

第1章 序論

1.1 はじめに¹

1.1.1 数値予報モデル開発を支援する基盤環境の整備の必要性

数年前から、数値予報課における数値予報モデル開発²は、単なるスキームの取り替えやパラメータチューニングを中心としたものから、モデル予測の誤差の原因やモデルの挙動を既知の事実や知見を論理的に積み上げることによって理解し、その理解に基づいてモデルの修正を試みる科学的なアプローチを重視するように変化しつつある。数値予報課報告・別冊第58号(気象庁予報部 2012)や第59号(気象庁予報部 2013)では科学的なアプローチによるモデルの問題発見とその解決に向けた手法についてまとめ、その後の開発をより科学的なものにする一つのきっかけとなった。

科学的なアプローチによる開発では、事実や知見を論理的に組み合わせる新たな結論を導き出し、その結論を別の知見などとさらに組み合わせることを繰り返し進める。そのため、科学的なアプローチによるモデル開発を行うためには、論理的な思考力を強化するとともに、現象の動態を把握すること、モデルの予測データを検証すること、現象とモデル予測に関する物理的考察を行うこと、気象学の知見を習得すること、また最新の研究動向を把握することなどを通じて、開発に利用できる事実や知見を増やす努力が日々必要で、大きな労力を要する。その上、モデルが高度化、複雑化している中で、モデルのさらなる精度向上を目指すためには、より高度な科学的知見が求められる。

一方、数値予報システムは多くのプログラムから構成されており、プログラム間で必要なデータをもれなく受け渡したり、依存関係を考慮したりしながらプログラムを実行し、その終了監視をして次のプログラムを実行するといったジョブ制御が必要となる。また、出力結果の検証、可視化のために膨大な出力データを取り扱う必要もある。計算機の計算能力の向上とともに、数値予報システムで扱うデータの量や種類は入力、出力ともに膨大になってきており、その扱いには高度な技術が要求されるようになってきている。また、モデルの問題点を見出すために、様々な観点で可視化を行うには、出力データが大きくなる中で、そのデータをより短時間で効率よく描画できる技術や、問題点を浮かび上がらせることができる表現方法の考案とその実装などが必要となっている。

従前の数値予報課では、上で述べた2つのスキル、すなわち、モデル開発に必要な気象学や物理学の知見

などのサイエンシ的なスキルと、実験システムの構築、データの可視化や検証のためのデータのハンドリング、ソースコードの最適化、並列化などのエンジニアリング的なスキルの両方がモデル開発者には求められた。しかし、近年、それぞれが高度化、複雑化する中で、両方の性質の異なるスキルを同時に求めることが難しくなっている。その結果、モデル開発者のスキルが中途半端な状態となり、データの可視化や実験システムの構築に時間が取られ、モデル改良のためのサイエンシ的な考察や思考に十分な時間を費やすことができず、効率のよいモデル開発を阻害している面があった。

原・高谷(2013)で紹介したように、欧州中期予報センター(ECMWF)や英国気象局(UKMO)では、開発環境を組織全体でできるだけ統一した上で、モデル開発にかかわる基盤環境の整備を専門に行う部署があり、モデル開発者はその基盤環境を利用して、科学的な思考に集中できる環境が整えられている。当庁においてもこのような部門の必要性が認識されて、実験システムの構築や可視化環境の整備といったモデル開発を支援するための基盤環境構築を担う基盤整備グループが、2011年に数値予報課数値予報班の一つのグループとして発足した。第3章で紹介する数値解析予報実験システム(NAPEX)は、以前は数値予報モデルの開発者が自ら構築して、スーパーコンピュータの更新のたびに新しい大きな変更が加えられてきた。第9世代スーパーコンピュータシステムの運用開始とともに導入されたNAPEXは、スーパーコンピュータシステムで実行する数値予報ルーチンの管理システムを構築した数値予報課プログラム班が構築に参加し、システム面の専門性が活かされた実験システムとなっている。

本報告では、このようなモデル開発環境の変化を踏まえ、モデル開発を支援するための基盤環境とその整備をテーマの一つとする。

1.1.2 開発管理

基盤整備の動きとともにモデル開発の効率を高めることを目的に進められたのが、開発管理のための環境構築とその実行である。

すでに室井(2013)で詳細に論じられているように、数値予報モデル開発が必然的に大規模かつ横断的なものになったことから、十分な議論を経て開発計画を作成の上、その工程管理や情報・成果の共有、ソースコードのバージョン管理を行うこと、すなわち「開発管理」が必要となった。開発管理は、開発の進捗状況を管理者の立場から把握することが主目的ではなく、どのような開発が進行しているのかを開発者間で把握しやすくすること、そして、開発の際に議論したこと、考えたこと、作業などを記録して、現在の開発者はもちろんのこと、未来の開発者への説明責任を果たせるように

¹ 原 旅人

² 本報告では、便宜上、解析システムなど、数値予報モデル以外の数値予報システムを構成する部分の開発を含めて「数値予報モデル開発」または「モデル開発」と呼ぶことにする。

することが大きな目的である。近年のモデル開発においては、組織横断的な開発が行われることも多くなり、開発者のコミュニティが大きくなりつつあるが、その結果、他の組織の開発の進捗は把握しにくい場合がある。また、モデル開発は継続的に行われるものであるが、その開発を未来永劫にわたって同じ開発者が行うことはなく、開発者の入れ替わりがある。そのような中で、開発者が替わったことで過去の知見が継承されないようなことがあれば、新規の開発者のスムーズな参入を妨げるとともに、すでに分かっている知見を再度、新しい開発者が自ら見出すために時間を費やすことになるなど、開発が非効率になる。このように、成果のまとめだけでなく、開発の記録を残しておくことは、継続的なモデル開発を実施するために必須のことである。

開発管理のための基盤になるシステムが第2.2節、第2.3節でそれぞれ紹介するプロジェクト管理システムとバージョン管理システムである。

バージョン管理システムは、ソースコードをはじめとするファイルの変更履歴を管理することが主たる役割の一つである。バージョン管理システムを利用する際には、ユーザがリポジトリと呼ばれるファイルの保管庫に変更内容を登録することで、ファイルに加えた変更がいつ、誰によって、どのような内容で行われたのか(変更差分)をそれに対する変更者のコメント(ログ)とともに記録する。また、変更記録に基づいて、最新版ではなく任意の時刻のファイルの状態を取り出すことも可能である。

バージョン管理システムを用いることで、ソースコードそのものの変更履歴を追跡できるようになる。しかし、その意図や背景、テストの結果などはバージョン管理システムに残された記録だけを見てもわからないことが多い。すでに述べたように、バージョン管理システムには変更者のコメント(ログ)を記録することができるが、記入できるのはテキストだけであり、図を用いるなどした詳細な説明をログに書き込むことは困難である。

さらに、モデル開発はソースコードの修正だけにとどまらない。実験や日々の予測結果から気がついた問題点の記録とその調査、実験結果の検証とその考察、それらに対する開発者間での議論もモデル開発における重要な過程の一つである。また、実験の結果、現業化に至った開発事項については、報告や刊行物などにその成果がまとめられることもあるが、結果が期待したものではなかった実験の情報もその後の開発においては重要な情報になりうる。しかし、従来、これらの情報は、そのときの開発者間のみで共有されることが多く、それ以外に共有されたり、将来の開発者のために記録として残されることが少なかった。また、添付資料なども開発者個人の保管場所に保存することが多かったために、人事異動とともに失われることも多かつ

た。このような状況は、時間が経過してから、当時の結果や議論を追跡することには困難を伴い、過去の開発者の知見も活かした効率的なモデル開発を阻害する一つの要因でもあった。

そのような背景のもと、開発過程の記録を残すことを目的に導入したのがプロジェクト管理システムである。プロジェクト管理システムは、ソースコードの修正はもちろんのこと、モデル開発におけるさまざまな過程を記録し、また、それを一覧にすることで進捗の把握にも利用できる。バージョン管理システムと併用することで、ソースコードの変更とその詳細な説明を一体として扱うことができる。また、作業にとりかかる際には、やるべき作業のリストアップにも用いることができる。このように作業過程、調査資料、議論を一か所に集約しておくことで、将来の開発者がそれらを見直すことも容易となる。その記録は庁内からは制限なく閲覧できるようになっているので、他の部署のモデルの開発状況を把握したり、それを基に共同開発につなげやすくなった。

現在、気象庁におけるモデル開発では、バージョン管理システム、およびプロジェクト管理システムを利用した開発管理が推進されている。こうした開発管理の実際を報告するのが、本報告のもう一つのテーマである。

本報告の2つのテーマである基盤整備と開発管理の強化は、冒頭で述べた科学的なモデル開発の推進に大きく寄与している。すでに述べたように、科学的なモデル開発のためには既知の事実や知見を論理的に組み合わせることが必要であるが、その論理的考察に用いた事実や知見、結論を導き出した論理を開発管理ツールに記述することで、開発記録だけにとどまらず、論理の確認、開発者間での共有にも活用している。また、NAPEX や汎用的な描画ツールの整備により、科学的な思考に費やす時間をより確保できるようになった。科学的な開発姿勢と、それを支える基盤整備と開発管理の強化の集大成が2016年の全球モデルの改良(米原 2016)や、次世代非静力学モデルの開発(気象庁予報部 2014)と言えよう。

1.1.3 本報告の内容・構成

本報告では、近年、気象庁のモデル開発で推進されてきた基盤整備、開発管理に焦点をあてて、その背景とこれまでの取り組みについてまとめる。

本節では、これまでに基盤整備や開発管理の必要性とその背景を説明してきた。開発管理の強化は数値予報モデル開発のプロセスの明確化とともに進められたものである。そこで、本章の後半では、数値予報課における開発プロセス、およびそのプロセスにおける開発管理の活用について論じる。また、最近の英国気象局における開発や現業化の判断などのプロセスの最新事情について、最近まで英国気象局に派遣されていた

