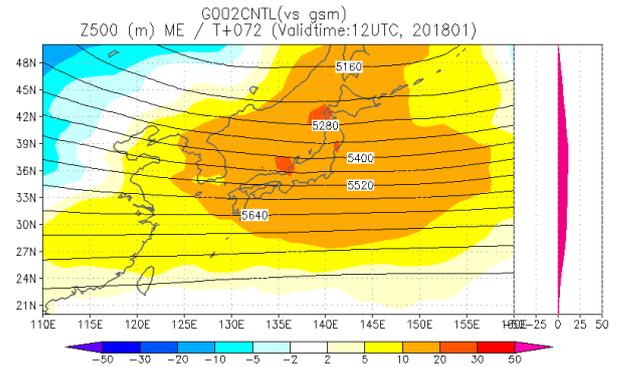


図 4.2.10 2018 年 1 月についての 500 hPa 高度場 [m] の ME (塗りつぶし) で、12UTC 初期値の FT=72 についてのもの。コンターは解析値の平均である。図の右側には東西平均したものを緯度ごとに表示している。

#### 4.2.9 台風構造の検証

DPSIVS は複数の検証パッケージの集合体であり、GSM 開発において新たに課題が認識され、検証の必要性が生じた場合、新しい検証ツールが作成されて追加される。ここでは、開発中の台風構造検証について紹介する。このツールは、GSM1603 には台風が過剰に発達し計算安定性の観点から問題となる事例があり、この解決が重要な課題であったことから開発されたものである。また、2018 年 10 月に策定された「2030 年に向けた数値予報技術開発重点計画」（気象庁 2018）では、今後台風の予測精度を向上させていくためには台風の内部構造の再現性に注目することの重要性が指摘されており、GSM が表現する台風について更に詳しく調査をするため、ツールの拡張を進めているところである。

具体的なイメージを図 4.2.11 に示す。図は台風を追跡し、台風中心を中心とする円周について平均を取った接線風及び気温アノマリーについてのものであるが、物理過程の各プロセスの加熱及び加湿率や雲などについても描画することが可能である。このような解析を通じて、今後モデルの更新に合わせて台風の構造がどのように変化していくのかを把握しつつ、また部外の専門家の意見も伺いながら、今後の台風予測精度向上に向けて取り組んでいく必要があると考えている。また、粗い解像度を持つ全球モデルにおいて、どこまで現実に近い台風を表現すべきかどうかも難しい課題であるため、領域モデルによる様々な分解能の結果との比較も今後有益であろう。現時点では、このツールで可能なことは台風構造の変化の把握にとどまるため、今後はリファレンスとして使える観測データや解析プロダクトが求められる。

#### 4.2.10 台風進路予測の成分別検証

台風予測については、実験期間を通じた進路予測精度や強度予測精度の比較、個別台風ごとの比較、強風

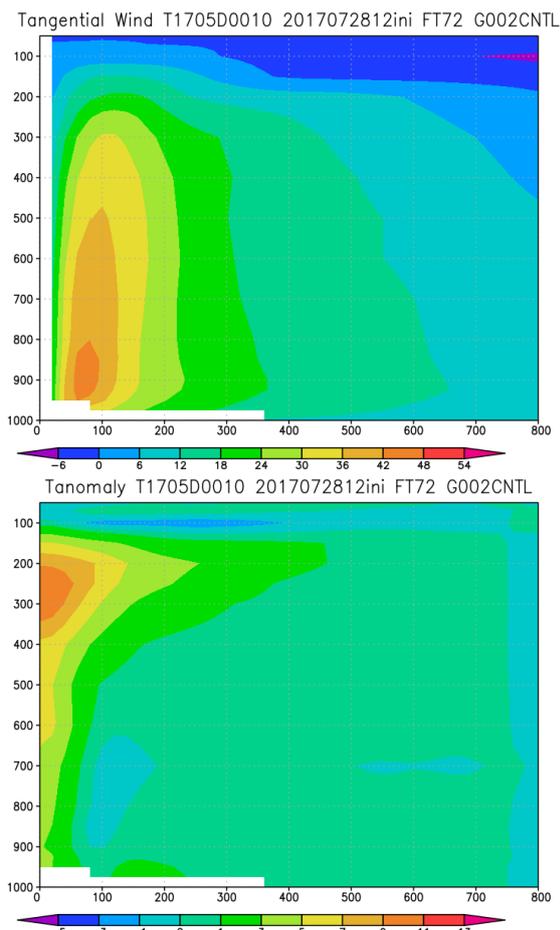


図 4.2.11 GSM1705 の台風表現例。2017 年 7 月 28 日 12UTC 初期値 FT=72 における台風第 5 号の予測。台風中心を中心とする円周に対して、接線風 [m/s] (上図) 及び気温アノマリー [K] (下図) を平均したもの。縦軸は気圧 [hPa] で横軸は台風中心からの動径距離 [km]。気圧面の円周上において、気圧が地表面気圧より高くなる点の一つでも持つ動径距離については、データを取り除いている (図中左下の空白域)。

半径の比較など様々な観点で検証を行っている。特に、進路予測については成分別検証と各事例の進路予測結果を突き合わせることでその特性を調べている。

台風進路予測誤差は台風の進行方向 (Along Track、以下 AT) とそれに直交する方向 (Cross Track、以下 CT) の成分に分解することで、進行速度と進行方向に関する誤差として把握することができる (梅津・森安 2013)。性能評価試験において進路予測誤差の成分別の検証を行っているが、ここではルーチンの検証結果を用いて近年の誤差特性の変遷を確認する。台風のステージを転向前・転向中・転向後に分け (定義は梅津・森安 2013 参照)、2014 年から 2017 年までの北西太平洋領域における年ごとの AT と CT の誤差分布を図 4.2.12 に示す。まず、各年に共通する誤差として、2014 年から 2016 年には転向後の進行が遅い傾向が見られる。ただし、2017 年にはこの傾向が不明瞭である。転向前や転向中の AT 成分の誤差では、進行が遅い事例の多い年

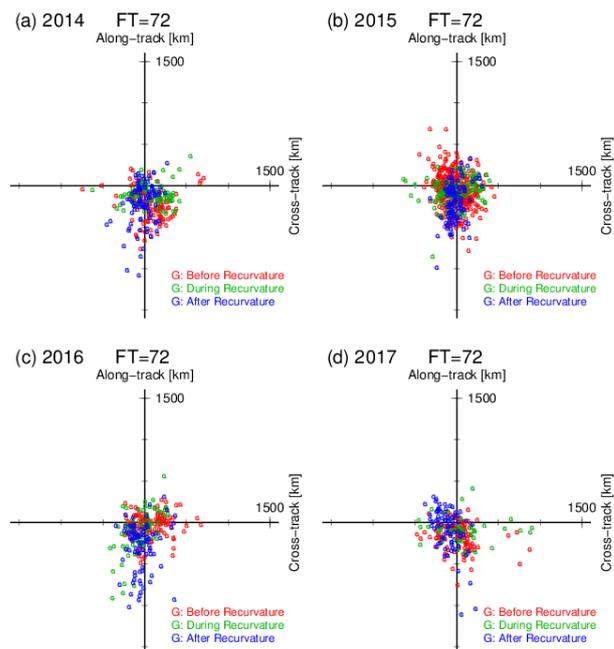


図 4.2.12 ルーチンの台風進路予測誤差の分布図。(a) 2014 年、(b) 2015 年、(c) 2016 年、(d) 2017 年の予測時間 FT=72、北西太平洋領域を対象としている。縦軸は台風の進行方向、横軸は進行方向に直交する方向の誤差 [km] を示す。赤色は転向前、緑色は転向中、青色は転向後を表す。

もあるが、特徴は年により異なる。また、CT 成分の誤差では、どのステージにおいても各年に共通した誤差特性は見られない。2014 年から 2017 年の間に GSM は 3 回更新されている (第 1.1 節)。このうち、GSM1705 への変更では進路予測誤差の特性変化は小さいことが性能評価試験により確認されている (米原 2017b)。このため、2017 年にそれまでの転向後の進行が遅い傾向が不明瞭となったことについては、GSM 更新の影響よりも 2017 年の台風事例の特徴に依存した結果と考えられる。台風予測の誤差特性を把握するためには、複数年の事例による評価が必要であることを示唆する結果といえる。

#### 4.2.11 まとめ

本節では、GSM の予測特性の課題についてその一部を検証手法とともに紹介した。紹介した誤差特性には、GSM が長年の課題としているバイアスや、近年の改良に伴って新たに注目されるようになったものが含まれている。近年の改良により、長年の課題であったバイアスは改善されてきているが、根本的な解決には至っていない課題も依然として存在する。

検証システムにおける今後の課題を記す。DPSIVS は、近年の GSM 開発において中心的課題である「compensating errors の解消」のため、改良項目の組み合わせ実験の評価検証を効率的に実施し、開発者へのフィードバックを増やすことを大きな目的として整備された。検証システムの発展はモデル開発の重要なテーマであ

り、GSM の更なる精度向上のため今後も拡充が求められる。まず、検証値として利用できる観測データや観測プロダクトをさらに収集し、評価可能な要素を地道に増やす必要がある。また、新しく提案された検証手法を取り込んでいくことが重要なのは言うまでもない。特に、雲についての3次元的な構造や日変化の検証は重要であり、衛星観測シミュレーションを利用した雲検証や、全球領域での降水の日変化特性の検証などの拡充を進める必要がある。また、現業数値予報センターである強みを最大限に活かし、データ同化で利用されている観測結果の評価検証での更なる有効活用も模索すべきであろう。他にも、大気現象に着目した検証を増やしていくことが望まれる。例えば、気象条件に応じたサンプリングに基づく検証や、現象を抽出して合成するなどが有益であろう。

数値予報モデルの改良に際しては、総合的な改善がみられる一方で、特定の要素や予測時間において悪化が見られる場合も多くあり、モデルの更新時には誤差特性の変化を網羅的に把握し直す作業が不可欠である。そのコストは決して低いものではないが、長期的な GSM の精度向上を支える屋台骨の作業でもある。今後も、検証システムの整備を進めて作業効率化を図りつつ、モデルの誤差特性を把握することが重要である。引き続き、検証システムの発展を通じて開発を効率的に行い、GSM の予測精度向上に努めていく。

## 参考文献

- 堀田大介, 原旅人, 2012: 物理過程開発のボトムアップ・アプローチとトップダウン・アプローチ. 数値予報課報告・別冊第 58 号, 気象庁予報部, 120–122.
- 石田純一, 2017: 数値予報システム開発のプロセス. 数値予報課報告・別冊第 63 号, 気象庁予報部, 4–10.
- 気象庁, 2018: 2030 年に向けた数値予報技術開発重点計画. 52pp.
- 宮本健吾, 中川雅之, 中村貴, 北川裕人, 小森拓也, 2009: 対流. 数値予報課報告・別冊第 55 号, 気象庁予報部, 68–82.
- Trenberth, K. E. and J. T. Fasullo, 2010: Simulation of Present-Day and Twenty-First-Century Energy Budgets of the Southern Oceans. *Journal of Climate*, **23**, 440–454.
- 梅津浩典, 森安聡嗣, 2013: WGNE 熱帯低気圧検証. 数値予報課報告・別冊第 59 号, 気象庁予報部, 98–111.
- 米原仁, 2017a: 活用例 (1)–全球モデル. 数値予報課報告・別冊第 63 号, 気象庁予報部, 29–49.
- 米原仁, 2017b: 全球数値予報システムの特性的変化. 平成 29 年度数値予報研修テキスト, 気象庁予報部, 8–30.

