

## 複数モデルによる気候予測の評価と結合に関する IPCC 専門家会合参加報告

江守正多

(国立環境研究所 地球環境研究センター 温暖化リスク評価研究室長)

高藪 縁

(東京大学 気候システム研究センター 教授)

### 1. はじめに

表題の会合は、2010年1月25～27日に米国コロラド州ボルダーの米国国立大気研究センター (NCAR) で行われた。IPCC 第5次評価報告書 (AR5) の第1作業部会 (WG1) 共同議長 Thomas Stocker (スイスベルン大学)、Dahe Qin (中国気象局)、第2作業部会 (WG2) 共同議長 Christopher Field (米国スタンフォード大学) を始め、IPCC WG1、WG2 に関係する分野の各国の専門家を中心に約60名が参加した。主に Thomas Stocker が中心となって会議を進行し、WG1 事務局 (TSU) が会議の運営にあたった。日本からは、筆者2名が参加した。

### 2. 背景

IPCC の気候変化予測は、主として各国の研究グループが開発する複数の気候モデルの結果に基づいている。第4次評価報告書 (AR4) では、複数のモデルの平均的予測を示したり予測の幅を示したりする際に、参加した全てのモデルを平等に用いていた (モデル民主主義、1モデル1票)。しかし、個々のモデルの結果を観測データと比較すると、観測データを再現する性能にはモデル毎に優劣がある。したがって、性能に応じたモデルの重み付けもしくは順位付けを行い、性能の高いモデルをより信頼する、あるいは比較的性能が高いモデルのみを用いることにより、予測を改善したり、予測の不確実性の幅を狭めたりできる可能性がある。ところが、どのような性能の指標 (メトリック) を用いれば、予測の信頼性を適切に評価できるのかは自明でない。

そこで、本会合では、AR5 において複数モデルの結果を基に予測結果を取りまとめるための、気候モデルの性能評価とそれに応じた予測結果の取り扱いについての指針が議論された。さらに、その結果を地域規模の気候予測や影響評価研究に用いる際の指針についても併せて議論された。

### 3. 会議の流れ

1日目は、午前中に開会のあいさつ等に続いて複数モデルの利用手法に関する4件の基調発表 (Reto Knutti、Claudia Tebaldi、Isaac Held、Ben Santer) があった。午後には、3つのグループに分かれて議論を行った。グループ1は複数気候モデルの解析手法について、グループ2は地域規模の予測や影響評価について、グループ3はモデルの順位付けの可能性や意味についてであった。

2日目は、午前に確率的手法に関する基調発表2件（Myles Allen、David Sexton）と地域予測や影響評価に関する基調発表3件（Bruce Hewitson、Tim Carter、Linda Mearns）があった。午後には1日目と同様の3つのグループに分かれて議論を行った。

1日目の夕方と2日目の昼過ぎにはポスターセッションが行われた。江守は環境省環境研究推進費 S-5 テーマ1で行っている不確実性評価研究の成果を、高藪は同テーマ2で行っているモデル性能評価研究の成果を発表し、どちらも何人かの主要な専門家と議論することができた。

3日目はグループ議論の続き（グループ1と3は議論されていた内容が近かったので、この時点で合併）を行った後、その結果を基に全体会で指針の骨子案を取りまとめた。ここでまとめられた骨子案を基にして、何人かの執筆メンバーが本会合の報告書の原稿を執筆する予定である。

本会合で議論された内容は、基調発表やグループ議論の内容も含めて、3日目の最後に全体会で議論された内容に集約されているので、以下ではその内容を基に議論の結果を報告する。

## 4. 会議の結果

### 4. 1 基本的な認識

一般に、複数モデルの平均は、個々のモデルの結果と比較して、より観測データに近くなることが多い。これは各モデル固有のバイアスが互いに打ち消し合うためと考えられているが、平均しても残るモデル共通のバイアスがある点にも注意が必要である。また、変動度などに注目すると必ずしも複数モデル平均が優れているとはいえない場合がある。

また、モデルの性能を評価するメトリックを用いてモデルに重み付けや順位付けをしようという試みはこれまでにいくつも行われてきている。しかし、万能の一般的なメトリックというのは見つかっていない。違うメトリックを用いれば違う順位付けが得られる。つまり、様々な変数の観測データからの誤差を足し合わせたようなメトリックを用いて、モデルの優劣を一義的に論じることはできない。