モデル開発者から報告する。

第2章では開発管理について取り上げる。気象庁における開発管理に対する取り組みを概観したのちに、そのためのシステムであるプロジェクト管理システムやバージョン管理システムの一般的な事項について簡単に説明する。その説明を踏まえて、数値予報課における開発管理ツールの活用について紹介する。加えて、検証およびそれを実施するツールのあり方についても触れる。検証ツールは数値予報システムの開発において行うべき評価を具体化したものであり、開発経験から得られた科学的知見が集積されていくものである。そのため、数値予報システムだけではなく検証ツールも開発管理の対象に含むことは重要である。

第3章では気象庁の数値予報モデル開発の重要な基盤の一つであるNAPEXについて述べる。NAPEXの必要性とその歴史を概観したのち、現在のNAPEXの設計思想と実装、利用について紹介する。

第4章では数値予報モデル開発で必要かつ重要な基盤技術であるデータハンドリングと可視化について取り上げる。気象庁の数値予報システムで用いられているデータフォーマット、数値予報課で用いられている可視化ツールについて概説する。

本報告の発行を通じて、開発基盤整備の重要性や開発管理への理解の深まりとそのさらなる普及を期待したい。

参考文献

- 原旅人, 高谷祐平, 2013: 海外数値予報センターの開発管理の例. 数値予報課報告・別冊第59号, 気象庁予報部, 195-199.
- 気象庁予報部, 2012: 物理過程の改善にむけて (I). 数値予報課報告・別冊第58号, 気象庁予報部.
- 気象庁予報部, 2013: 物理過程の改善にむけて (II). 数値予報課報告・別冊第59号, 気象庁予報部.
- 気象庁予報部, 2014: 次世代非静力学モデル asuca. 数値予報課報告・別冊第60号, 気象庁予報部.
- 室井ちあし, 2013: 開発管理の必要性. 数値予報課報告・別冊第59号, 気象庁予報部, 192-194.
- 米原仁, 2016: 全球数値予報システムの物理過程改良の概要. 平成28年度数値予報研修テキスト, 気象庁予報部, 1-3.

1.2 数値予報システム開発のプロセス¹

1.2.1 はじめに

数値予報は、国内外の観測データの収集に始まり、次にそのデータの品質管理（誤差の大きいデータの除去や補正）を行い、観測値と第一推定値を用いてデータ同化システムにより初期値を作成する。そして、初期値から数値予報モデルを用いて将来の大気状態を予測し、その結果は統計的手法を用いて補正される（ガイダンス）。これら一連の処理はスーパーコンピュータシステム等で自動で実行され、そのための数多くのプログラムが用いられる（室井 2012）。

本節では上記の流れの中で観測データの品質管理から数値予報モデルによる予測までを数値予報システムと呼ぶこととし、その開発プロセスについて述べる。数値予報システムは気象学、物理学、計算機科学、数値流体力学、数学、統計学等の様々な科学分野の知見が必要となる。そのため、多くの開発者の知見を集める必要がある。また、室井（2013）に述べられているように、近年、数値予報システムの目的が多様化し、開発は必然的に大規模かつ横断的なものになっている。そのため、開発管理が重要となり、その中の一つである開発プロセスの策定が重要となっている。

気象庁で現業運用される数値予報システムは気象業務の根幹を支えるものであり、まず第一に定められた時間内に処理を終え、障害を起こさない（万一の場合は速やかな復旧や代替手段を取る）ことが重要である。その上で、予測精度の向上を図る必要がある。

そこで、本節ではこれまで数値予報課で検討してきた数値予報システムの開発プロセスについて述べる。まず、予測精度向上のためにはそれを阻害する誤差要因を把握する必要があり、伝統的に認識されてきた誤差要因を第 1.2.2 項 に簡単にまとめた。そして、これまでにあまり認識されてこなかった誤差要因を含めて今後の開発で検討すべき事項について第 1.2.3 項 でまとめた。第 1.2.4 項 ではこれらの検討すべき事項を踏まえて実際の開発の進め方（フロー）についてまとめた。

1.2.2 数値予報システムの誤差要因

精度向上のためには、誤差を軽減させる必要がある。誤差要因には様々なものがあるが、伝統的に認識されてきた主たるものだけでも以下が挙げられる。

- 初期条件の誤差
 - － 観測の時間・空間分解能に起因する誤差
 - － 観測機器の持つ誤差
 - － 観測の代表性に起因する誤差
 - － 品質管理手法に起因する誤差
 - － データ同化手法に起因する誤差（データ同化システムの分解能による制限や観測演算子に起因するものを含む）
- 数値予報モデルの誤差

- － 空間・時間分解能の制限による誤差（離散化手法に起因する誤差を含む）
- － 物理過程の不確実性に起因する誤差
- － 数値予報モデルが考慮しない大気中の素過程による誤差
- 境界条件の誤差
 - － 地形の表現の制限に起因する誤差
 - － 地表面状態の不確実性に起因する誤差
 - － 海面・海氷状態の不確実性に起因する誤差
 - － 人為的な上部境界条件に起因する誤差
 - － 領域モデルにおける側面境界条件に起因する誤差

有限の数値演算による数値予報システムではこれらの誤差（及びここに記載しない要因による誤差）を全てゼロにすることができないものの、全ての誤差を可能な限り小さくしていくことにより予測精度向上が期待できる。

効率的な開発のためにはこれまでの開発において誤差がある程度減少してきた要因に対する開発よりも相対的に大きな誤差要因に対する開発により注力することが望ましい²。しかし、どの誤差要因が大きいか特定することは困難であり、それを発見するための開発も必要である。その際には、伝統的には認識されていなかった新たな誤差要因が見つかることもある。次項では、ここで述べた誤差要因を念頭において、今後の開発において考慮すべき事項について述べる。

1.2.3 数値予報システムの開発において今後考慮すべき事項

現業数値予報システムの開発は気象庁において 50 年以上の歴史があり、その開発の歴史の中で様々な誤差が減少している。しかし、全ての誤差が一律に減少してきたわけではないため、時代によって相対的に大きな誤差を生じる要因は異なっている。また、これまでの多くの先人の努力により現業数値予報システムが成熟してきた現状では、これまでは想定しなかったことが誤差の要因となることや相対的に小さかった誤差が大きくなることもあるだろう。

数値予報課では第 1.2.2 項 で述べた問題意識から、数値予報課における開発を通じて得られた教訓や海外の数値予報センターの動向を基に、今後の開発で着目すべき事項について議論を行ってきた。現時点における議論を通じて得られた知見を以下の通りまとめた。

(1) 科学的根拠と統計的根拠

現業数値予報システムの開発において真の予測精度向上を担保するためには、科学的根拠が重要なことはいうまでもない。問題は、どのように真の予測精度向上を担保するかである。ここで、真の予測精度向上と

² ただし、それぞれの誤差を減らすための開発は、開発者の人材育成までを考慮すると数年から十数年は必要となるため、現時点では相対的に小さな誤差要因に対しても継続的に開発を行っていく必要があることにも留意しなくてはならない。

¹ 石田 純一

は、開発成果が実用化された後の事例（すなわち、未来の事例）に対する精度向上を指す。

後述の開発フロー（図 1.2.1）にあるように基礎開発の段階で様々な手法により科学的根拠を積み上げていく。その後、変更の導入前後で比較実験を行い、統計検証を通じて検証スコアが向上する結果が得られたことも真の予測精度向上に対する統計的根拠の一つとなりうる。ここで、事前の基礎開発を通じて積み上げた根拠に対して比較実験により得られた統計的根拠をどのように扱うべきか議論の余地がある。まず、比較実験は過去のデータを用いて実施するより他はないが、現業数値予報システムの改善で求められるのは前述の通り、未来における予測精度向上であり、本質的な制約がある。例えば、パラメータ等について実験期間に過剰に合わせることで、実験期間では精度が向上してもそれ以外の期間ではむしろ精度が悪化するというオーバーフィッティングの問題が生じる懸念がある。また、これまでの開発により数値予報システムの誤差は非常に小さくなってきたため、真の改善の影響（シグナル）は予測精度の変動（ノイズ）と比べて小さくなってきた（即ち、S/N 比が低くなった）。このような問題意識に対して、Geer (2016) による欧州中期予報センターにおける検討結果があり、以下に簡単に抜粋して紹介する。

- 数値予報の開発で典型的な 0.5% のスコア向上を信頼できる精度で検証するには半年程度のサイクル実験が必要だが計算機資源的にどの予報センターでも困難。
- そのため、伝統的なスコアは予測における科学的な開発の質の指針に常に用いることができない。
- 比較実験前の基礎開発のみを根拠とする（即ち比較実験を行わない）ことはバグの可能性や基礎開発の科学的根拠の不完全性を考えると不適切。
- 従って、より少ない実験設定で多くのサンプルによる実験を行うことが最適であろう。
- 様々な変更のそれぞれに対して異なる実験を行い、予測精度（スコア）により良い変更を選択する戦略は無駄が多い。
- 実験設定の選定においては長期間の実験に先立って個々の変更を注意深く評価し、良い変更に対してフルの実験を行うのが良い。

即ち、従前と比して比較実験を通じて統計的根拠を得るためには個々の変更に対して事前の基礎開発の段階での科学的根拠を重視する必要がある。基礎開発の段階で妥当と考えられる変更をまとめた（パッケージ化した）比較実験が必要である。スコアの良し悪しはその大小で判断できる³ であろう。しかし、基礎開発の段階での科学的根拠の妥当性は自明でないこともありうることから、議論を積み重ねる必要がある。