## 4.3 再予報型の予測実験<sup>1</sup>

### 4.3.1 はじめに

再予報とは、過去の多数の事例を対象に同一のモデルを用いて行う数値予報である (Hamill et al. 2006; 高谷 2012)。GSM の開発プロセスの中では、その精度評価のために再予報型の予測実験 (以下、再予報実験) を行っている。通常、GSM の開発では直近の特定の年の夏・冬期間を対象に性能評価試験を行う (米原 2017a)。しかし、性能評価試験だけではその対象とする実験期間・予報時間以外での予測特性・精度の確認、年々変動を伴う現象の評価を行なうことは出来ない。第 4.1 節で述べたように、再予報実験は性能評価試験だけでは把握できないこれらの点を補う役割をもつ。

本項では、GSM の開発プロセスの一部として行っている再予報実験について、その仕様、検証内容、検証結果から見える GSM の予測特性や課題を述べる。

### 4.3.2 実験仕様

表 4.3.1 に、再予報実験の仕様を示す。GSM の性能評価試験と異なる点として、予報モデルの水平解像度を低くしている点が挙げられる。低解像度に変更することで実験を行なうために必要な計算機資源を大幅に削減し、性能評価試験よりも多数の初期時刻に対してより長い予報時間の実験を行なうことが可能になっている。再予報実験では、GSM の性能評価試験では評価出来ない実験期間・予報時間を補えるように、過去 30 年 (1981~2010 年) の各季節の初期時刻に対して予報時間を 34 日とした予測実験を行なっている。大気初期値は気象庁 55 年長期再解析 (JRA-55; Kobayashi et al. 2015) の解析値を水平・鉛直内挿して作成している。本来は性能評価試験と同様に、解析予報サイクルを通じた初期値の変化も含めて評価出来ることが望ましいが、過去の複数年に渡って解析予報サイクル実験を行なうことは計算機資源の制約上困難であるため、既存の再解析値を用いている。

再予報実験は全球アンサンブル予報システム (以下、全球 EPS: Global Ensemble Prediction System) で用いる予報モデルとしての予測特性の確認も兼ねている。このため、全球 EPS の開発における精度評価や系統誤差補正用統計量の計算に用いられる再予報の仕様 (金浜 2017a) とほぼ同様の仕様としている<sup>2</sup>。ただし、性能評価試験と同等の実時間で実験を行なうことが出来るように初期時刻数を減らして実行している。全球 EPS の再予報では過去 30 年に渡って毎月 3 初期時刻 (10 日、20 日、月末の 12UTC) の予測を行うが、再予報実験ではその一部の各季節 1 初期時刻 (3 月、6 月、9

月、12 月の月末の 12UTC) の予測を行う。2 初期時刻の予測を並行して実施する場合 5 日程度で終了し、検証・評価を行うことができる。特定の季節について詳細に GSM 改良のインパクトを確認したい場合には、その季節の初期時刻の予測実験を集中的に行う。

### 4.3.3 検証内容

再予報実験に対して行っている主な検証内容を表 4.3.2 に示す。再予報実験では、まずはアンサンブル平均の予測場の平均誤差に着目する。性能評価試験だけでは、予測 2 週目以降の平均誤差の拡大傾向や、春・秋期間における想定外の平均誤差の増加を見落とす可能性がある。各季節について再予報実験を行い、予報時間別に検証を行うことで、そのような見落としの可能性を減らすことが出来る。また、予測場のアノマリー相関係数や RMSE の空間分布に着目して、大きな空間スケールで顕著な改善・悪化傾向が見られるかどうか確認する。これらの検証では解析値の不確実性も考慮して、JRA-55 解析値を検証値として評価するだけでなく、ERA-Interim (Dee et al. 2011) を検証値とした評価も実施している。

平年偏差の予測精度は、循環指数<sup>3</sup> や日本周辺域の平年偏差のアノマリー相関係数で評価している。これらの指数に大きな変化が見られた場合、その指数の対象としている領域に着目して予測特性の変化を確認する。

性能評価試験が対象としている実験期間だけでは十分に評価できない現象に関する検証も行なっている。例えば、アジアモンスーン域の降水や循環場の表現、マッデン・ジュリアン振動の表現、ブロッキング高気圧の発生頻度、成層圏突然昇温の予測特性の変化を確認している。これらの予測特性を確認することは、全球モデルとしての性能を評価する観点だけでなく、予報作業で利用する上での留意点を把握する観点でも重要である。

### 4.3.4 検証結果から見える予測特性

再予報実験の検証例として、最新の GSM のバージョンである GSM1705 (米原 2017b) を用いた実験の検証結果を示す。なお、ここでは初期時刻が 3 月 31 日 12UTC の予測を春初期日、6 月 30 日 12UTC の予測を夏初期日、9 月 30 日 12UTC の予測を秋初期日、12 月 31 日 12UTC の予測を冬初期日と呼ぶ。

#### (1) 平均誤差

各季節の予測 1~4 週目で時間平均したアンサンブル平均予報の平均誤差に着目する。図 4.3.1 に大気上端上向き短波放射の平均誤差を示す。夏・冬初期日を見ると、概ね第 4.2 節の性能評価試験の検証結果と同様

<sup>1</sup> 越智 健太 (地球環境・海洋部 気候情報課)、関口 亮平 (地球環境・海洋部 気候情報課)

<sup>2</sup> 下部境界条件とする海面水温・海水データの初期値には解析値を用いるが、その時間変化は気候値で表現している。

<sup>3</sup> 循環指数とは、大気大循環の状態を見るためにその特徴をよく表すように作られた指数である (藤川 2013)。例えば、偏西風が南北に蛇行しているか東西の流れが卓越しているか確認するための東西指数や、北極域の寒気の蓄積の度合いを確認するための極渦指数がある。

表 4.3.1 再予報実験の仕様

予報モデル	水平解像度	TL479 (約 40 km) (～18 日目)、TL319 (約 55 km) (18 日目～)
	鉛直層数	100 層 (最上層 0.01 hPa)
初期値	大気	JRA-55 再解析値
	陸面 (土壌水分を除く)	オフライン陸面解析値 (JRA-55 強制)
	土壌水分	気候値 (オフライン陸面解析に基づく)
下部境界値	海面水温	MGDSST 初期偏差持続 (摂動あり)
	海水	MGDSST に利用した海水データセット (予測期間中は統計的推定)
アンサンブル手法	大気初期摂動	SV 法 (Initial SV + Evolved SV)
	モデル摂動	確率的物理過程強制法
	境界摂動	SST 摂動
アンサンブルメンバー数		5
予報時間		34 日
実験期間		1981～2010 年 (30 年間)

表 4.3.2 再予報実験の主な検証内容

	検証項目	検証要素
予測場の決定論的検証	平均誤差	高度 (500 hPa)、気温 (850 hPa、地上)、海面更正気圧、速度ポテンシャル (200 hPa)、流線関数 (200 hPa, 850 hPa)、降水、放射フラックス、帯状平均東西風・気温
	アノマリー相関係数・RMSE	高度 (500 hPa)、気温 (850 hPa、地上)、海面更正気圧、速度ポテンシャル (200 hPa)、流線関数 (200 hPa, 850 hPa)
年々変動の評価	循環指数のアノマリー相関係数	東西指数、極渦指数など (詳細は藤川 (2013) 参照)
	日本周辺域の平年偏差のアノマリー相関係数	北日本、東日本、西日本、沖縄・奄美周辺における領域平均気温 (850 hPa、地上)
現象に着目した検証	アジアモンスーン	アジアモンスーン域の降水の季節変化や WY 指数 (Webster and Yang 1992) の時系列
	ブロッキング高気圧	出現頻度
	成層圏突然昇温	極域成層圏の気温の時系列
	マッデン・ジュリアン振動	MJO 指数 (Wheeler and Hendon 2004) の相関係数、RMSE、振幅・位相の平均誤差
確率検証	ブライアスキルスコア	高度 (500 hPa)、気温 (850 hPa、地上)、海面更正気圧
	ROC 面積スキルスコア	
スプレッド	スプレッド / RMSE 比	高度 (500 hPa)、気温 (850 hPa、地上)、海面更正気圧、速度ポテンシャル (200 hPa)

に、CERES プロダクトに比べて夏半球側の陸上や南大洋で短波放射の反射が過少であることが分かる。また、性能評価試験の対象としていない春初期日ではアリューシャン列島付近を中心に短波放射の反射が過少であるといった、他の初期日では明瞭に見られない傾向があることも分かる。大気上端上向き長波放射については、どの初期日についても概ね同様に過剰傾向が見られている (図略)。性能評価試験と共通して見られる特徴が多いことから、放射フラックスの予測特性は予報モデルの解像度や予報時間によって大きく変わらないと考えられる。

熱帯域の対流活動や循環場が予測期間中にどの程度

維持できているか確認するために、図 4.3.2、図 4.3.3 にそれぞれ降水量、200 hPa 速度ポテンシャルの平均誤差を示す。各季節の降水量の平均誤差を比べると、どの季節にも共通してインド洋西部で GPCP<sup>4</sup> の月別解析値 (Adler et al. 2003) に比べて降水量が多い傾向が見られる。また、季節特有の傾向として夏初期日のフィリピン海周辺の降水量が少ない傾向や、春期から秋期にかけてフィリピン海から日本の南海上で降水量が少ない傾向が見られる。200 hPa 速度ポテンシャルの平均誤差 (図 4.3.3) を見ると、降水量の平均誤差に対応して、インド洋西部を中心に上層発散が強い傾向

<sup>4</sup> Global Precipitation Climatology Project

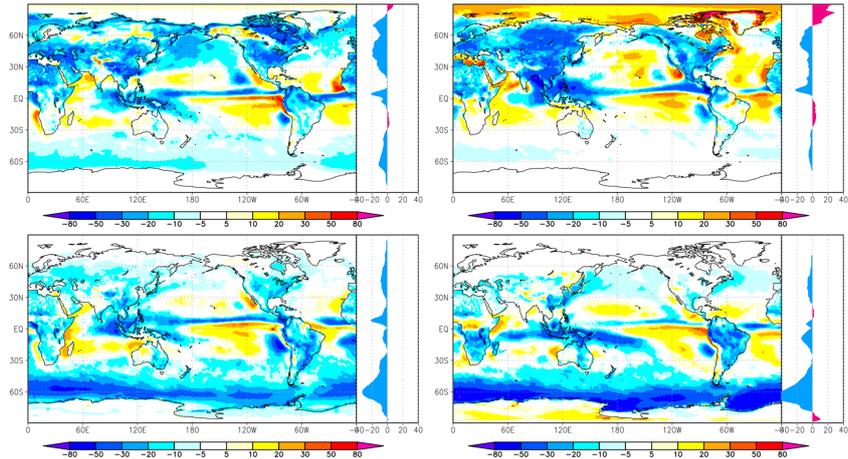


図 4.3.1 大気上端上向き短波放射の平均誤差 [ $\text{W m}^{-2}$ ] (予測 3~30 日目の 28 日平均場)。予測の初期時刻は、左上：3 月 31 日 12UTC (春初期日)、右上：6 月 30 日 12UTC (夏初期日)、左下：9 月 30 日 12UTC (秋初期日)、右下：12 月 31 日 12UTC (冬初期日)。平均誤差は CERES プロダクトの月平均値からの差を示す。図の右側には東西平均したものを緯度ごとに表示している。