むしろ、目的が異なれば、適切なメトリックも異なるであろう。将来予測の不確実性のうちのある部分は、観測可能な量と関連付けることができ、これを用いてメトリックを定義することができる。特に、プロセスやフィードバックの理解は予測の信頼性と関連しており、これがメトリックの基礎を与える可能性がある。（例えば、モデルの現在気候の季節変化における雪氷アルベドフィードバックの大きさは、そのモデルの将来予測における温暖化の進行に伴う雪氷アルベドフィードバックの大きさときれいな相関関係があることが明らかになっている。従って、現在気候の季節変化において同フィードバックの大きさが観測データに近いモデルほど、将来予測における同フィードバックの大きさの信頼性が高いと考えられる。）

また、各国の研究グループが開発したモデルを集めたアンサンブルは、系統的にサンプリングされたものではないことに注意が必要である。このため、重み付きモデルアンサン

ブルを確率分布と解釈するのは危険である。

#### 4. 2 複数モデル取り扱いの指針

(以下は、主として AR5 の執筆者にとっての指針として議論されたが、個々の研究者にとっての指針も実際には含んでいる。)

1. 第一ステップとして、まず順位付けや重み付けを行わないで、解析の対象となり得る全てのモデルを平等に用いた場合の統計量について記述しておくのがよい。
2. 順位付けや重み付けを行う場合、プロセスや現象に注目して観測可能な量と将来予測を関連させたようなものを使うのが一つの方法である。
3. プロセスに基づくメトリックは複数モデルの解析から得られる場合もあるし、プロセス研究から得られる場合もある。
4. 順位付けや重み付けを基に選択された複数モデルの統計を行う場合、全てのモデルを平等に用いた場合との結果の差を確認するのがよい。また、メトリックや観測データなどの選択によって結果がどう変わるか感度を見ておくのがよい。
5. アンサンブルの解釈をする場合、アンサンブルの性質（内部変動の違いなのか、モデルパラメータの違いなのか、モデル構造の違いなのか等）を理解し、明示しておく必要がある。これらに関して、非専門家に対する助言や例示があるのが望ましい。
6. モデルデータベースに結果を提出するためのモデル性能についての制限は無い。しかし、研究者が特定の解析を行う際に、モデルの一部を選択することはあるだろう。ただし、その場合には選択の理由を明示するべきである。
7. 研究者がモデル性能のメトリックを考える場合、複数の活動（WCRP など）によって開発されている様々なメトリックを検討するとよい。複数のメトリックの結果を一か所に置く場所を作ることも考えるべきである。
8. 各々の研究グループのモデルによる「ベスト実験」と一つのモデルを基にして物理過程やパラメータを変えるなどした「物理摂動実験」とは、区別して扱った方がよい。後者は、将来の気候変化の可能性の広がりについて有用な情報を提供し、ベスト実験のみでは得られないモデルの多様性を知るために役立つ。また、10 年スケール予測は長期予測とは異なる。
9. 「パーフェクトモデル」テストを行って手法をテストするのが推奨される。つまり、モデルのうちの一つを「正解」と仮定して、他のモデルが「正解」を予測する際の信頼性をメトリックで期待通り評価できるかテストする。
10. 古気候シミュレーションは、モデル評価において独立な情報を提供しうる。
11. 設定の違いに注意すれば、CMIP3 のアンサンブルを CMIP5 と合わせて使ってもよいだろう。また、モデルが世代を経るごとに改善することを調べるのもよい。
12. モデル間の結果の一致度は、必ずしも確からしさを意味しないことに注意が必要。
13. モデル毎のアンサンブルメンバー数が違う場合について注意する必要がある。モデル毎のアンサンブルメンバー数がモデルの重みに影響するのはよくない。

14. 透明性確保のため、モデルのデータはすべて PCMDI などのデータベースに置かれ、観測データセットも PCMDI に置かれるかそこからリンクを貼られるとよい。複数モデルの加工したデータを置く場所を作ってもよい。解析や作図はあとで再現できるように適切に記述を残すことが必要。

#### 4. 3 地域予測と影響評価への指針

地域的な気候変化を評価する場合、1. 歴史的な変化、2. プロセスの変化、3. 全球気候モデルによる予測、4. ダウンスケールされた予測、を四本柱としてできるだけ全て考慮すべきである。気候モデルの不確実性、ダウンスケーリングの不確実性も網羅的に考慮されるべきである。しかし、個々の研究がこのような不確実性を網羅することは実際には難しい。そこで、個々の研究が特定の気候予測に基づいている場合、その特定の気候予測がこのような複数の情報源による広い「文脈」のどこに位置するかを確認することが、評価の上で重要である。