³ ただし、複数のスコアがある場合に、その優先度を決定することは困難が伴う。

今後の開発においてさらに予測誤差が減少する（さらに S/N 比が低くなる）ことを期待すれば、益々このような観点での開発が重要となると考えられる。

(2) compensating errors を考慮することの重要性

compensating errors とは堀田・原 (2012) に述べられている通り、過程 A の持つ誤差と過程 B の持つ誤差の符号が逆で、合計の誤差が小さくなることを意味する。このことは、合計の誤差が小さく見えている場合に、仮に過程 A の誤差が小さくなるような修正を施した場合に、合計の誤差がかえって増大することを意味する。従って、ボトムアップ・アプローチ (堀田・原 2012) により個々の過程の問題点を見つけ出して修正することは精度向上の必要条件であっても十分条件ではなく、これに加えて別の検討が必要であることを意味する。

このことは数値予報モデルの個々の物理過程の間で生じうるだけでなく、物理過程と境界条件の間、あるいはデータ同化サイクルにおいて、数値予報モデルとデータ同化システム（観測データの品質管理を含む）の間等でも生じうる。

過去においても、compensating errors は存在していたと考えられるが、数値予報システムが成熟し、合計の誤差が小さく見えることが多くなってきたことから、近年になってこの誤差が顕在化してきたと考えられる。

(3) 実装の細部まで検討することの重要性

数値予報システムの開発では最終的に解くべき方程式系（多くの場合は時間・空間方向に連続な偏微分方程式系となる）を導出する必要がある。例えば、サブグリッドスケール現象のパラメタリゼーションではサブグリッドスケールの現象がそれより大きいスケールの現象に与える影響をモデル化し、それに沿って方程式系が決まる。この方程式系の導出段階では如何にして適切なモデル化を行うかが重要であると共に計算機資源等の制約を念頭において何らかの近似を用いることも検討する必要がある。続いて、この方程式系を離散化し、最終的には Fortran 等のプログラムコードに落とし込んでいく。この離散化やプログラムコードへの変換においては多くの任意性がある。また、方程式系では単なる定数として表されるものが実際には限られた観測事実から得られた幅を持つ値しか得られない場合であっても、数値予報システムに実装する際には一意に定める必要がある。このような例として、物理過程の様々なパラメータ、観測データの品質管理における間引き手法や間引き間隔、初期値や境界値等の作成時の水平・鉛直内挿手法等多くのものが挙げられる。さらに、実装においては処理順序やループ範囲等まで一意に定めなくてはならない。

従前はパラメタリゼーションにおけるモデル化や方程式系の妥当性が予測誤差に大きな影響を与えると認識の下、これらについての検討に多くの注意が払われてきた。この検討の重要性は引き続き残るものの、

これまでにあまり検討されなかった実装の細部におけるわずかな違いが予測誤差に与える影響が大きいことがありうるということが認識されてきた。このことを指して、「Minor treatment の重要性⁴」と呼ぶこともある。

これも、数値予報の開発の黎明期ではシステム全体の誤差が大きく、例えば、モデルで考慮されていなかった過程を追加する、といった開発により誤差が減少してきたところ、近年ではシステム全体の誤差の減少に伴い、実装の細部に至るまで検討することが重要となってきたものである。

注意すべき事項として、論文等においてはモデル化や定式化について触れられていても紙幅の制限から実装の細部まで触れられていないことがある。しかしながら、実装の細部の重要性に鑑みて、何らかの形で検討結果を記録する必要がある。このために第 2.2 節で述べられているプロジェクト管理システムを重要なシステムとして位置づけている。

(4) スケール分離の認識の重要性

一般に格子間隔の 5-8 倍のスケールの現象は数値予報モデルでは陽に表現され、格子間隔より小さいスケールの現象はサブグリッドスケールとしてパラメタリゼーションで表現する。そして、その間に位置する格子間隔と同程度から 5 倍程度のスケールは陽に表現することもサブグリッドスケールとしてパラメタライズすることもできないため、新たな扱いが必要となり、これは Grey Zone と呼ばれる。Grey Zone については原 (2012a) で触れられており、境界層、対流さらには地形性重力波についての検討 (Vosper et al. 2016) がなされている。この記述で分かるように、どの現象が Grey Zone と呼ばれるかは数値予報モデルの格子間隔によって異なってくる。すなわち、数値予報モデルの格子間隔を小さくして分解能に起因する誤差を減少させようとする場合は、これまでにパラメタリゼーションとして扱ってきた現象が部分的に陽に表現されるようになり、新たな扱いを検討する必要がある (当然、この場合も第 1.2.3 項 (3) で述べた通り、実装の細部に至るまで検討が必要である)。

(5) 不確実性の大きいパラメータ・予報変数を極力増やさないこと

数値予報システムの開発はこれまでの先人の開発の上に成り立っており、多くの過程が精緻化・高度化されてきた。このことにより、数値予報モデルにおいては特に精度に大きな影響を与えうる大気中の多くの現象がパラメタライズされて取り込まれるようになってきた。即ち、前述の「数値予報モデルが考慮しない大気中の素過程による誤差」や「物理過程の不確実性に起因する誤差」は大きく減少したことを意味する。

そこで、既存の過程の精緻化・高度化に取り組むこ

ととなるが、不確実性の大きいパラメータを増やすと精度向上につながるどころか精度が悪化することがありうることに注意する必要がある。パラメータを追加した場合には何らかの観測値 (あるいは観測値や他の予報変数による関数) でその値を与える必要がある。予報変数を追加する場合には、その時間変化を記述する方程式 (予報方程式) を追加する必要がある。しかし、第 1.2.3 項 (3) で述べた通り、観測値が限られた観測事実から得られた幅を持つ (不確実性を持つ) 値しか得られないことがあり、予測誤差に与える影響が大きいことがある。また、予報方程式は適切な離散化を行わないとノイズが生じて予測精度を悪化させる可能性もある。

不確実性を持つパラメータを決定する際にはパラメータをわずかに変更した複数の比較実験を行うことがある。しかし、これは前述の Geer (2016) が報告した無駄が多い戦略となりうる。特に不確実性の大きいパラメータを多く導入すればするほど組み合わせの数は膨大となる。このような戦略では、多くの開発期間を要することになるか、少数の事例による比較実験に基づいたオーバーフィッティングの問題が生じうる⁵。また、仮にこれらの問題を解決したとしても複数の不確実性の大きいパラメータによる compensating errors の懸念を払拭することは容易ではない。

このことは時としてパラメータや予報変数を減らすという仕様のダウングレードが精度向上をもたらすという逆説的なことがありうることを意味する。事例として、原 (2015) で示されたメソモデルにおける境界層過程の改良が挙げられる。これは、メソモデルにおける境界層過程を Mellor-Yamada-Nakanishi-Niino (MYNN) レベル 3 モデルから MYNN レベル 2.5 モデルに変更したものである。これは名前からも分かる通り、仕様としてはダウングレード (実際にも予報変数を減らしている) であるが、精度向上につながっている。この問題はパラメータの不確実性ではなく予報方程式の解法に起因していたものであるが、闇雲にパラメータや予報変数を増やすことが精度向上につながらないことがありうる例であろう。

精緻化・高度化に向けた開発においては、極力不確実性の小さい物理量を用いてモデル化・方程式系の導出を行っていくこと、不確実性の大きい物理量を導入せざるを得ない場合はいかにして不確実性を減らすかを (非常に困難であるが) 検討することが重要である。また、時には精緻化・高度化ではなく新しい知見を活かして既存のパラメータの不確実性を減らすことに注力することが良いこともあるだろう。

⁵ この最たる例がいわゆる champion case によるパラメータの推定である。champion case により得られたパラメータの組み合わせが普遍的でないことは直感的にも理解しやすいが、一つの事例でなく複数の事例であっても少数である場合は問題が生じうることに注意が必要である。

⁴ http://www.mpimet.mpg.de/fileadmin/atmosphaere/WCRP_Grand_Challenge_Workshop/presentations/GC.Kawai.pdf

(6) 分解能向上のための開発期間

数値予報システムの分解能向上への社会的要請は大きい。これは、より詳細な分布の情報が欲しいという要望もあれば、分解能の制限によりこれまで表現できなかった現象をも予測対象にして欲しいという要望もある。しかし、分解能の制限による誤差は、これまでの数値予報システムの歴史において、大きく改善された誤差要因であり、誤差要因の観点では相対的に誤差の大きさは低下してきていると考えられる。

すなわち、社会的要請を満たしつつ効率的に誤差を軽減させなくてはならない。分解能を向上させるためには、スーパーコンピュータシステムの更新等により得られた計算機能力の向上を利用することとなる。単純に分解能を向上させるだけであれば容易であるが、これまでに述べた事項を念頭におくと精度向上を得るには、相当の開発期間を要する可能性がある。新しい計算機での分解能の仕様が決まってから導入までの5、6年の期間を持つ開発計画として位置づけて開発を行うことが重要であろう。

1.2.4 数値予報課における数値予報システムの開発プロセス

第 1.2.3 項で述べた考慮すべき事項を踏まえて検討した、現在の開発のフローを図 1.2.1 に示す。

開発のフローは数値予報課において「ルーチン変更のガイドライン」(以下、「ガイドライン」という)としてまとめてきた。それに基づいて、数値予報課と気象研究所との連携を促進するため、2015 年度に気象庁技術開発推進本部モデル技術開発部会メソグループにおいてメソ・局地モデルの開発を対象として議論したものを同じく 2015 年度に開催された気象庁予報部・観測部と気象研究所の研究懇談会で数値予報システムの開発全体を対象を広げた上で懇談したものである。

数値予報課における数値予報システムの開発を担当する数値予報班には複数のグループ(全球・台風グループ、メソモデルグループ、観測データ処理グループ、基盤整備グループ)がある。グループの中にもいくつかの開発コミュニティがあり、取り扱う開発課題や開発コミュニティの規模が異なっている。さらに、開発課題により適切な開発プロセスが異なりうる。ここで示した開発フローはいわば最大公約数として集約したものであり、特に基礎開発の段階では具体的な開発プロセスは開発コミュニティにより異なっている。これらについては第 2 章の活用例をご覧ください。