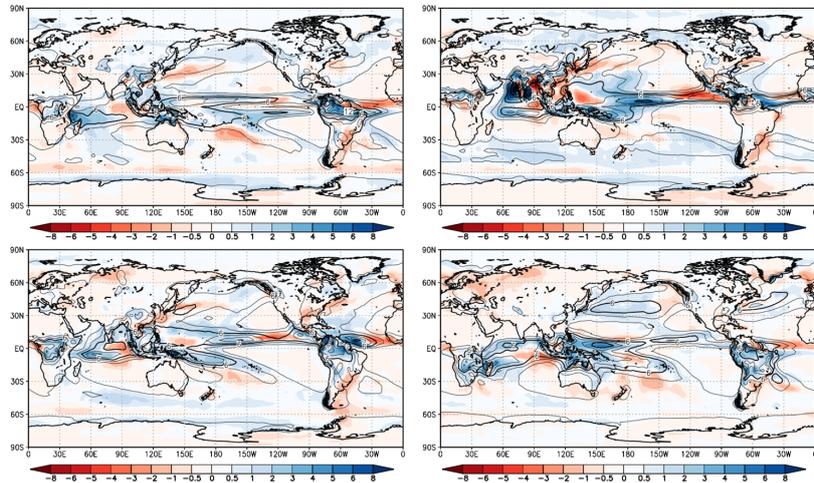


図 4.3.2 降水量の平均誤差 (色) とモデル平年値 (等値線) [ $\text{mm/day}$ ] (予測 3~30 日目の 28 日平均場)。予測の初期時刻は、左上：3 月 31 日 12UTC (春初期日)、右上：6 月 30 日 12UTC (夏初期日)、左下：9 月 30 日 12UTC (秋初期日)、右下：12 月 31 日 12UTC (冬初期日)。平均誤差は GPCP 月平均値からの差を示す。

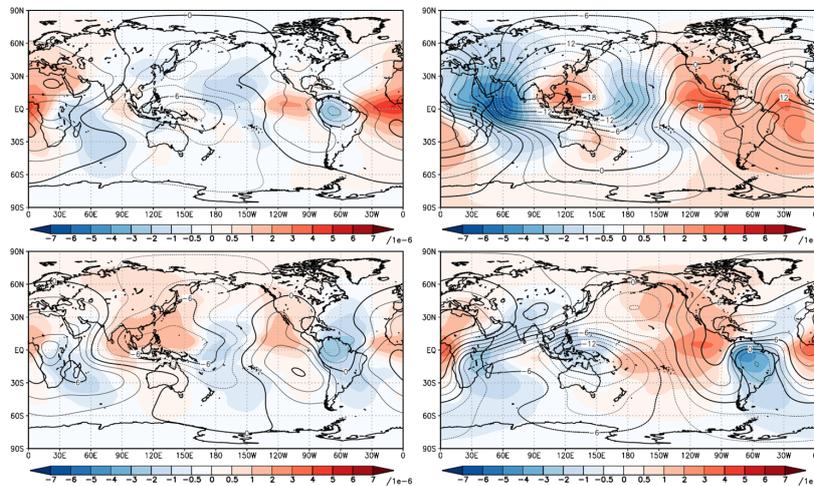


図 4.3.3 図 4.3.2 と同様。ただし、要素は 200 hPa 速度ポテンシャル [ $\times 10^6 \text{ m}^2 \text{ s}^{-1}$ ]。平均誤差は JRA-55 からの差を示す。

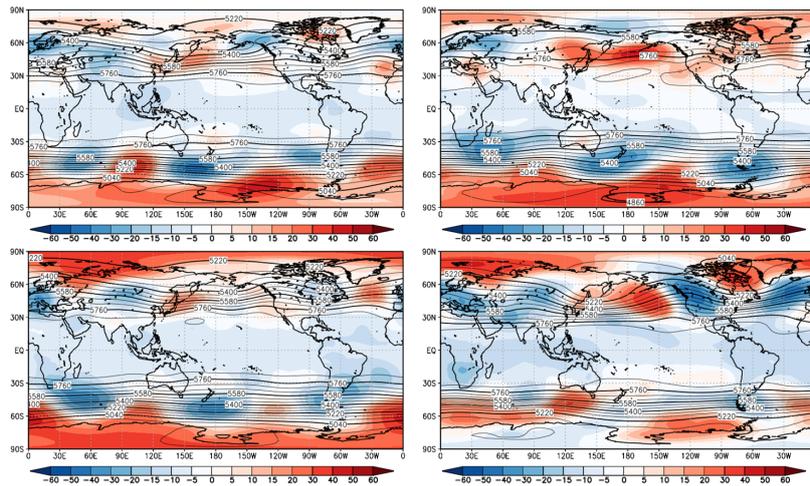


図 4.3.4 図 4.3.2 と同様。ただし、要素は 500 hPa 高度 [m]。平均誤差は JRA-55 からの差を示す。

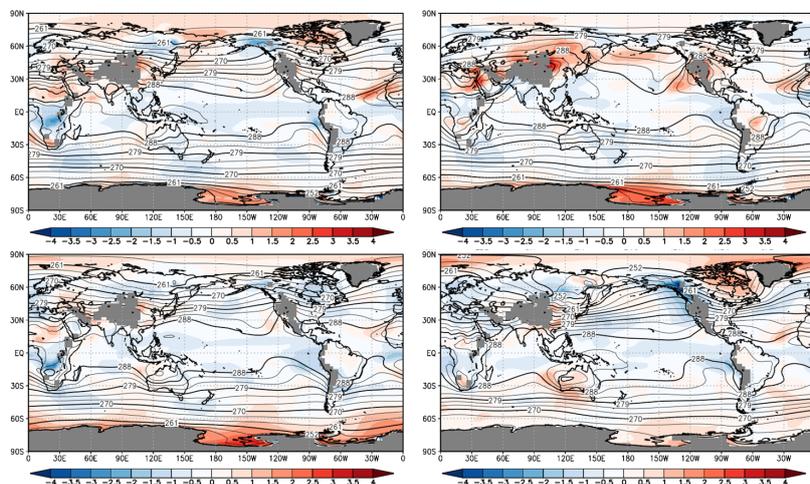


図 4.3.5 図 4.3.2 と同様。ただし、要素は 850 hPa 気温 [K]。平均誤差は JRA-55 からの差を示す。

や夏初期日のフィリピン海周辺の大規模発散場が弱い傾向があることが分かる。

中・高緯度の平均誤差の特徴を確認するために、図 4.3.4、図 4.3.5 にそれぞれ 500 hPa 高度、850 hPa 気温の平均誤差を示す。500 hPa 高度の平均誤差を見ると、年間を通してアリューシャン列島周辺で 500 hPa 高度が高い傾向が見られる。この誤差は夏期の太平洋高気圧の北縁で気圧が高い傾向、冬期のアリューシャン低気圧が弱い傾向に対応する。冬初期日の日本周辺に着目すると、第 4.2 節で指摘されたトラフの前面が浅く予想される傾向に対応する正バイアスはそれほど大きくないように見える。予測を週別に確認すると、予測 1 週目は第 4.2 節と同様の傾向が見られているが、予報時間の経過とともに正バイアスの傾向は緩和している (図略)。予測の誤差には様々なプロセスが関わっているためその原因は明確ではないが、このように予報時間によって平均誤差の傾向が変わって見えること

もある。850 hPa 気温の平均誤差を見ると、夏初期日のシベリア付近に他の季節には見られない正バイアスがあることが分かる。現在の全球 EPS (予報モデルは GSM1705 より前のバージョンである GSM1603E; 山口 2017) ではこの正バイアスは予測 1 週目では小さいものの、予報時間の経過とともに GSM1705 よりも大きく拡大している。GSM1705 の開発時にはその平均誤差の拡大傾向にも着目しながら物理過程の改良が行ったことで、平均誤差を縮小させることが出来ている。

## (2) 現象に着目した検証

現象に着目した検証の例として、アジアモンスーンに伴う降水の季節変化、ブロッキング高気圧の出現頻度、成層圏突然昇温の検証結果を示す。なお、ここで示す検証結果は各季節 1 初期時刻の予測だけでなく、他の初期時刻の予測も合わせて評価したものである。

図 4.3.6 にインド付近とフィリピン付近で経度方向に平均した日平均降水量平年値 (1998~2010 年) の

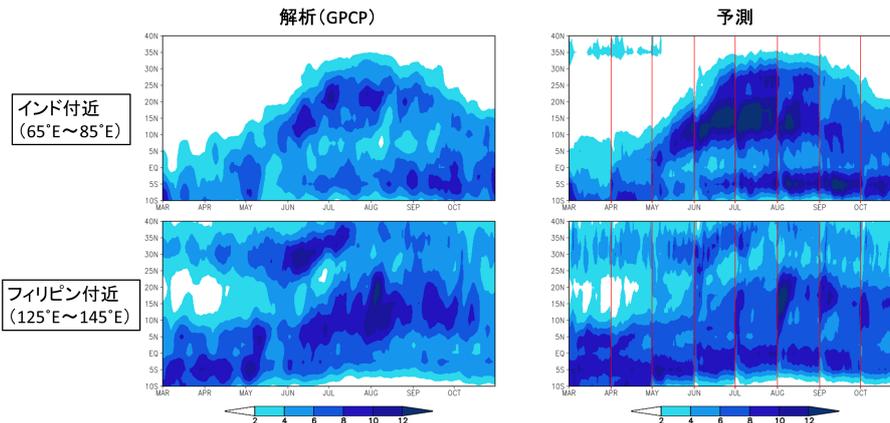


図 4.3.6 降水量平年値の緯度・時間断面図。上段はインド付近（東経 65~85 度）、下段はフィリピン付近（東経 125~145 度）で経度平均した日別降水量 [mm/day] の 1998~2010 年平均値の季節変化（3~10 月）。左：GPCP、右：各月末初期日の予測値。予測値は各初期日の予測をつなぎ合わせて示している（赤線は初期日の位置を示す）。

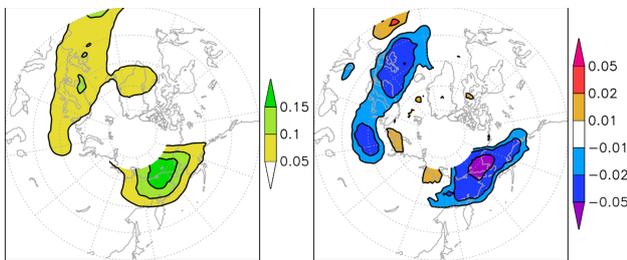


図 4.3.7 北半球冬季におけるブロッキング高気圧の平均出現頻度分布。左：解析値 (JRA-55)、右：予測値と解析値の差。1981/1982~2009/2010 年 (29 年) の 12~2 月、リードタイム 4~31 日目を対象としている。