以下では、開発フローの各フェイズで実施すべきことを述べる。なお、前述の通りこのフローは数値予報課と気象研究所の共同開発を念頭において作成したものである。数値予報課が単独で実施する場合でもフローそのものは同じであるが、一部違いがあるため、適宜補足する。

(1) 開発課題検討

開発課題の設定においては、現業数値予報における課題を確認することが必要である。ここで、開発に着手する場合に何をすべきか自明でないが何らかの開発が必要である課題と、何に着手すべきか自明であるものの必要性が必ずしも明確でないものの両者がありうる。

前者には、例えば数値予報のユーザにとって望ましくない予測特性の解決といった課題や安定運用を確保するための計算安定性の向上といった課題が挙げられる。このような課題に対しては、人的・計算機資源とユーザにとっての利益を勘案して優先すべき課題を定め、その上で、具体的な開発課題に落とし込むことが必要である。

一方、後者には、例えば、現在の数値予報システムで用いられている定式化や離散化手法の不十分な点の改善及び新たな観測データの利用に向けた開発等が挙げられる。さらに、国内外で研究ベースで開発された手法を利用することもある。これらの開発課題に対しては、ユーザにとっての利益につながるかどうか、あるいは compensating errors も考慮した上でそもそも予測精度が向上するかどうかを精査する必要がある。その上で、課題として設定するかどうかを決定する。

続いて、開発工程を作成する。開発課題に対して、その段階で十分な科学的根拠が積み上げられているか、開発により想定される予測特性の変化をどれくらい事前に把握できるか、変更により影響の及ぶ範囲等を考慮し、必要な工程を作成する。その場合に関連する文献のレビューがこれらを考慮する際に有益である。ここでいう文献のレビューとは必ずしも論文に限らず、これまで開発で得られた知見も重要であることから、第 2.2 節で述べられているプロジェクト管理システムでの知見の集積が有益である。特に、持続的な開発においては、集積された過去の知見は欠かせない。

(2) 基礎開発及びその評価

開発課題が設定されれば基礎開発及びその評価のフェイズに移る。ここでは、基本的な開発として、開発プログラムの性能・インパクトを評価する。ルーチンと同等の環境で開発する必要はない(ガイドライン)。このフェイズは科学的根拠を積み上げるもっとも重要なフェイズである。従って、この段階で多くの開発時間を要することが通常である。

この段階での開発手法には開発コミュニティや開発対象により様々なものがある。特に実データでの実験に先立って検討すべきことが多い。例えば、力学過程の開発ではフルモデル(実際に現業数値予報システムで用いられるモデル)への組み込みに先立った理想実験(一様流による移流のテスト(石田ほか 2014)や浅水モデルによるテスト等)やフルモデルへ組み込んだ理想実験(山岳波(石田ほか 2014)等)がある。また、物理過程では SCM (Single Column Model) によるテストといった理想実験(原 2012b)がある。このような

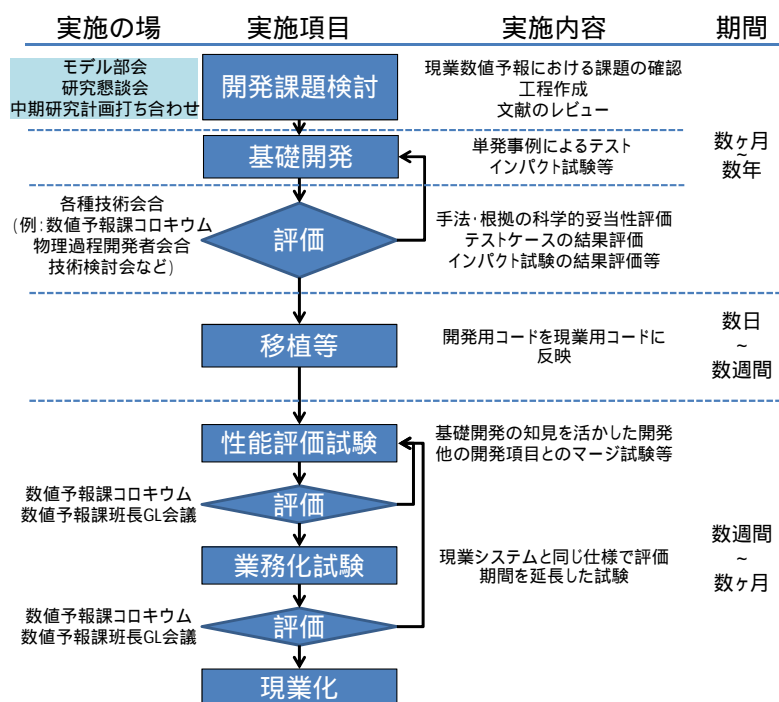


図 1.2.1 現業化までの開発フロー

テストでは解析解や特別観測等との比較が行われる他、国内外の多くの数値予報システムの開発者が参加するモデル相互比較プロジェクトにおいてモデル間の比較を行うこともある。データ同化システムの開発でも簡略化したモデルを用いた新しいアルゴリズムの効果の確認や1点観測データ同化実験といった理想実験が行われることがある。また、観測データの品質管理においては、異常データや同化手法の定式化より要請される仮定（観測データ同士に誤差相関がない、観測データにモデルに対するバイアスがない等）を満たさないデータの除去が必要である（佐藤 2011）。品質管理は多くのプロセスから成り立っており、実データでの実験を実施する前に不適切なデータが除去されているか、あるいは利用すべき適切なデータが誤って除去されていないか入念な確認が必要となる。

実データでの実験でルーチンと同等の環境でない実験の例として、数値予報モデルの開発におけるデータ同化サイクルを実施しない実験、全球モデル(GSM)開発における低分解能モデルによる実験（事例数を増やす、もしくは予報期間を延長する）等が挙げられる。さらに、計算機資源が許す範囲で、開発フローの次のフェーズをにらんで性能評価試験と同等の期間の実験を行うこともある。

実データの実験による結果から、重要な一部の指標（例えば、全球数値予報システムにおける台風進路予

測やメソ・局地数値予報システムにおける降水予測）もしくはそれらを含む複数の指標での精度の良し悪しのみをもって何らかのパラメータを決定するいわゆるチューニングはなるべく行わないことが望ましい。これは、第 1.2.3 項 で述べた問題点があるためである。とはいえ、観測により一意に定まらないパラメータもあり、実モデルでの評価に先立った基礎開発の段階でも決められないことがある。その場合は、比較実験により定める他ないことから、それに備えて観測の幅の範囲でパラメータの変更による影響の度合いを調査して知見を積み上げておくことも行われる。

このフェーズでは、評価とその結果に基づいた基礎開発のサイクルを入念に行うことが必要である。この評価のため、数値予報課コロキウム⁶、物理過程開発者会合⁷、グループミーティング⁸が開催され、議論が行われる。その他に、開発コミュニティにより様々なミーティングが開催されている。また、ミーティング以外にも開発管理サーバ(第 2.4 節)の Redmine(第 2.2 節)のチケット上で議論を行うこともあり、ミーティング

⁶ 原則、週に 1 回開催し、各自の開発課題の途中経過報告や性能評価試験・業務化試験の報告等を行う。

⁷ 技術開発推進本部モデル技術開発部会物理過程グループが年に数回程度開催し、物理過程の詳細な議論を行う。

⁸ 数値予報班の各グループがそれぞれ月に 1 回主催し、進捗報告等を行う。

を実施した場合でも、その議論結果をチケットに記載しておくことも有益である。

(3) 移植等

このフェイズでは、別のシステムで開発が行われた開発課題をルーチンと同等の仕様に移植すること等を想定している。基礎開発の段階で移植まで済んでいれば特に実施する必要はない。

従前は移植に要する開発コストは大きかった。しかし、近年は数値予報システムで用いられる言語がFORTRAN77からFortran90(もしくはそれ以降)となり、Fortran90以降で利用可能となるmodule文を利用したモジュール化・パッケージ化が進んできたこと、数値予報システムの運用で使われるスーパーコンピュータシステム以外のシステムやコンパイラの利用がしやすくなったこと等により、数値予報システムとは別のシステムからの移植の困難性は低下している。また、バージョン管理システムのマージ機能も移植に係るコスト低減につながっている。数値予報システムは複数のプログラムから成り立つシステムであるが、NAPEX(第3章)の整備により開発したプログラムをシステムに移植するコストも低減している。

(4) 性能評価試験及びその評価

ガイドラインでは「基本的な開発で成果が得られた場合、ルーチンと同等もしくは準ずる仕様で性能評価試験を行う。」としている。性能評価試験はシステムにより異なっており、全球数値予報システムでは、「現業システムと同等の仕様(ただし、全球速報解析は行わず、初期値として全球サイクル解析値を用いて1日1回(12UTC初期時刻のみ)の予報を行う)による、原則として夏期と冬期の各1か月程度のサイクル実験」とし、メソ数値予報システムでは、「現業システムと同等の仕様による、原則として夏期と冬期の各1週間程度のサイクル実験」としている。局地数値予報システムは毎時実行されることから、1日あたりの実行回数を減らして様々な事例に対する影響を評価することとしている。また、これらにおいて、実験の比較対象(コントロール)は原則として最新の現業システムによる実験結果とするとしている。

ここでは、NAPEXを利用した実験が行われる。複数の開発課題がある場合、独立した試験を行うこともあるが、前述のcompensating errorsの存在や次のフェイズを念頭におき、複数の開発課題を統合して試験を行うと、より手戻りが少なくなることが期待される。このフェイズでは、第1.2.3項(1)で述べた通り、信頼できる統計的根拠を与えるだけの事例数が十分でないことから、統計検証の結果を過度に重視せず、基礎開発での知見との整合性を重視することが望ましい。