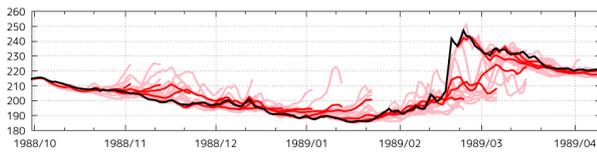


図 4.3.8 1988 年 10 月から 1989 年 4 月にかけての北極上空 30 hPa 気温 [K] の時系列。黒線は解析値 (JRA-55)、赤線はアンサンブル平均した予測値。桃色線は各アンサンブルメンバーの予測値を表す。

緯度時間断面図を示す。解析は GPCP の日別解析値 (Huffman et al. 2001)、予測は 2~9 月の月末初期日の予測を並べたものであり、図の赤線は初期日の位置を示している。まず、インド付近（上段）に着目すると、北緯 5~30 度付近を中心に GPCP に比べて予測の降水量が過剰であることが分かる。また、GPCP ではアジアモンスーンの季節変化に伴って降水の極大域が北緯 20 度まで北上しているが、予測では北緯 15 度付近に極大域があることが分かる。これらの傾向は予測初期から予測期間を通じて見られている。次に、フィリピン付近（下段）に着目すると、北緯 5~15 度や北

緯 25~35 度付近を中心に GPCP に比べて予測の降水が過少であることが分かる。北緯 5~15 度付近の過少傾向は予測期間を通して一貫して見られるものではなく、6~8 月末初期日では予測時間の経過とともに降水量が減少する傾向が明瞭に見られている。

図 4.3.7 に解析値 (JRA-55) における北半球冬季のブロッキング高気圧の平均出現頻度分布<sup>5</sup>と、予測における平均出現頻度分布とその解析値との差を示す。解析値では大西洋からヨーロッパにかけての領域と極東からアラスカにかけての領域で平均出現頻度が多くなっているが、予測では解析値に比べると頻度が少ないことが分かる。予測を週別に確認すると、この傾向は特に予測 1 週目から 2 週目にかけて明瞭になっている (図略)。

図 4.3.8 には、成層圏突然昇温の予測の一例として 1988 年から 1989 年にかけての冬季の 30 hPa 気温の時系列を示す。解析値 (黒線) と比べると、予測 (赤線：アンサンブル平均、桃色：各メンバー) では解析値よりも 2 月後半の突然昇温時の気温の上昇幅が小さいことが分かる。このように大きな昇温が十分に予測できていない傾向は他の事例でも概ね同様に見られる (図略)。金浜 (2017b) では、全球 EPS の予報モデルの水平解像度の高解像度化、鉛直層数の増強 (最上層は 0.1 hPa から 0.01 hPa に変更)、物理過程の変更などによって成層圏突然昇温の予測がより解析値に近くなっていることが示唆されている。

ここで示した現象の予測には予報モデルの様々なプロセスが関わっており、その改善のための方法は必ずしも明確になっていない。今後も予報モデルの改良に合わせて、適宜、再予報実験を行い、これらの現象に対する影響を確認しながら理解を深めていくことが必要である。

<sup>5</sup> ブロッキング高気圧の検出方法の詳細は佐藤ほか (2015) を参照いただきたい。

#### 4.3.5 今後の課題

本項では再予報実験のねらいやその仕様・検証内容、予測特性について紹介した。数値予報モデルの継続的な改良のためには、多角的な検証によって予測特性の変化やそのメカニズムを正しく理解することが重要である。そこで得た知見によって、次の改良に向けた開発にフィードバックを行うことが出来る。本項で紹介した予測特性の中には、今後の改良に向けて継続的に取り組みが進められているものもある。引き続き、性能評価試験だけでなく再予報実験やその検証を通じて課題を明らかにしながら、GSMの精度向上に向けて取り組む必要がある。

再予報実験の仕様にも検討の余地があるだろう。例えば現在のメンバー数は5メンバーであるが、アンサンブル平均の予測や確率予測精度の評価を行なう上では不十分であることが知られている (Ma et al. 2012)。また、各季節それぞれ1初期時刻の実験では大きな空間スケールの平均誤差の変化傾向を捉えることはできるが、局所的なスケールの予測精度の変化傾向は十分に捉えることが出来ないことがある。予報モデルの変更の効果を適切に把握出来るように、計算機資源を勘案しながら、より評価に適した再予報の仕様や検証項目についても検討を進めていきたい。

#### 参考文献

- Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P. P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor.*, **4**, 1147–1167.
- Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597.
- 藤川典久, 2013: 季節予報用語集. 平成 24 年度季節予報研修テキスト, 気象庁地球環境・海洋部, 319–361.
- Hamill, T. M., J. S. Whitaker, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46.
- Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind, 2001: Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeor.*, **2**, 36–50.
- 金浜貴史, 2017a: 再予報の仕様. 平成 28 年度季節予報研修テキスト, 気象庁地球環境・海洋部, 9–11.
- 金浜貴史, 2017b: QBO、成層圏突然昇温. 平成 28 年度季節予報研修テキスト, 気象庁地球環境・海洋部, 45–46.
- Kobayashi, S., Y. Ota, Y. Harada, A. Ebata, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48.
- Ma, J., Y. Zhu, R. Wobus, and P. Wang, 2012: An effective configuration of ensemble size and horizontal resolution for the NCEP GEFS. *Advances in Atmospheric Sciences*, **29**, 782–794.
- 佐藤均, 宮岡健吾, 長澤亮二, 新保明彦, 高谷祐平, 松枝聡子, 杉本裕之, 2015: ハインドキャストによる検証. 平成 26 年度季節予報研修テキスト, 気象庁地球環境・海洋部, 22–45.
- 高谷祐平, 2012: 再予報・ハインドキャスト. 天気, **59(6)**, 493–495.
- Webster, P. J. and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quart. J. Roy. Meteor. Soc.*, **118**, 877–926.
- Wheeler, M. C. and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932.
- 山口春季, 2017: 全球アンサンブル予報システムの導入. 平成 29 年度数値予報研修テキスト, 気象庁予報部, 35–41.
- 米原仁, 2017a: 活用例 (1) 全球モデル. 数値予報課報告・別冊第 63 号, 気象庁予報部, 29–49.
- 米原仁, 2017b: 全球数値予報システムの改良の概要. 平成 29 年度数値予報研修テキスト, 気象庁予報部, 1–7.

## 4.4 1年積分共通評価ツール<sup>1</sup>

### 4.4.1 はじめに

GSMの開発時に実行する解析予報サイクル実験を伴わない予測実験型の検証として、1年積分共通評価ツール(COOL: Common evaluation tOOL for one-year model integration)がある。COOLは下部境界値に海面水温(SST)及び海水密度(SIC)の解析値を与えて現業仕様よりも低解像度で1年積分を実行する共通実験環境である。GSMに限らず全球域を予測対象とした数値予報モデルの開発に使われていることや、必要な計算機資源を抑えた仕様にすることで手軽に実行できることが特徴である。本節では、開発におけるCOOLの役割を述べた後、COOLを用いた実験(以下、COOL実験)の実験仕様及び実験結果の評価・検証といったツールの概要をまとめ、モデル開発者が利用する際に留意している点について触れる。また、気象庁が現業運用してきたGSMに対するCOOL実験の検証結果のモデル間比較や、第4.2節や第4.3節で示した性能評価試験や再予報実験から見える特性との比較を通して、モデル開発における今後の課題を述べる。

### 4.4.2 開発におけるCOOLの役割

気象庁は全球域を予測対象とした現業数値予報システムとして、GSM、全球アンサンブル予報システム(全球EPS: Global Ensemble Prediction System)、季節アンサンブル予報システム(季節EPS)を運用している。これらの数値予報システムに何らかの変更を加える際には、基礎評価試験(鉛直1次元モデルによる理想実験や低解像度実験など)を通して期待される結果が得られることを確認した後、現業システムと同じ仕様で行う性能評価試験や業務化試験を経る<sup>2</sup>(原2013; 石田2017)。COOL実験は基礎評価試験の一つとして位置づけられており、開発初期段階での鉛直1次元モデルなどの理想実験による評価が完了した後、性能評価試験や再予報実験に進む前に実行することを想定している。また、モデルの予測対象や利用目的を問わず全球数値予報モデル開発者が共通の仕様で実験を行うことで、開発成果や課題の共有、科学的な議論を促すことが期待される。

性能評価試験や再予報実験とは異なり、AMIP<sup>3</sup>型の実験<sup>4</sup>を行うことで下部境界の時間発展に由来する誤差を除いた評価が可能となる。AMIP型の実験でモデ

<sup>1</sup> 松川 知紘

<sup>2</sup> 気候情報課では、全球EPSや季節EPSに関する性能評価のために過去30年分の再予報実験を実施している。

<sup>3</sup> Atmospheric Model Intercomparison Projectの略で、観測された月別海面水温・海水分布を境界条件に与えて積分した過去実験の結果をモデル間で比較する、大気モデル相互比較プロジェクト(Gates1992; Gates et al. 1999)。

<sup>4</sup> 本節では、SST及びSICの推定値を下部境界値に与える性能評価試験や再予報実験と対比する意味で、解析されたSST及びSICを下部境界値に与えた実験をAMIP型と呼ぶ。

ルの特性を評価するためには複数年を対象とした実験を行うことが望ましいが、COOL実験は計算機資源や実行時間の面で手軽に実行できることを優先して実験期間を1年にしている。性能評価試験や再予報実験と比較して実験にかかる時間が短くなる<sup>5</sup>ことで、実験開始から短い時間で結果を確認することができる。

COOL実験を行う目的として以下の点が挙げられる。

- 加えた変更から想定される変化が現れているか、意図しない変化が現れていないかといった確認
- 基本場の維持性能として、季節変化に卓越するような極端なドリフトが見られないかといった確認
- 全球エネルギー収支や降水量などの基本的な物理量の変化や予測特性の把握

一方で、時空間スケールの細かいインパクトや予測精度に関する議論、年々変動などの気候学的な議論を目的にCOOLを使うことはできないため、性能評価試験や再予報実験、AMIP型の複数年を対象とした長期積分実験などを目的に応じて使い分けしている。

近年の全球数値予報モデル開発におけるCOOLの利用実績について紹介する。大きく分けて二つの用途(i)開発時における変更のインパクトの確認や基礎的な予測特性の評価、(ii)現業化されたバージョンなどの複数モデル間の相互比較、で利用されている。一つ目の用途については既に述べた通りである。二つ目に挙げた相互比較は、実験仕様を統一したことによって可能となった。性能評価試験は通常は直近の夏及び冬を実験期間とすることから、実験期間が異なる実験間の比較が非常に困難であったのに対し、COOL実験では実験対象年も含めて実験仕様を揃えている。また、例えばGSMと季節EPSなど、実験仕様の違いから比較が困難であったモデル同士の比較も可能となり、全球数値予報モデルの開発者間での成果や課題の共有が進んでいる。