評価にあたっては、数値予報課コロキウムにおける主として技術的な観点での議論と数値予報課班長会議⁹

⁹ 数値予報課長を筆頭に、課内の班長や各グループのリーダー等が参加する会議。原則週に1回開催される。

における主としてマネジメントの観点での議論・意思決定が行われる。マネジメントの観点とは他の開発課題を含めた全体の開発工程の確認・調整等を指す。次のフェイズでより多くの計算機資源及び時間を費やすことからその評価は慎重にする必要があり、数値予報課全体として複数の議論を経て次のフェイズに進むこととしている。

(5) 業務化試験及びその評価

業務化試験は性能評価試験で期待される改善を提示した上で、実験期間を延長する(全球数値予報システムにおいては夏期・冬期の各3か月、メソ数値予報システムにおいては夏期・冬期の各1か月)ことを主とする。局地数値予報システムにおいては、1日24回のテストを行い、現業運用に対して問題がないか確認する役割を持たせている。

ここでは、後続のシステム(ガイダンスや全球数値予報システムにおけるメソ数値予報システム)の評価が必要となる場合に備えて、必要なデータ(境界値等)の出力を行う。

このフェイズでは実際に現業化したことを想定した予測精度の評価が重要となる。近年、数値予報システムの用途が広がり、多くの要素・指標等が利用されている。開発において、全ての要素・指標が精度改善することは困難であるが、様々な観点で評価を行う。

評価にあたっては、性能評価試験同様に数値予報課コロキウムと数値予報課班長GL会議において議論・意思決定が行われる。

(6) 現業化

現業化の意思決定が行われた後は実際に現業化するためのフェイズに入る。ここでは、変更による予測特性の影響を庁内のユーザへ説明する。総合的な改善が現業化に必要であるが、前述の通り、全ての要素・指標において精度を改善することは困難である。また、精度が改善した場合でも特性が異なる場合は、ユーザの利用方法が変わりうることを考慮する必要があり、幅広く解説する必要がある。また、庁外のユーザへ技術情報の提供も行う。

その他に、現業化に向けた作業として、精度に影響のない高速化、必要な出力データの精査、時にはジョブ構成の見直し等も行われる。

ここで述べた開発フローは順調に開発が進んで、実際に現業化される場合を表している。しかし、開発課題によっては評価フェイズをパスできないこともあり、この場合は仕切り直すこともありうる(ただし、フロー図には明記していない)。

1.2.5 諸外国の動向

開発プロセスの検討においては、諸外国の現業気象機関における動向も参考にしている。英国気象局における開発プロセスは第1.3節や原・高谷(2013)を、欧州中期予報センター及び英国気象局における物理過程

開発におけるプロセスは高谷・原 (2012) を、参照いただきたい。

まず、米国環境予測センター (NCEP) の動向について簡単にまとめる。これは、数値実験作業部会 (WGNE)¹⁰ の第 31 回会合 (平成 28 年 4 月開催) で報告¹¹ された。NCEP では、数値予報システムの精度改善をなるべく速やかに実施して欲しいという要望と、改良後にポストプロセス等の開発が必要なことから、なるべくゆっくり改善して欲しいという相反するユーザの要望がある。そこで、NCEP では開発プロセスを以下のように定めている。まず、開発サイクルの当初にワークショップを開催し、優先すべき開発課題を費用対効果の面から検討し、スケジュールを策定する。このワークショップでは開発者だけでなく、カスタマー (ユーザ) も参加する。続いて、開発 (2-4 か月) と評価 (2 週間) を 1 つのサイクルとして繰り返す。なお、評価にはユーザも参加する。最終的にこのサイクルを繰り返して得られた変更が承認されると導入される。

次に WGNE 会合の動向も報告する。これまでの会合では各数値予報センターにおける改良の網羅的な概略報告が行われていた。これは、ある数値予報センターで導入された手法が改善につながるという情報だけでもその他の数値予報センターにおいて、同じ手法を導入することによる改善が期待されることから、このような情報交換は有益であったと考えられる。しかし、近年では、第 1.2.2 項で述べた通り、compensating errors の存在や実装の細部の検討の重要性が増したことにより、他の数値予報センターで改善が得られた手法をそのまま導入するだけでは精度向上につながらないことが多い。そこで、第 31 回会合から網羅的な概略を報告するよりはむしろ、何らかの開発項目に焦点をあてて開発で得られた知見を共有するように変更された。これは、数値予報課における問題意識が WGNE と共通していることを意味していると考えている。

1.2.6 まとめ

本節で述べた事柄はここ数年間で数値予報課の多くの開発者が検討・議論して得られた知見をとりまとめたものである。

以下では、今後検討すべきと考える項目について述べる。まず、多くの課題に対する開発が個別に行われているが、早い段階で統合して評価することを促進する必要があるだろう。個々の開発課題の単独の評価では精度改善の効果が見えにくいという問題への対応、compensating errors への対応が図れることが期待される。また、必要となる計算機資源を減らすことも期待

される。現在は、予報モデルの開発では基礎開発のフェイズで統合した評価が行われている。一方で、観測データの利用においては業務化試験のフェイズで統合することが多い。計算機資源を減らすことや第 1.2.3 項 (1) で述べた S/N 比が低くなったことを踏まえると、性能評価試験のフェイズから統合することも視野に入りたい。さらに、数値予報モデル・データ同化システム・観測データの利用の開発を統合した評価も今後の検討課題である。なお、改良をまとめて現業化することにより、ガイダンスの開発コストや特性の変化に対応するユーザ負担が減少するといった効果も見込める。また、ユーザーニーズをどのように汲み取り、開発プロセスにどのように反映させていくかの検討も必要である。

数値予報システムの開発プロセスは今後も継続して検討していく必要があると考えている。引き続き議論を行い、より良いものにしていきたい。

参考文献

- Geer, A., 2016: Significance of changes in medium-range forecast scores. *Tellus A*, **68**, 30229.
- 原旅人, 2012a: 数値予報モデルにおける物理過程の役割. 数値予報課報告・別冊第 58 号, 気象庁予報部, 2-7.
- 原旅人, 2012b: 理想実験による物理過程の評価. 数値予報課報告・別冊第 58 号, 気象庁予報部, 130-137.
- 原旅人, 高谷祐平, 2013: 海外数値予報センターの開発管理の例. 数値予報課報告・別冊第 59 号, 気象庁予報部, 195-199.
- 原旅人, 2015: 境界層過程の改良. 平成 27 年度数値予報研修テキスト, 気象庁予報部, 24-27.
- 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, 河野耕平, 松林健吾, 2014: 理想実験を通じたドライモデルとしての評価. 数値予報課報告・別冊第 60 号, 気象庁予報部, 62-87.
- 室井ちあし, 2012: 数値予報の流れ. 平成 24 年度数値予報研修テキスト, 気象庁予報部, 2-3.
- 室井ちあし, 2013: 開発管理の必要性. 数値予報課報告・別冊第 59 号, 気象庁予報部, 192-194.
- 佐藤芳昭, 2011: 数値予報データ同化システム. 数値予報課報告・別冊第 57 号, 気象庁予報部, 1-6.
- 高谷祐平, 原旅人, 2012: ECMWF および UKMO における物理過程開発. 数値予報課報告・別冊第 58 号, 気象庁予報部, 123-128.
- Vosper, S. B., A. R. Brown, and S. Webster, 2016: Orographic drag on islands in the NWP mountain grey zone. *Quart. J. Roy. Meteor. Soc.*, **142**, 3128-3137.

¹⁰ 数値予報モデルと大気気候モデルによる数値実験に関わる研究開発を推進するため、世界気象機関大気科学委員会と世界気候研究計画合同科学委員会の合同部会として設置されている専門部会。世界の主要数値予報センターの開発者が委員となっている。

¹¹ http://www.wmo.int/pages/prog/arep/wwrp/new/documents/5_wgne_ncep_M.Ek_April2016v2.pdf

1.3 英国気象局における全球モデルの開発プロセス¹

第 1.2.4 項では、数値予報課における数値予報モデル開発プロセス、およびそのプロセスにおける開発管理の活用について述べてきた。本節では、海外の数値予報センターにおけるモデル開発プロセスの例として、英国気象局における全球モデルの開発プロセスについて紹介する。本節に書かれている内容は英国気象局科学諮問委員会²等の公開資料や、筆者が 2014 年 6 月から 2 年間、英国気象局に滞在した際に実際に見聞きしたこと・経験したことに基づいている。

英国気象局では、短期予報から気候予測まで複数の時間スケールをカバーする高精度の全球モデルを開発する目標を掲げている (Brown et al. 2012)。その目標を達成するために、モデルの開発プロセスや必要な試験を明確にしている点、プロジェクト管理システム (第 2.2 節) を有効活用している点、実験システム・標準評価ツール等の開発基盤整備にも力を入れている点など、モデル開発コミュニティが開発管理を強化する上で参考にすべき点が多い。実際に、プロジェクト管理システムについては英国気象局の取り組みを参考に、気象庁でも利用が進んでいる (第 2.2.2 項)。

1.3.1 英国気象局の全球モデルの標準設定：GA (Global Atmosphere)

英国気象局では、全球・領域統一モデルである、UM (Unified Model) を開発・維持している。UM は共通の入出力システムのもと、全球モデル、領域モデル、移流拡散モデル等多種多様な用途に使われる基盤モデルとしての位置づけになっており、様々なオプション設定が含まれる。その中でも、UM の全球大気モデルとしての標準設定は GA (Global Atmosphere) と呼ばれる (Walters et al. 2011)。具体的には、使用する力学過程や物理過程の選択、パラメータの設定、入力データとしての地形や気候値などが GA に含まれる。英国気象局内での全球モデルを使った研究開発は、全て GA をベースに行われる。力学過程、物理過程等の改良のパッケージの導入によって、全球大気モデルとしての予測性能が上がると GA のバージョンが上がっていく。GA に関連するコンポーネントの設定として、陸面モデル GL (Global Land)、海洋モデル GO (Global Ocean)、海氷モデル GSI (Global Sea Ice) もあり、それらを結合したものが GC (Global Coupled) と呼ばれる。2017 年 1 月時点での最新バージョンは GA7, GC3 である。領域モデルについても全球モデルに倣い、RA (Regional Atmosphere) と呼ばれる標準設定を開発・管理するよう、準備が進められている。また、各バージョンのトラ

ックに相当する標準設定であり、次期バージョンの開発の雛形になるような設定であることを強調する際は、例えば “GA7.0” のようにバージョン番号の後に “.0” をつける。

GA の実行プログラムは UM 本体そのものである。そのため、GA の開発は UM 本体と並行かつ連動して行われる。開発者は何らかのモデル変更を GA に導入したい場合、まず UM にその変更を導入する。変更と前後して全球モデルとしての試験を行い、後述 (第 1.3.4 項) のプロセスを経て GA の設定として正式に採用される。その変更が反映されるバージョンのリリースは UM 本体と GA でタイミングが異なる。UM, GA のバージョンが別に存在することは一見奇異に感じるかもしれない。これは、UM のバージョン番号は全球モデルとしての性能を識別する番号ではないためである。前述したように UM は多様な用途で使用されており、多数のオプションが存在する。そのため、UM の各用途での設定は UM 本体の管理とは別に管理する必要がある。また、英国気象局全体として全球モデルの開発・改良を進めるためにも、GA のような標準設定を厳密に管理することが必要となる。全球モデルにおいては、UM, GA のバージョンアップはそれぞれ、ソフトウェア、モデルの予測特性の更新を示している側面が強い。そのため、例えば「UM10.2 での GA6.0 設定」、「UM10.3 での GA6.0 設定」といった呼称が存在する。しかし、この例の場合、GA のバージョンが同じであるので、全球モデルとしての予測特性は変わらない。深刻なバグ修正等がなければ、異なる UM バージョンでの同じ GA 設定の結果は厳密一致または無視できる範囲の差となる。

2017 年 1 月における UM 本体の開発プロセスは原 (2013) による解説から大きく変化していないため、本節の以降の項では GA の開発プロセスに絞って解説する。

1.3.2 GA と現業数値予報モデル

GA は現業数値予報モデルの設定としても使用される。図 1.3.1 は GA の更新と現業数値予報モデルの更新の関係を表した模式図である。GA を現業数値予報モデルとして採用する場合、パラメータやオプションについて小規模な変更を施すことがある。この場合、現業数値予報モデルとしての設定は GA のバージョン番号を x とし、“GA $x.1$ ” のように、GA $x.0$ から分岐させる形で区別して呼ばれる。現業数値予報モデルが GA $x.1$ からさらに分岐またはアップデートされることはなく (不具合修正や UM のソフトウェアとしてのバージョンアップ等を除く)、現業数値予報モデルを更新する際はより新しいバージョンである GA $y.0$ (y は x より新しいバージョン番号) から分岐された GA $y.1$ が採用される。GA $y.0$ を現業モデル向けに微調整を行う必要がない場合、GA $y.1$ は GA $y.0$ と全く同じものになる。GA が現業数値予報モデルとして採用されたのは、2011 年 3 月に導入された GA3.1 (Walters

¹ 氏家 将志

² Met Office Scientific Advisory Committee (MOSAC)
<http://www.metoffice.gov.uk/learning/library/publications/science/met-office-scientists/mosac>

et al. 2011) が最初である。さらに、2014年7月に力学コアの更新と物理過程の大規模な改良を含む、GA6.1が現業化された。原則的にGAのバージョンが進むごとに基本的な性能は向上していくが、予報スコアやプロダクトへの悪影響の問題等で現業数値予報モデルとしては採用を見送られることもある。実際、GA4.0とGA5.0は気候平均場の再現性は良いものの、予報スコアの面でGA3.1を上回ることができず、現業数値予報モデルとしては採用されていない。しかし、その間もGAの開発は最新バージョンをベースとして行われていた。

GAの更新の意思決定と、その現業導入の可否の意思決定は別途行われている点に注意したい。これはGAの利用形態のひとつは現業数値予報システムであるが、そのほかにも気候予測や研究などの用途にも利用されるためである。GAの更新の意思決定は本節で述べるプロセスを経てなされる。一方、現業数値予報システムへの導入の最終判断は、モデル・解析を問わず、科学部門、顧客サービス部門、IT部門の長によって構成されるWGOS (Working Group on Operational Suites) が行う。WGOSで承認された変更内容はPS (Parallel Suite) と呼ばれる並行運用を経て、現業数値予報システムに導入される。

1.3.3 GA開発におけるプロジェクト管理システムの利用

英国気象局における技術開発、例えばモデル開発やデータ同化システム開発、数値予報ルーチンシステム開発等においては、プロジェクト管理システムのひとつである、Trac (第2.2.1項) を用いて開発の情報の記録・共有や進捗管理がなされている。GAおよび関連するコンポーネントの開発においても、GMED (Global Model Evaluation and Development) と呼ばれる専用のTracが用いられており、各開発項目の記録と管理、開発の進捗管理、ミーティングの記録の共有などに活用されている。UMのソフトウェアとしての管理はUM system team と呼ばれる専門のチームがUM用のTrac上で行うが、GAの設定の管理や実験環境、評価・検証ツールの整備は全球モデル開発評価セクションが行う。GAの開発において、管理者(全球モデル開発グループのマネージャー)は次のバージョンに導入される開発項目、各項目のステータスを漏れ無く把握するためにTracを活用している面が強い。そのため、Tracのチケットの属性(第2.2.3項)のうち「マイルストーン」(チケット終了の目標)、「ステータス」(チケットの段階)を適切に記載することが重視されていたり、意思決定の記録としてTracのWiki上に各種会議の議事録が記載されている。

GAの開発に関わる人数は、バージョンにもよるが、力学過程、物理過程、実験システム、評価・検証システムの開発者、モデルの評価者を含めて40-50人程度で

ある³。気象庁の全球モデル開発(第2.5.2項)より規模は大きく、セクションをまたいだ開発も行われている。そのため、プロジェクト管理システムを活用した開発管理やモデル変更のより厳密な手続が必要となる。

1.3.4 GAシリーズの開発サイクル

GAシリーズの開発は概ね1年、長くても2年を1サイクルとして行われ、そのプロセスや必要な実験などの手続は明確に示されている。開発の1サイクルは、図1.3.2のような流れになっている。まず、サイクルの最初の「Prioritisation Meeting」で、次のバージョンのGAへの導入項目やその優先順位付けがなされる。その後、「個々の開発」の中で各項目の性能が評価される。仕様が固まった項目については、「Science Assurance Meeting」で各開発項目の導入の承認を諮る。承認された個々の項目については、取りまとめられ、「パッケージテスト」としてさらに総合的な性能評価が行われる。最終的な変更内容が固まれば「Code Freeze」とし、最終テストを経て、新バージョンとして「リリース」される。リリース後はユーザも交えた「Global Model Evaluation Workshop」で新バージョンの性能や課題が議論される。議論された内容は次々期のPrioritisation Meetingにも反映される。また、各フェイズで標準実験とその評価が行われており、次々期の課題設定にもフィードバックされている。開発サイクルに関連して、開発内容のUMのトランクへのマージや現業数値予報システムへの導入等も行われる。以下では個々のフェイズとその中で行われる標準実験について、詳細に説明する。

(1) Prioritisation Meeting

Prioritisation Meeting は開発サイクルの最初に行い、既存のモデルの問題点やこれまでの進捗、モデル開発の将来性の観点を勘案して、どの項目が優先事項かが議論される。この会合には力学過程・物理過程・全球モデル開発・部外連携セクションの各ヘッドやマネージャー⁴が出席し、議論および意思決定を行う。次期バージョンに導入される項目は“Implementation cycle”というカテゴリに所属し、このミーティングで各開発項目に対して、“critical”, “high”, “normal”などの優先順位付けがなされ、Tracに記録される。次期バージョンに導入できるほどの進展が得られていない開発項目については“Research cycle”というカテゴリに属し、基礎開発が継続され、次々期バージョンでのPrioritisation Meetingで再度議論される。

³ GA6.0/6.1の解説論文である、Walters et al. (2016)の著者として記載されている人数。さらにGA6の開発において、UMのシステム管理者、大気・海洋結合モデルの開発者、現業数値予報システムへの導入に関わったデータ同化システムの開発者等を含めると100人を超える。

⁴ 英国気象局の科学部門は各セクションごとに「ヘッド」と呼ばれる長があり、セクション内のグループごとに、グループを取りまとめる「マネージャー」を置いている。

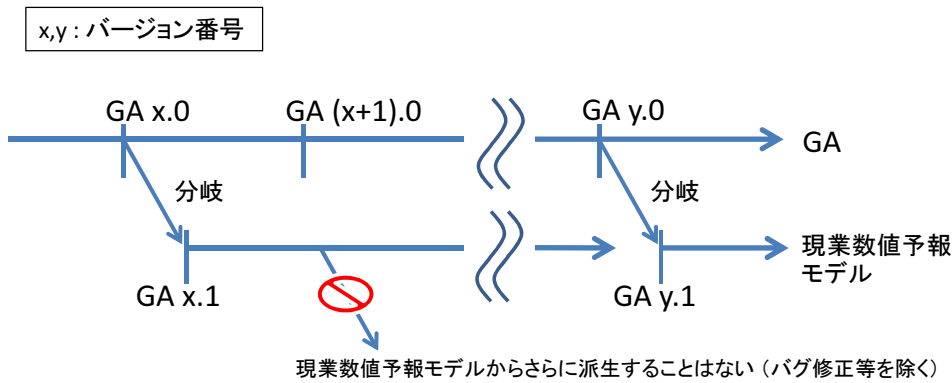


図 1.3.1 GA と現業数値予報システムで使われる全球モデルの関係を表した模式図。図中の水平の矢印はバージョンが上がる方向を、斜めの矢印は現業数値予報モデル用のブランチとして分岐することを表している。x, y は GA のバージョン番号 ($y > x$) を示している。