### 4.4.3 実験仕様

表4.4.1にCOOLの実験仕様を、図4.4.1にCOOL実験の模式図を示す。水平解像度はTL159(約110km)と現業仕様よりも低解像度にする<sup>6</sup>ことで、計算機資源を抑えている。鉛直層数は任意に設定することとしており、現業数値予報システムの鉛直層数と揃えることが一般的である<sup>7</sup>。大気初期値には気象庁55年長期

<sup>5</sup> 第10世代スーパーコンピュータシステム上において、2018年12月時点で現業化されているバージョンのGSMでのCOOL実験は、表4.4.1の正規版が約8時間、簡易版が約3時間で終了する。

<sup>6</sup> 2018年12月現在での現業数値予報システムの水平解像度は、GSMがTL959(約20km)、全球EPSの積分18日目までがTL479(約40km)、全球EPSの積分18日目以降がTL319(約55km)、季節EPSの大気モデルがTL159である。

<sup>7</sup> 2018年12月現在での現業数値予報システムの鉛直層数は、GSM及び全球EPSが100層、季節EPSの大気モデルが60層である。

表 4.4.1 COOL の実験仕様

水平解像度	TL159 (約 110 km)	
鉛直層数	任意 (各モデルの鉛直層数に依存)	
初期値	大気	JRA-55 再解析値
	陸面	JRA-55 再解析値もしくはオフライン陸面解析値 (JRA-55 強制)
下部境界値	海面水温	MGDSST の解析値
	海水	MGDSST に利用した海水データセットの解析値
アンサンブル手法	時間間隔を 30 時間とした LAF 法	
アンサンブルメンバー数	正規版	12
	簡易版	4
スピニングアップ	1 か月	
評価期間	2001 年 1 月～2002 年 2 月	

再解析 (JRA-55; Kobayashi et al. 2015) を、陸面初期値には JRA-55 陸面解析値もしくはオフライン陸面解析値<sup>8</sup> を、それぞれ解像度変換して用いる。下部境界値には、気象庁海洋気象情報室によって整備された MGDSST 再解析値 (栗原ほか 2006) 及び、海水解析値データセット (Ishii et al. 2005; Matsumoto et al. 2006) の日別解析値を時間内挿して与える。初期値に由来しないモデル本来の予測特性を評価するために、初期値の影響が及ぶ積分開始後の 1 か月間をスピニングアップとして評価に用いないこととし、その後の 14 か月間<sup>9</sup> を評価期間とした。

COOL 実験では、実験結果の不確実性を小さくすることを狙って複数メンバーによるアンサンブル実験結果をアンサンブル平均して評価する。アンサンブルメンバー数と計算機資源や実行時間はトレードオフの関係にあり、手軽に実行できるツールとするために適切なメンバー数を選択する必要がある。これらの制約から正規版 (12 メンバー) と簡易版 (4 メンバー) のどちらかを利用者が選択するようにした。

エルニーニョ・南方振動などの顕著な現象が発生していない、実験結果の評価に用いる衛星観測データなどの検証データが比較的多く利用できる、といった観点から 2001 年を実験対象年とした。実験期間は、性能評価試験や再予報実験のように変更することはせずに固定とする。実験対象年を変更することで実験結果にどの程度影響を及ぼすのかを把握するため、同じモデルを用いて対象年を 1997 年 (エルニーニョ年)、1999 年 (ラニーニャ年)、2004 年 (平常年) に変更した実験結果を比較したところ、降水量の誤差分布については

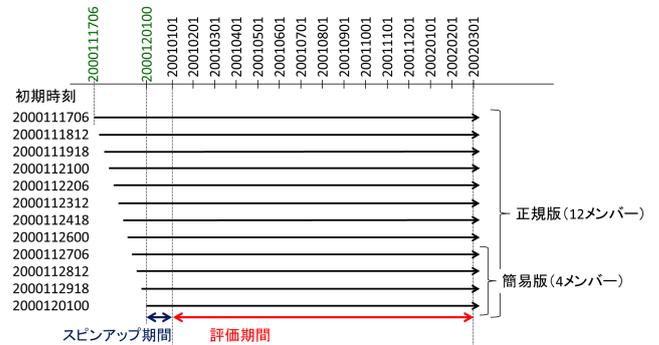


図 4.4.1 COOL 実験の模式図。個々の矢印はアンサンブルにおける各メンバーの予測を表す。

差が比較的小さい一方で、500 hPa ジオポテンシャル高度や海面更正気圧は、特に内部変動が大きい中高緯度において年ごとに誤差傾向が異なる (図略)。COOL 実験は複数年を対象としておらず、単一年の実験結果から得られる誤差やモデル間の差のみの評価が可能となる。モデル気候値や気候学的な系統誤差を評価するためには、複数年を対象とした実験を行う必要があることに留意していただきたい。

#### 4.4.4 評価検証システム

GSM の開発においては、性能評価試験を主な対象とした全球決定論予測についての標準検証環境 (DPSIVS) が、モデル開発者自らによって整備・拡張されている (米原 2017)。COOL における実験結果の評価・検証においても、考え方は DPSIVS 開発当時の思想と同じであり、統合的な検証環境を整備することを目指して DPSIVS のファミリーとなる検証システム (COOLIVS) を整備している。検証システムのあり方や管理方針のみならずディレクトリ構成や起動方法などを DPSIVS と揃えることで、検証システムの利用や機能拡充への敷居を低くすることを試みた。数値予報課、気候情報課、気象研究所などのモデル開発者によって個別に維持・管

<sup>8</sup> GSM1603 (第 1.1.7 項参照) で導入された陸面モデルの変更により、JRA-55 と GSM で陸面モデルの層数や予報変数が異なるようになったため、必要に応じてオフライン陸面解析値を利用している。

<sup>9</sup> 第 4.4.4 項で示すように、COOL 実験の結果を 3 か月平均場で評価するため、越年する 3 か月平均場についても連続した 3 か月間で評価できるように 2 か月余分に積分する。

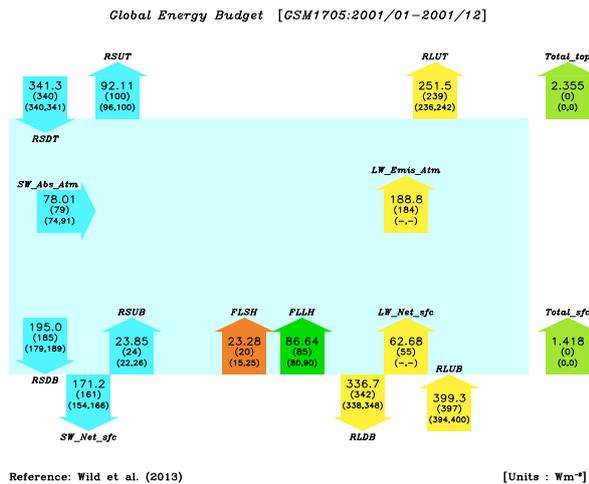


図 4.4.2 全球エネルギー収支の模式図の例。図中の矢印は、大気上端下向き短波放射 (RSDB)、地表面下向き短波放射 (RSUT)、大気上端上向き短波放射 (RSUB)、地表面上向き短波放射 (RLUT)、顕熱 (FLSH)、潜熱 (FLLH)、地表面下向き長波放射 (RLDB)、大気上端上向き長波放射 (RLUB) の各フラックス及びそれらから計算される正味のフラックスを模式的に図示している。矢印中の値 [W m<sup>-2</sup>] は、上段にモデルの 1 年平均値、中段と下段の括弧の中に Wild et al. (2013) の値とその不確か幅を示している。

理されてきた検証ツールを COOLIVS のパッケージとして統合することで、様々な目線で実験結果を評価できる検証システムになることが期待される。

以下に、2018 年 12 月時点で COOLIVS に含まれるパッケージの概要を簡単に紹介する。検証値の追加、新しい検証パッケージの追加や機能拡張など、今後も検証システムを拡充していく予定である。

### Ebudget

全球エネルギー収支は全球数値予報モデルにとって重要な指標の一つであり現実とずれていないことが望ましいことから、放射フラックス、顕熱及び潜熱フラックスから計算される、1 年平均値での全球エネルギー収支を評価する。図 4.4.2 に例を示したエネルギー収支の模式図に加えて、各フラックスの値の一覧表や 3 か月平均値の季節変化などから、エネルギー収支の実験間比較や検証値 (Wild et al. 2013) との比較ができる。

### Zonalmean

各実験結果、実験間の差分、各実験の誤差 (検証値との差分) について、3 か月及び 1 年平均場に対して帯状平均した高度 1 hPa までの南北-高度断面図を描画する。描画する要素は東西風、南北風、気温、比湿、相対湿度、鉛直 p 速度、雲水量などで、再解析データ (JRA-55 や ERA-Interim (Dee et al. 2011) など) を検証値としている。

### Errmap

Zonalmean と同様に、各実験結果、実験間の差分、各実験の検証値との差分を、3 か月及び 1 年平均場の空間分布として描画することで、実験間での変更点のインパクトや誤差の空間分布を把握することができる。検証値として、再解析データ、CERES<sup>10</sup> の放射フラックス、GPCP<sup>11</sup> の降水量、OAFflux<sup>12</sup> の顕熱・潜熱フラックス、SOC<sup>13</sup> の顕熱・潜熱フラックス及び運動量フラックスなどのプロダクトを用いて評価している。

### 4.4.5 予測特性

GSM1403, GSM1603, GSM1705 (それぞれ第 1.1 節参照) について、COOL 実験の評価結果を放射収支に関する特性を中心に比較した後、GSM1705 の予測特性について第 4.2 節や第 4.3 節で示した性能評価試験や再予報実験と比較しながら述べる。また、実験仕様の共通化により比較が可能となった、季節 EPS で利用されている大気海洋結合モデルの大気モデル (本項では CPS2-atmos と呼ぶ)<sup>14</sup> の評価結果も参考に示す。本稿では断らない限り、3~5 月を春、6~8 月を夏、9~11 月を秋、12~2 月を冬と表記する。

表 4.4.2 に各モデル及び検証値 (Trenberth et al. 2009; Wild et al. 2013) の全球エネルギー収支一覧を、図 4.4.3 及び図 4.4.4 にそれぞれ夏期及び冬期の地表面下向き短波放射フラックス、大気上端上向き長波放射フラックス、降水量の誤差空間分布を示す。

GSM1403 の特徴として、全球平均では大気上端上向き短波放射、顕熱、大気上端上向き長波放射がいずれも過剰であることが挙げられる。空間分布で見ると、海上を中心に地表面下向き短波放射が過剰である一方で、陸上ではやや過剰であることが分かる。

GSM1603 の全球収支では、地表面下向き短波放射が過剰、大気上端上向き短波放射が過剰となる改悪が見られる。原因として、放射過程における雲のオーバーラップの方法の変更などにより短波放射に対して大気の光学的厚さが薄くなったため、地表面に入射する短波放射が増加したことが挙げられる。GSM1403 で見られた対流活発域での短波放射の入射が過剰な誤差が反転し、逆に過剰となっていることが空間分布から分かる。大気上端上向き短波放射が過剰になったことは、短波放射の入射が過剰になったことと整合的である。

GSM1705 では、放射過程に関する改良として、水雲粒有効半径診断方法の改良や、積雲上昇流域の放射雲診断スキームの導入、エーロゾルの取り扱いの精緻化などが含まれており、GSM1603 で見られた短波放射

<sup>10</sup> Clouds and the Earth's Radiant Energy System

<sup>11</sup> Global Precipitation Climatology Project

<sup>12</sup> Objectively Analyzed air-sea Fluxes

<sup>13</sup> Southampton Oceanography Centre

<sup>14</sup> 2015 年 6 月に現業化された季節アンサンブル予測システム (JMA/MRI-CPS2; Takaya et al. 2018) の大気海洋結合モデルのうち、大気モデル部分のみを用いたモデルを指す。

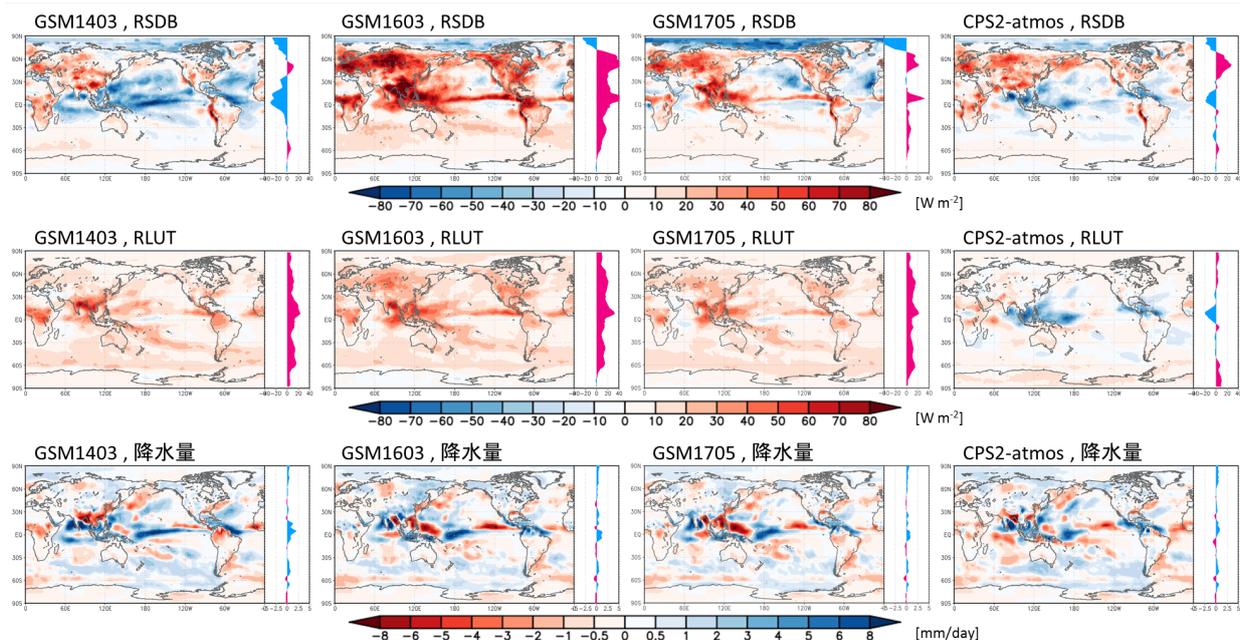


図 4.4.3 COOL 実験における夏 (6~8 月) の 3 か月平均値の誤差の空間分布。上から順に、地表面下向き短波放射フラックス (RSDB) の対 CERES 平均誤差 [ $\text{W m}^{-2}$ ]、大気上端上向き長波放射フラックス (RLUT) の対 CERES 平均誤差 [ $\text{W m}^{-2}$ ]、降水量の対 GPCP 平均誤差 [ $\text{mm/day}$ ] (降水量のみカラーバーの色が逆であることに注意)、左から順に、GSM1403, GSM1603, GSM1705, CPS2-atmos の誤差を示しており、各図の右側には帯状平均した平均誤差を緯度ごとに表示している。いずれも 4 メンバーのアンサンブル平均値である。

表 4.4.2 COOL 実験における各エネルギーフラックスの全球平均値 [ $\text{W m}^{-2}$ ]。GSM1403 以降のバージョンの GSM (GSM1403, GSM1603, GSM1705) 及び CPS2 大気モデル (CPS2-atmos) の 1 年平均値に加えて、検証値として Trenberth et al. (2009) (T09; 2000 年 3 月から 2004 年 5 月の観測値を基にした値) 及び Wild et al. (2013) (W13; 21 世紀初めにおける気候状態での推定値) の値を示している。W13 の値については不確実性の幅を括弧の中に示す。各フラックスの英語表記は図 4.4.2 と同様であり、検証値を除いていずれも 4 メンバーのアンサンブル平均値である。

	RSDT	RSDB	RSUT	RSUB	FLSH	FLLH	RLDB	RLUT	RLUB
GSM1403	341.2	182.1	103.0	24.3	26.5	85.6	340.9	252.0	398.8
GSM1603	341.2	209.9	78.6	24.9	23.3	89.0	335.6	254.6	399.5
GSM1705	341.2	194.9	92.3	24.0	23.3	86.9	336.8	251.5	399.3
CPS2-atmos	341.2	188.6	99.1	25.4	19.1	80.4	338.5	239.1	398.8
T09	341.3	184	101.9	23	17	80	333	238.5	396
W13	340	185	100	24	20	85	342	239	397
	(340-341)	(179-189)	(96-100)	(22-26)	(15-25)	(80-90)	(338-348)	(236-242)	(394-400)

を中心とした誤差が大幅に改善されている。GSM1403 と比べても、総合的に見れば全球エネルギー収支が改善している。GSM1603 で見られた、夏期におけるユーラシア大陸や北米全域、冬期におけるアフリカ大陸や南米大陸の短波放射の入射が過剰な誤差が大幅に改善しており、改良の狙いが明瞭に現れている。

GSM1403 以降の GSM において一貫して見られる特徴として、冬期の南大洋における短波放射の入射が過剰な誤差と、対流活発域を中心とした大気上端上向き長波放射が過剰な誤差が挙げられる。これら二つの誤差は GSM1705 でやや軽減できているが、GSM1403 以降継続して見られている今後解決していくべき課題の

一つである。

次に降水量について述べる。表 4.4.3 に示した全球平均降水量一覧から、GSM1403 以降の全ての GSM で夏冬ともに GPCP よりも降水過多となっているが、モデルを更新するたびに過剰な誤差が小さくなる方向に変化する改善が見られる。夏期 (図 4.4.3 の下段) は、GSM1603 でインド洋の降水過多が改善している一方で、フィリピン東海上における降水過少が顕著となり、その傾向は GSM1705 でも継続している。冬期 (図 4.4.4 の下段) の南米大陸上の降水分布はバージョンごとに大きく変化しており、GSM1603 では地表面下向き短波放射が過剰となった結果として降水量も過多になった

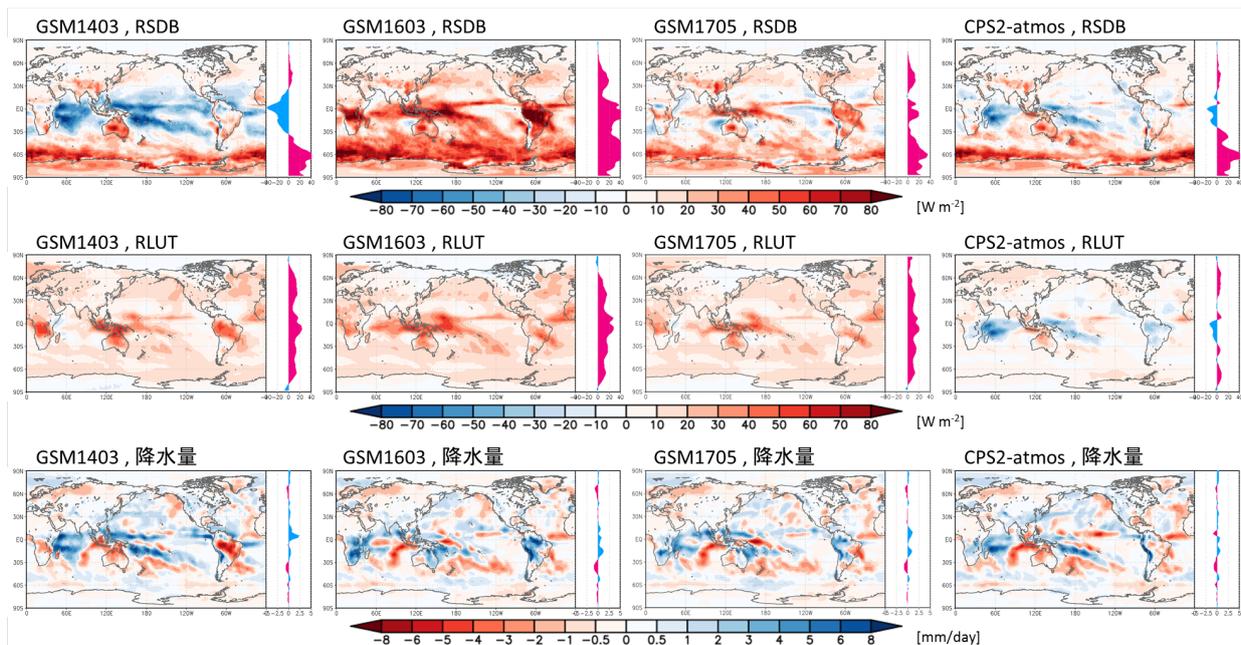


図 4.4.4 図 4.4.3 と同様、ただし冬（12～2月）の3か月平均値の誤差を示す。

表 4.4.3 COOL 実験における3か月平均降水量の全球平均値 [mm/day]。GSM1403 以降のバージョンの GSM (GSM1403, GSM1603, GSM1705) 及び CPS2 大気モデル (CPS2-atmos) に加えて、検証値として GPCP の値を季節ごと（春：3～5月、夏：6～8月、秋：9～11月、冬：12～2月）に示しており、検証値を除いていずれも4メンバーのアンサンブル平均値である。

	春	夏	秋	冬
GSM1403	2.94	3.05	2.97	2.94
GSM1603	2.81	2.92	2.82	2.84
GSM1705	2.76	2.86	2.78	2.80
CPS2-atmos	2.75	2.86	2.75	2.75
GPCP	2.66	2.72	2.69	2.71

と考えられるが、GSM1705 では改善されている。

CPS2-atmos の全球エネルギー収支は GSM と比べると検証値との一致が非常に良く、対流活発域における地表面下向き短波放射及び大気上端上向き長波放射が過剰といった GSM の予測特性とは異なる特徴が見られる。夏期のフィリピン東海上の降水過少はほとんど見られず、日付変更線付近の降水過多も GSM と比べて軽減されている。季節予報に用いる大気海洋結合モデルの開発において、全球エネルギー収支や熱源となる西部太平洋の降水分布などを重要視して大気モデルを開発していることが、この結果から見て取れる。