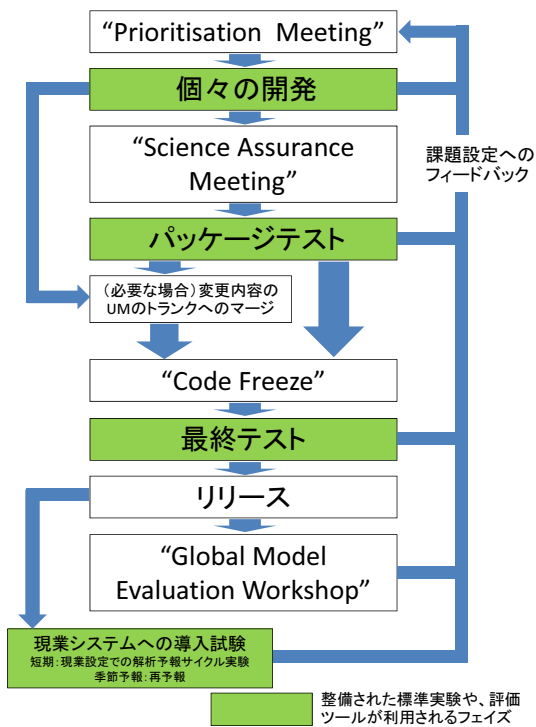


図 1.3.2 GA の開発 1 サイクルの流れ。図中の緑で塗られたボックスは、標準実験や評価ツールを利用するフェイズを示している。

(2) 個々の開発

個々の開発項目については、UM 本体の開発 (原 2013) と同様、その内容が Trac 上のチケットに必ず登録されていなければならない。また、UM 本体の開発と GA の開発は並行かつ連動して行われているため、GA のチケットから、UM の関連チケットにリンクが張られていることも多い。チケットには開発の目的を記述し、進捗や使用した UM のブランチの所在、実験結果等を記録する。また、変更内容が UM のトランクに既に反映されているか否かも記録する。変更内容が固ま

り、パッケージテストに導入する準備が整った場合は、開発者は各チケットにリンクされた、“TicketDetails” と “TicketSummary” と呼ばれる Wiki ページを作成する。TicketDetails では、変更の概要や、物理的基礎 (参照論文がある場合はそのリンク)、解像度依存性の有無を記述する。TicketSummary には、標準実験の結果を掲載し、変更のインパクトを改善点、改悪点含めて記述する。また、計算安定性・計算時間などに影響があればその旨も記述する。TicketDetails, TicketSummary を作成したら、レビューの段階に進み、開発者または管理者から指名されたレビューアによって変更内容が評価される。通常、レビューアは変更内容の分野に詳しい者が指名される。レビューアは変更内容が適切であるか、適切な実験がなされているかを審査する。レビューアからコメントや指摘があった場合、開発者はそれに対応する必要がある。レビューにおいては、精度向上とともに、物理的に適切な変更であるかも評価の対象となる。単体の変更で精度向上が見られなくても、物理的に正しい変更であり、さらに他の変更との組み合わせで精度向上が見込まれる場合は、レビューを通過することもある。

(3) 標準実験とその評価

標準実験は、短期予報実験と気候予測実験の 2 種類があり、開発者は両方を行わなければならない。短期予報実験では、水平格子間隔約 55 km のモデルを用いた、“Case Study” と呼ばれる ECMWF の解析値を初期値とした、夏冬合わせて 24 事例が標準事例として整備されている。気候予測実験については水平格子間隔約 130 km のモデルを用いた、AMIP ラン (Atmospheric Model Intercomparison Project : 大気モデルの相互比較プロジェクトで指定された実験で、大気モデルに海面水温解析値を与える) での 20 年積分が標準実験として行われる。短期予報実験と気候予測実験では、エロゾルの扱い (短期予報実験では気候値、気候予測実験ではエロゾルスキームで計算) や積雲対流スキ-

ムのパラメータの1つ、質量・エネルギー補正スキーム適用の有無などの違いはあるが、その他は原則的に共通の設定を用いる。

評価・検証についても標準パッケージが準備されており、短期予報実験については TRUI (Trial User Interface) と呼ばれる平均誤差、平方根平均二乗誤差等の標準的な予報スコアや平均予報場を計算する検証パッケージが、気候予測実験については Validation Note (気候平均場の検証や、放射収支の確認) Auto Assess (モンスーン、マッデン・ジュリアン振動、赤道波、ブロッキングなどの現象に応じた評価パッケージ) と呼ばれるパッケージが整備されている。これらの評価・検証ツールは標準実験を行えば誰でも簡単に実行できるよう整備されており、図 1.3.2 の緑で塗られたボックスのフェイズで使用される。また、開発者は必要に応じて追加の実験や、追加の解析結果も TicketSummary に記載することができる。解析予報サイクル実験の実施は個々の開発の過程では任意だが、大規模なモデル変更や陸面過程の変更など、解析場への影響が大きいことが予想される場合は、この段階で解析予報サイクル実験が行われることもある。

(4) Science Assurance Meeting

各チケットのレビューと前後して、各チケットで検討された変更をパッケージテストに導入してよいかの判断を Science Assurance Meeting と呼ばれる会合で行う。この会合には力学過程・物理過程・全球モデル評価開発・部外連携セクションの各ヘッドやマネージャー、データ同化・アンサンブル予報のセクションのヘッド、領域モデル開発グループのマネージャー等が出席し、意思決定を行う公式な会合である。パッケージテストへの導入項目はチケット単位で議論される。通常、チケットの担当者が会合の出席者に対してプレゼンテーションを行い、導入可否を諮る。モデルの予測精度に影響のある変更は原則的にすべて Science Assurance Meeting にかげられる。Science Assurance Meeting で議論されたすべての項目が承認されるとは限らず、次期バージョンに導入するには事前調査や準備が不十分であると判断され、“Research cycle” のカテゴリに差し戻される項目も実際に存在する。Science Assurance Meeting はパッケージの最終仕様が固まるまで随時開催される。この会合の議事録、決定事項、次回の会合までの課題等は Trac 上の Wiki に記録され、英国気象局の職員及び共同研究契約を結んでいる外部の協力者に公開される。

(5) パッケージテスト

レビューを通過し、Science Assurance Meeting で承認された項目はパッケージテストに組み込まれる。この段階ではレビューを通過した項目を統合して、性能評価試験が行われる。パッケージテスト自身も Trac のチケットの一つとして扱われ、実験結果を記録していく。性能評価のための標準試験は個々の開発の中で行

われるものと同じであるが、パッケージテストでは解析予報サイクル実験も行われる。パッケージテストの最初のバージョンに取り込まれる内容は優先度や進捗を考慮し、議論して決定される。その後、準備ができた項目やパッケージテストの中で発覚した問題の対応版を順次導入し、テストを行う。GA7 の開発では、最終的に GA6 に対して約 40 (チケット数) の変更がパッケージに取り込まれた。パッケージテストの段階で、新バージョンが前バージョンに対して予測精度が向上しているかどうかが強ク問われるようになる。もちろん、ただ精度が向上していればそれでよし、ということではなく、どの変更がどの要素の精度向上に寄与しているかの切り分けも行う。パッケージテストの実行は全球モデル評価開発セクション内の全球モデル開発グループが取りまとめて行うが、テスト結果については個々の変更の開発者も交えて頻繁に議論が行われる。必要に応じて、パラメータチューニングもこの段階で行われる。

(6) Code Freeze

パッケージが固まったら、Code Freeze とし、問題が発覚した場合を除き、原則的にそれ以後の変更は次々期のバージョンの開発に反映させる。また、新バージョンに導入される変更内容は、この段階で UM のトラックに反映されていなければならない。

(7) 最終テスト

Code Freeze 後は全球モデル評価開発セクションが 1 年程度の長期解析予報サイクル実験、さまざまな水平分解能での大気・海洋結合モデルでの長期積分、季節予報の再予報等を実行し、総合的な評価を行う。また、現業数値予報システムに導入するための試験 (現業数値予報システムと同じ水平分解能での解析予報サイクル実験) もこの段階で開始される。ただし、現業数値予報システムに導入するための判断は GA の開発サイクル本体とは分岐して行われる。

(8) リリース

Code Freeze 後、x をバージョン番号とすると、GA x.0 の名称で新しいバージョンがリリースされ、英国気象局内および外部のユーザが利用可能となる。新バージョンを現業システムに反映させる場合、現業用に細かい修正を加えたものはブランチとして GA x.1 の名称でリリースする。また、新しいバージョンがリリースされると、英国気象局内外のユーザがそのバージョンのモデルを使った研究に関する論文で引用できるよう、GA x.0/x.1 の概要や性能を解説した査読付き論文を投稿・出版する。2017 年 1 月現在、大気モデルについては GA3.0/3.1 (Walters et al. 2011), GA4.0 (Walters et al. 2014), GA6.0/6.1 (Walters et al. 2016) が、大気・海洋結合モデルについては GC2.0 (Williams et al.

2015) の解説論文が出版または投稿されている⁵。

(9) Global Model Evaluation Workshop

リリースと前後して、外部ユーザによる評価結果も含め、Global Model Evaluation Workshop が開催される。この会合では英国気象局の職員に加え、外部の研究協力者も出席し、新しいGAの性能や改善点・問題点が議論および共有される。また、優先して改善すべき点もリストアップされ、次々期の Prioritisation Meeting に向けた基礎資料となる。

1.3.5 GA シリーズの開発プロセスに関する分析

前項までは、英国気象局における、GAの開発プロセスを概観してきた。この項では、まとめにかえて、英国気象局の取り組みに関して、モデル開発コミュニティが参考にすべき点、気象庁における全球モデル開発との共通点など筆者なりの見解を述べたい。

まず、開発サイクルの初期段階で、重要開発項目の関係者間での認識合わせ、開発項目の具体的な優先順位付けをしている点は特筆すべきである。モデルの既存の問題点、その問題点に対するアプローチの進捗、さらにユーザからの要望を反映した上で、次の全球モデルの更新で何を改善するかを明確にしている。このことは、現業気象機関においてより良い気象・気候情報サービスを提供し続けるために非常に重要であるといえるだろう。気象庁においては、台風進路予測の精度向上や、予報現業を通じて見られる問題点（例えば、原(2015)で指摘された南岸低気圧事例におけるトラフの遅れなど）に対するアプローチ、その他全球モデルとしての精度向上（対流圏・成層圏の気温や風のバイアス、降水分布、熱帯循環の維持）前バージョンで残された（または副作用として顕在化した）問題などを勘案して、開発サイクルの初期段階で全球モデルの次期バージョンの目標を明確にすることが、開発サイクルを活性化する上で必要であろう。

次に、短期予報・気候予測両方での実験を義務付けていること、さらにそれら実験の検証・評価を標準化している点も注目すべきである。この点には、短期予報から気候予測まで複数の時間スケールをカバーする全球モデルを開発するという、英国気象局の方針が強く反映されている。また、(i) 複数の時間スケールの数値実験から、多角的にモデルの評価を行っている、(ii) 目標を達成するための開発環境の基盤整備がなされている、という意味で、モデル開発コミュニティが参考にすべき点でもある。

(i) について、筆者が英国気象局に滞在した際に、積雲対流過程における固体降水の融解過程の改良を行った。この際、短期予報実験と気候予測実験を両方行うことで見える改善点（例えば、対流圏中層における気

温のバイアス改善が時間スケールを問わず見られること）や問題点（例えば、北西太平洋の循環場のバイアス拡大が気候予測実験で顕著になること）などが見つかった。複数の時間スケールでの数値実験は、モデルの改良に問題がないかの確認や、compensating errors がないかを探すべき⁶ となり、改良を進める上で役立つことを実感した。短期予報モデルの開発においても、長期積分や気候予測実験による評価からバイアスの特性を探るのは有益である。

(ii) について、英国気象局においては

- 複数の時間スケールをカバーする統一的な全球モデルを開発する

という目的を達成するために

- 共通のモデル設定による、短期予報・気候予測実験システム
- 気候平均場や大気現象の再現性の評価に重点をおいた評価ツール

が整備されている。これは、英国気象局は気候研究を行うハドレーセンターや応用気候のセクションを有しており、気候研究や気候サービス業務の規模が大きいこと、英国内の天気予報作成支援は領域モデルが主な役割を担っていることから、全球モデルの評価は気候予測モデルとしての評価に重点を置いている傾向があることの反映であろう。

気象庁においては、全球モデル・全球データ同化システムを含む全球数値予報システムは、天気予報や台風進路予報、週間天気予報の作成支援が重要な役割であること、全球モデルは全球 EPS (経田 2016) 等のアンサンブル予報システムの予報モデルとしても利用されること、さまざまな水平分解能での実験や長期積分によって見えるモデルのバイアスの確認の重要性も認識されつつあることから、

- 天気予報、台風進路予報、週間天気予報等を支援する、全球数値予報システムを開発する
- 多角的な評価を行い、全球モデルの総合的な予測精度を向上させる

ことを目指すために

- 解析予報サイクル実験システム
- 気候予測実験システム
- 上記システムの評価・検証ツール

を整備・拡充することが重要になるだろう。実際に、第 2.5.6 項で述べられているように全球モデルの開発と並行して評価・検証ツールの開発も進められているところである。

また、最終的な予測精度はパッケージテストの中で評価し、個々の開発項目単体では、科学的妥当性や基礎調査が十分行われているかを重視している点も注目すべきである。このことは、第 1.2.2 項で述べられている、数値予報システムの誤差要因が数値予報の発展と

⁵ GA5.0 の論文は未出版であるが、力学コアの更新が最も大きな変更であったため、GA5.0 の代替として Wood et al. (2014) が引用されることがある。

⁶ 最終的に積雲対流スキームの他の部分の改良や、地形性重力波スキームの調整との組み合わせで大きな問題とならなくなることがわかった。

ともに変わってきていることと関係していると考えられる。かつては、モデルの中で考慮されていない重要な過程（例えば、1980年代以前における、全球モデルの重力波抵抗など）が数多くあり、一つの変更で劇的に精度が向上するということがあったが、現在はそのようなことは稀で、一つの項目を改善すると別の隠れていた問題が顕在化することのほうが多い。そして、多くの場合、その問題は Minor treatment（第 1.2.3 項）と関連している。最終的な予測精度をパッケージで評価するということは、大気現象をある程度再現するのに必要なスキームが出揃い、一つのスキームの追加や変更だけで精度が向上する時代はすでに終わったこと、今後は各過程を洗練させつつ、複数の変更や過程間の相互作用を考慮したうえで精度改善を図る時代になったことの反映であると考えられる。第 2.5 節に詳しく述べられているように、数値予報課での全球モデル開発もここ数年は同様のことを行っており、英国気象局での取り組みと整合している。パッケージテストをモデルの各パーツの基礎調査の次に行うステップとして重視することは、今後、継続的な予測精度向上を進める上で重要であると言える。もちろん、複数の変更を単純に組み合わせて精度が向上すれば良い、ということではなく、精度向上の原因の切り分けや、どのエラーが他のエラーを隠蔽しているのかの正しい見極めが必要になる。そのためには、担当者が自らの専門性を高めるだけでなく、周辺分野の相互理解やモデル開発全般に関する広い知識と経験が必要になる。また、モデル開発と評価を融合してモデルの総合的な性能を上げていくアプローチは、数値予報課コロキウムをはじめとした各種会合や数値予報研修テキストおよび数値予報課報告・別冊（付録 A）において、数値予報に関して幅広く理解し、議論する機会が多い気象庁では、より効果を発揮するのではないかと筆者は考えている。

気象庁と英国気象局では、数値予報システムの開発規模や関連する業務を取り巻く環境に違いはあるものの、英国気象局でのモデル開発の取り組みのうち、合理的と思われるものは気象庁の事情に応じてアレンジした上で反映することで、開発サイクルの活性化や数値予報システムの継続的な精度向上に資することができると考えている。

参考文献

Brown, A., S. Milton, M. Cullen, B. Golding, J. Mitchell, and A. Shelly, 2012: Unified Modeling and Prediction of Weather and Climate: A 25-Year Journey. *Bull. Amer. Meteor. Soc.*, **93**, 1865–1877.

原旅人, 2013: 英国気象局の開発管理. 数値予報課報告・別冊第 59 号, 気象庁予報部, 195–197.

原旅人, 2015: 事例検討. 平成 27 年度数値予報研修テキスト, 気象庁予報部, 82–99.

経田正幸, 2016: 全球アンサンブル予報システムの開発. 数値予報課報告・別冊第 62 号, 気象庁予報部, 52–57.

Walters, D., M. Brooks, I. Boutle, T. Melvin, R. Stratton, S. Vosper, H. Wells, K. Williams, N. Wood, T. Allen, A. Bushell, D. Copsey, P. Earnshaw, J. Edwards, M. Gross, S. Hardiman, C. Harris, J. Heming, N. Klingaman, R. Levine, J. Manners, G. Martin, S. Milton, M. Mittermaier, C. Morcrette, T. Riddick, M. Roberts, C. Sanchez, P. Selwood, A. Stirling, C. Smith, D. Suri, W. Tennant, P. L. Vidale, J. Wilkinson, M. Willett, S. Woolnough, and P. Xavier, 2016: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geosci. Model Dev.*, in review.

Walters, D. N., M. J. Best, A. C. Bushell, D. Copsey, J. M. Edwards, P. D. Falloon, C. M. Harris, A. P. Lock, J. C. Manners, C. J. Morcrette, M. J. Roberts, R. A. Stratton, S. Webster, J. M. Wilkinson, M. R. Willett, I. A. Boutle, P. D. Earnshaw, P. G. Hill, C. MacLachlan, G. M. Martin, W. Moufouma-Okia, M. D. Palmer, J. C. Petch, G. G. Rooney, A. A. Scaife, and K. D. Williams, 2011: The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations. *Geosci. Model Dev.*, **4**, 919–941.

Walters, D. N., K. D. Williams, I. A. Boutle, A. C. Bushell, J. M. Edwards, P. R. Field, A. P. Lock, C. J. Morcrette, R. A. Stratton, J. M. Wilkinson, M. R. Willett, N. Bellouin, A. Bodas-Salcedo, M. E. Brooks, D. Copsey, P. D. Earnshaw, S. C. Hardiman, C. M. Harris, R. C. Levine, C. MacLachlan, J. C. Manners, G. M. Martin, S. F. Milton, M. D. Palmer, M. J. Roberts, J. M. Rodríguez, W. J. Tennant, and P. L. Vidale, 2014: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations. *Geosci. Model Dev.*, **7**, 361–386.

Williams, K. D., C. M. Harris, A. Bodas-Salcedo, J. Camp, R. E. Comer, D. Copsey, D. Fereday, T. Graham, R. Hill, T. Hinton, P. Hyder, S. Ineson, G. Masato, S. F. Milton, M. J. Roberts, D. P. Rowell, C. Sanchez, A. Shelly, B. Sinha, D. N. Walters, A. West, T. Woollings, and P. K. Xavier, 2015: The Met Office Global Coupled model 2.0 (GC2) configuration. *Geosci. Model Dev.*, **8**, 1509–1524.

Wood, N., A. Staniforth, A. White, T. Allen, M. Diamantakis, M. Gross, T. Melvin, C. Smith, S. Vosper, M. Zerroukat, and J. Thuburn, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quart. J. Roy. Met. S.*, **140**, 1505–1520.