GSM1705 の放射に関する特性を詳しく述べる。図 4.4.5 に夏期及び冬期における GSM1705 の放射フラックスの対 CERES 誤差空間分布を示す。対流活発域や夏半球の陸上、冬期の南大洋などで短波放射の反射が

弱く、地表面下向き短波放射を反転した誤差が海水上を除いて見られ、雲が不足していると考えられる。海水上に目を向けると、夏期の北極域においては地表面短波放射が下向き（図 4.4.3 の上段）と上向きの両方で過少である一方で、冬期の南極域では地表面の下向き短波放射が過剰であるにもかかわらず上向き短波放射が過少である。このことから、海水面アルベドや大気の光学的厚さなどに課題があることが示唆される。地表面下向き長波放射の誤差から、夏期北極域では雲が多く長波放射が過剰である一方で、その他の領域では雲が少なく長波放射が過少である可能性がある。これらの誤差特性を放射収支のみならず多面的により詳細に調査することで、今後の改善に繋がることが期待される。

ここで放射フラックスについて、予測時間や評価期間などが COOL 実験と異なる性能評価試験及び再予報実験と本項で示した評価結果を比較する<sup>15</sup>。図 4.2.4 に示されている性能評価試験や再予報実験（図略）における大気上端上向き長波放射フラックスは、ベンガル湾からフィリピンの東海上にかけてや、太平洋上の対流活発域など、大気上端上向き長波放射が過剰である誤差特性が類似している。一方で、大気上端上向き短波放射について、性能評価試験、再予報実験、COOL 実験における平均誤差は見え方が異なる点がある。性能評価試験による誤差特性（図 4.2.5）を見ると、夏冬

<sup>15</sup> 性能評価試験は 00, 06, 12, 18UTC 初期値の予測時間 FT=6 について 2017 年 8 月及び 2018 年 1 月の月平均誤差であり、再予報実験は予測 3～30 日目の 28 日平均場の 1981～2010 年の 30 年平均系統誤差であるのに対し、COOL 実験は 2001 年 6～8 月及び 2001 年 12 月～2002 年 2 月の 3 か月平均誤差である。

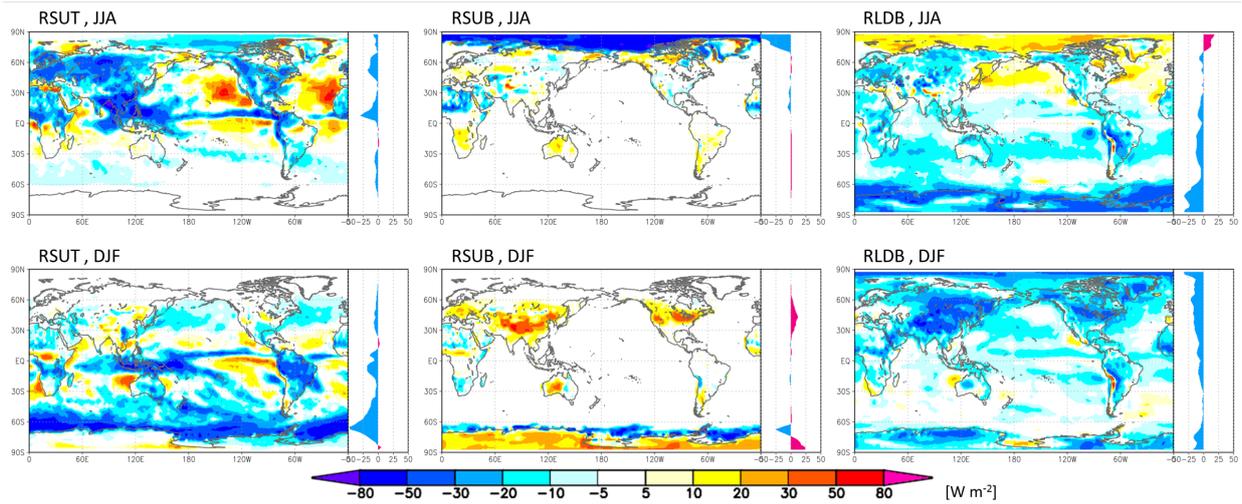


図 4.4.5 GSM1705 の COOL 実験における 3 か月平均値の誤差の空間分布。上から順に、夏 (6~8 月)、冬 (12~2 月)、左から順に、大気上端上向き短波放射フラックス (RSUT)、地表面上向き短波放射フラックス (RSUB)、地表面下向き長波放射フラックス (RLDB) の対 CERES 平均誤差 [ $\text{W m}^{-2}$ ] を示しており、各図の右側には帯状平均した平均誤差を緯度ごとに表示している。いずれも 12 メンバーのアンサンブル平均値である。

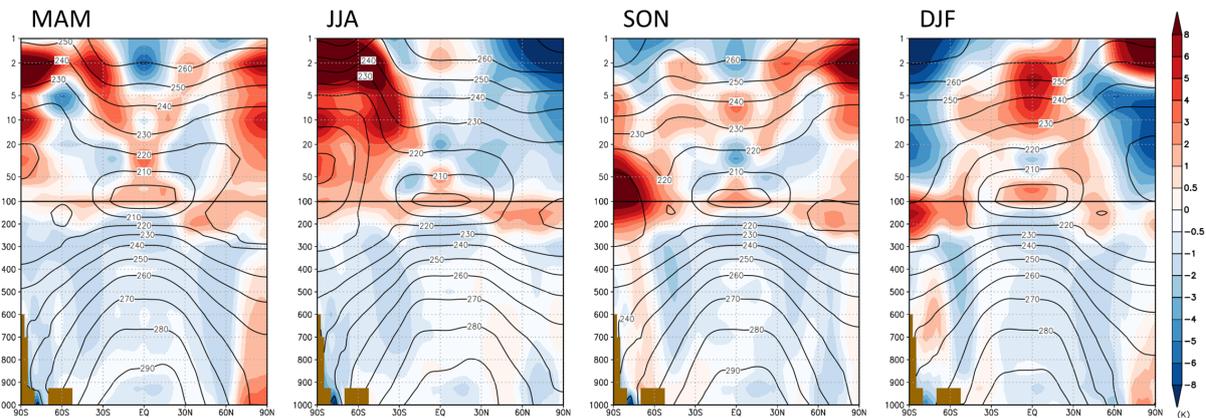


図 4.4.6 GSM1705 の COOL 実験における帯状平均した気温 [K] の 3 か月平均値の高度-緯度断面図。左から順に、春 (3~5 月)、夏 (6~8 月)、秋 (9~11 月)、冬 (12~2 月) の実験結果について、色は対 ERA-Interim 平均誤差を、等値線は予測結果を示しており、縦軸は気圧 [hPa] を表す。黄土色の塗りつぶし域は地上気圧の帯状平均がその高度よりも低い領域を表している。いずれも 12 メンバーのアンサンブル平均値である。

ともに東部太平洋の夏半球側を中心とした上向き短波放射が過剰であるのに対し、再予報実験 (図 4.3.1) や COOL 実験では対流活発域に沿って過少な誤差が見られる。また、ペルーやチリの沿岸といった海洋性層積雲が発生しているとみられる領域において、性能評価試験や再予報実験では顕著な短波放射の反射不足が見られる一方で、COOL 実験では特に夏期にほとんど見られない。全球平均で見ると、上向き短波放射が過少な誤差特性が COOL 実験で最も顕著に見える。性能評価試験、再予報実験、COOL 実験で評価結果が異なる原因について、水平解像度の違いに起因する差や、FT=6 では見えない誤差が長期積分をすることで現れていること、対象年による誤差特性の違いが大気上端上向き長波放射と比べて大きいことなどが可能性として考えられるが、原因は分かっていない。

CERES のプロダクトにはある程度の誤差が含まれていることを念頭に置く必要があり、特に地表面における値は大気上端よりも相対的に不確実性が大きい<sup>16</sup>。また、本項では CPS2-atmos との比較はあまり触れなかったが、GSM と類似の特性や、GSM と比べて改善されている点などを精査することで、今後のエネルギー収支改善の手がかりになるかもしれない。

次に、第 4.2 節で示した GSM の対流圏下層低温バイアスについて、COOL 実験での評価結果を確認する。図 4.4.6 に、GSM1705 の気温の誤差の高度-緯度断面図を季節ごとに示す。

季節に関係なく見られる誤差として、高緯度の一部を除く対流圏の低温や対流圏界面付近の高温傾向があ

<sup>16</sup> <https://ceres.larc.nasa.gov/dqs.php>

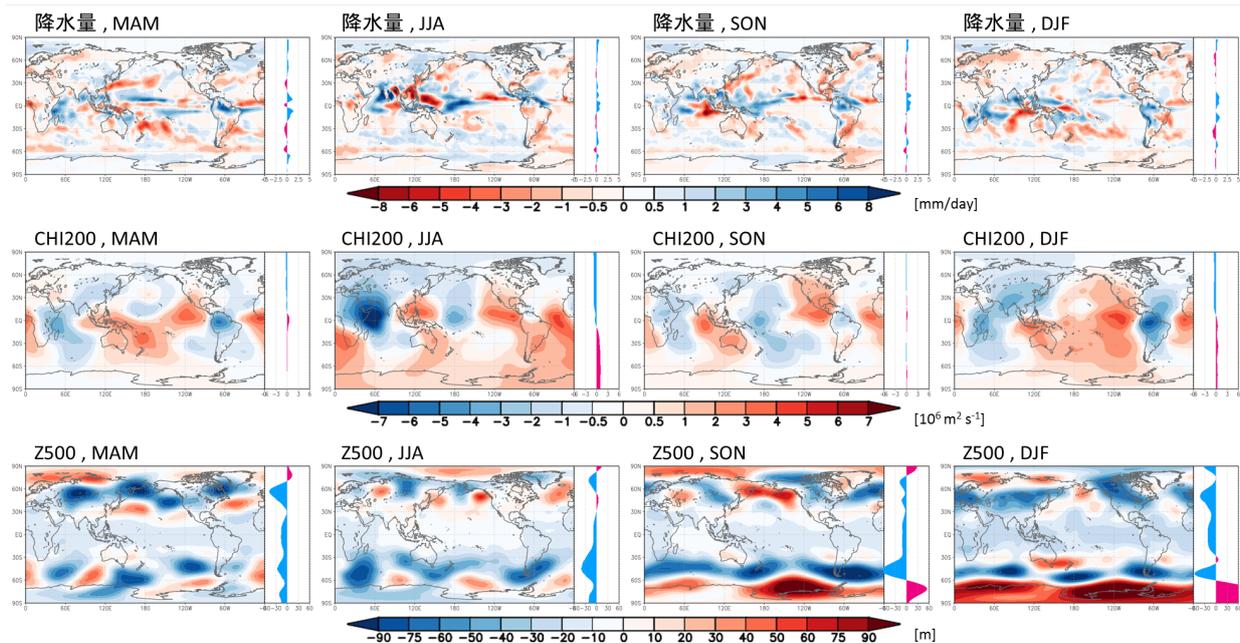


図 4.4.7 GSM1705 の COOL 実験における 3 か月平均値の誤差の空間分布。上から順に、降水量の対 GPCP 平均誤差 [mm/day]、200 hPa 速度ポテンシャルの対 JRA-55 平均誤差 [ $10^6 \text{ m}^2 \text{ s}^{-1}$ ]、500 hPa ジオポテンシャル高度の対 ERA-Interim 平均誤差 [m]、左から順に、春 (3~5 月)、夏 (6~8 月)、秋 (9~11 月)、冬 (12~2 月) の平均誤差を示しており、各図の右側には帯状平均した平均誤差を緯度ごとに表示している。いずれも 12 メンバーのアンサンブル平均値である。

る。夏期及び冬期の気温に関して、性能評価試験及び再予報実験の帯状平均誤差 (図略) と比較すると、対流圏における低温傾向の中でも特に熱帯対流圏上層や冬半球側中緯度対流圏における低温は実験間でほぼ共通している誤差である一方で、高緯度域や対流圏下層の高温傾向は実験間で差異がある。また、COOL 実験の方が対流圏の低温傾向や成層圏の誤差の絶対値が大きいことから、図 4.2.1 で示した初期値における対流圏の低温バイアス傾向が予測が進むとともに拡大し、性能評価試験や再予報実験の予報期間を超えて誤差の拡大が持続している可能性がある。上部成層圏における冬 (夏) 半球側の極域高温 (低温) バイアスなども実験間で程度共通しているが、成層圏は対流圏と比べて誤差の年々変動が大きく検証データ間での差異も相対的に大きいので、複数年による評価が望ましい。

図 4.4.7 に示した GSM1705 の COOL 実験における降水量、200 hPa 速度ポテンシャル (CHI200) 及び 500 hPa ジオポテンシャル高度 (Z500) に関する誤差の空間分布について、再予報実験における系統誤差 (図 4.3.2、図 4.3.3、図 4.3.4) と比較する<sup>17</sup>。降水量について、第 4.3 節で指摘しているインド洋西部の過多バイアスや夏期のフィリピン東海上における過少バイアス、春期から夏期にかけての日本の南海上の過少バイアスは COOL 実験でも概ね一致した誤差傾向が見える。その他にも、アマゾン周辺や夏期の日付変更線付近の降

水過剰なども共通して見られる特徴である。降水量に対応して、大規模収束発散を表す CHI200 の誤差分布も熱帯域の特徴は概ね再予報実験と一致している。

一方で、Z500 に関する誤差傾向は降水量や CHI200 と比べると再予報実験との違いが大きいように見える<sup>18</sup>。低緯度帯のほぼ全域で高度が低い誤差は共通であるのに対し、COOL 実験では中緯度帯では帯状平均で負の誤差傾向となっているものの波列のような誤差パターンが多く見られる。COOL 実験では、再予報実験で年間を通じて見られたアリューシャン列島付近での Z500 が高い誤差が秋を除いて不明瞭である。850 hPa 気温 (T850) についても、中高緯度における誤差傾向は再予報実験と一致しない部分が多く、第 4.3 節で指摘している T850 の夏期シベリア付近の高温バイアスは 2001 年を対象とした COOL 実験では見られない (図略)。

最後に、性能評価試験では評価が難しい成層圏準二年周期振動 (QBO: Quasi-Biennial Oscillation) について述べる<sup>19</sup>。GSM1403 で非地形性重力波を表現するパラメタリゼーションとしてレイリー摩擦に代わって Scinocca (2003) によるスキームが導入されたことにより、モデル気候場や中層大気の子報精度、QBO の表現が改善された (第 3.4.3 項)。そこで、GSM1705 (CNTL) 及び GSM1705 の非地形性重力波スキームをレイリー摩擦に戻した実験 (TEST) における COOL 実験での

<sup>17</sup> 再予報実験では初期時刻を基準に季節を決めるのに対し、COOL 実験は予測対象期間で季節を決めるため、両者で対象月が異なる。

<sup>18</sup> 図 4.3.4 とカラーバーが異なる。

<sup>19</sup> 再予報実験を用いた QBO の評価は金浜 (2017) を参考にしたい。

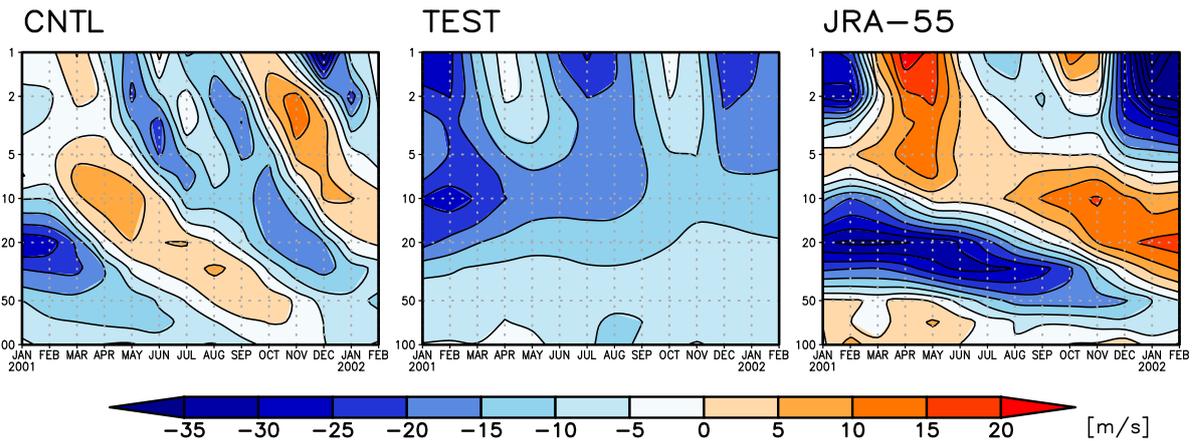


図 4.4.8 5°S–5°N で平均した帯状平均東西風 [m/s] の時間–高度断面図。左から順に、CNTL (GSM1705)、TEST (GSM1705 の非地形性重力波スキームをレイリー摩擦に戻した実験)、JRA-55 であり、縦軸は気圧 [hPa] を表す。TEST 及び CNTL はいずれもある 1 メンバーの結果である。

QBO の表現を確認する。図 4.4.8 に示した熱帯域で平均した東西風の高度–時間断面図を見ると、TEST では表現されていない熱帯下部成層圏における東西風の周期変化が CNTL では見られることから、COOL 実験の結果からも非地形性重力波スキームの改良によって QBO の表現が改善したことが分かる。しかしながら、GSM1705 の QBO の表現には周期が短く振幅が弱いという特性があり、これらの表現の改善は今後の課題である。季節よりも長い時間スケールで変動する現象は複数年にわたる長期積分環境を用いて評価することが望ましいが、QBO に関して言えば COOL を用いても振幅や周期の大まかな評価は可能だと考えられる。解析予報サイクルにおける成層圏の第一推定値と観測値の整合性を確認するために性能評価試験を、東風・西風位相の転換を評価するために再予報実験を用いるなど、目的や評価項目が異なる複数の実験により特性を把握することが重要である。

#### 4.4.6 まとめと今後の展望

COOL は低解像度で AMIP 型の実験を行う 1 年積分共通評価ツールであり、開発において性能評価試験や再予報実験とは異なる役割を果たす基礎評価試験として利用されている。気象庁の各部署の全球数値予報モデル開発者が、共通の仕様で実験及び評価・検証を行うことで、開発効率の向上や、成果や議論の共有を促すことが期待される。再予報実験と同様に夏冬以外の季節に関する評価が可能となる点は、通常性能評価試験にはない利点の一つである。手軽に実行できるツールとするために計算機資源や実行時間を抑えることを優先しているため、実験対象年が 1 年である点や少数のアンサンブルメンバーで評価している点で評価結果の解釈に注意が必要である。

例えば放射フラックスや降水分布など、第 4.2 節や

第 4.3 節で示した性能評価試験や再予報実験における評価と類似の誤差特性が見える要素も多く、COOL は基礎的な評価に有用だろう。一方で、中高緯度における Z500 や T850 など、特に内部変動の大きい要素・領域では実験対象年を変更することで誤差特性が大きく変わり得るため、基礎評価試験の段階で COOL のみを用いて評価・検証を行うことは誤った予測特性の評価に繋がりがかねない。多面的な評価・検証を行うために今後も評価検証システムの拡充を進めていく一方で、目的や用途に応じて様々な実験による評価を組み合わせることで開発を進めていき、科学的な議論を通じた全球数値予報モデルの精度向上に努めていきたい。

#### 参考文献

- Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balsameda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kállberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597.
- Gates, W. L., 1992: AMIP: the Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **73**, 1962–1970.
- Gates, W. L., J. S. Boyle, C. Covey, C. G. Dease, C. M. Doutriaux, R. S. Drach, M. Fiorino, P. J. Gleckler, J. J. Hnilo, S. M. Marlais, T. J. Phillips,

- G. L. Potter, B. D. Santer, K. R. Sperber, K. E. Taylor, and D. N. Williams, 1999: An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55.
- 原旅人, 2013: 概論. 数値予報課報告・別冊第 59 号, 気象庁予報部, 1–5.
- 石田純一, 2017: 数値予報システム開発のプロセス. 数値予報課報告・別冊第 63 号, 気象庁予報部, 4–10.
- Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto, 2005: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe Collection. *Int. J. Climatol.*, **25**, 865–879.
- 金浜貴史, 2017: QBO、成層圏突然昇温. 平成 28 年度季節予報研修テキスト, 気象庁地球環境・海洋部, 45–46.
- Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48.
- 栗原幸雄, 桜井敏之, 倉賀野連, 2006: 衛星マイクロ波放射計、衛星赤外放射計及び現場観測データを用いた全球日別海面水温解析. 測候時報, **73**, S1–S18.
- Matsumoto, T., M. Ishii, Y. Fukuda, and S. Hirahara, 2006: Sea ice data derived from Microwave radiometer for climate monitoring. *AMS 14th Conference on Satellite Meteorology and Oceanography*, P2.21.
- Scinocca, J. F., 2003: An Accurate Spectral Nonorographic Gravity Wave Drag Parameterization for General Circulation Models. *J. Atmos. Sci.*, **60**, 667–682.
- Takaya, Y., S. Hirahara, T. Yasuda, S. Matsueda, T. Toyoda, Y. Fujii, H. Sugimoto, C. Matsukawa, I. Ishikawa, H. Mori, R. Nagasawa, Y. Kubo, N. Adachi, G. Yamanaka, T. Kuragano, A. Shimpo, S. Maeda, and T. Ose, 2018: Japan Meteorological Agency/Meteorological Research Institute–Coupled Prediction System version 2 (JMA/MRI-CPS2): atmosphere–land–ocean–sea ice coupled prediction system for operational seasonal forecasting. *Clim. Dyn.*, **50**, 751–765.
- Trenberth, K. E., J. T. Fasullo, and J. Kiehl, 2009: Earth’s global energy budget. *Bull. Amer. Meteor. Soc.*, **90**, 311–323.
- Wild, M., D. Folini, C. Schär, N. Loeb, E. G. Dutton, and G. König-Langlo, 2013: The global energy balance from a surface perspective. *Clim. Dyn.*, **40**, 3107–3134.
- 米原仁, 2017: 活用例 (1)-全球モデル. 数値予報課報告・別冊第 63 号, 気象庁予報部, 29–49.