地域的な評価をする場合、その地域の気候の（あるいは関連するプロセスの）再現性能が悪い気候モデルを外すことがあり得るが、その基準は自明ではない。また、ダウンスケールの際に、計算量の都合や利用可能なデータの都合で一部の気候モデルしか用いられない場合があるので、そのような場合は明示されるべきである。

#### 5. その他

会議の導入部で、IPCC AR5 に引用される論文の締め切りが以下のように明示された。（これは WG1 のスケジュールのはず）

2012 年 7 月 31 日までに投稿された論文

2013 年 3 月 15 日までに受理された論文

が、評価の対象となる。

また、WG1 AR5 の章の構成で、筆者が理解できていなかった点の一つわかったので、ついでに記しておく。AR4 では地域毎の気候予測の章（11 章）があったが、AR5 ではこれに代わり地域気候に関連する現象毎の予測の章（14 章）になる。この場合、地域毎の予測の記述はされないのだろうか、と思っていたが、これについては（全球気候予測についてのまとめと共に）付録として「世界地図」(Atlas of Global and Regional Climate Projections) を付けるそうである。合意された章の構成に確かに Annex I として Atlas が入っている。

しかし、この構成では、地域気候モデルを用いた研究などの引用される箇所はどこかという疑問が依然として残る。関連する専門家に個人的に質問してみたが、この点はまだ十分に議論されていないという印象を受けた。

#### 6. 所感

この会議で、気候モデルの成績付けの基準や「足切り」の基準が議論されるのではないかと心配していた向きもあるかもしれないが、さすがによくわかっている専門家が集まっ

で議論しただけあり、妥当な（無難な？）認識を共有して終わったといえるだろう。

複数モデルを用いて不確実性の幅を把握することの重要性、性能評価のためのメトリックがプロセスの理解に基づくことの重要性、メトリックの有効性を統計的にテストすることの重要性などは、いずれも筆者たちが環境省環境研究推進費 S-5 等を進めるにあたって持ち続けていた認識と合致するので、国内のモデル評価研究、不確実性評価研究が国際的な専門家の認識に照らして妥当な方向に進んでいることを確認できた。今後はその成果を国際的に発信していくことにさらなる努力をしたい。

IPCC AR4 にいくつかの間違いが見つかったことなどをきっかけに、IPCC に社会的な批判の声が浴びせられていたタイミングでの会議であったので、データ公開、解析・作図の再現性確保など、IPCC の透明性を神経質に意識した議論もみられたのが印象に残った。

（江守）

会議の初っ端の基調講演において Reno Knutti 博士が 20 数個の「ジェット飛行機」の写真を並べ、これを全部「平均」したらどうなるか！という合成写真を見せた。すぐ想像できるように、およそ飛行機と判別するのは困難なぼやけた映像が現れ、会場の笑いをとった。20 数個の「ジェット飛行機」の写真には、いくつかのヘリコプターやプロペラ機の写真があり、「これらは明らかにジェットではない」という話であった。つまり、最低限の選択はあってもよいというメッセージとして印象に残った。その後の会議は「それではどう選択していくのか」という論調になっていくのかと思ったが、さにあらず、全体として、将来予測のスプレッドを知るためなどにも、多様性（diversity）を重視するという意見に収束していった。ただし、特定の視点毎に言えば、モデルに優劣があるのは確かに認められることであり、多様な複数モデル出力の中から、正当な理由の下に選択的な利用を行うのは、それぞれの研究者の力量と経験に任されるということである。

Jerry Mheel 博士の最後のまとめは、「前回のレポートでは、ある国の 2 つのモデルが最大感度と最小感度を示し、それらを統計から除くべきかどうかという議論があったが、平均を大きく変えることはないので多様性を残すことにした」などという言葉から始まった。やや身構えたが、特に明らかな物理的欠陥などがない場合にはむやみに除外すべきではないという論調だったようで、多くのモデルの答えが一致したからと言ってそれが確実性を保証するものではないという指摘とともに、複数モデルの評価・利用には、個々の研究者の繊細な検討力が問われるということが多くの専門家の意見として印象に残った。また、複数モデルの扱いにおいては純粋な統計的利用のみでなく、プロセス再現の妥当性についても考慮すべきという認識が共通していることも印象的であった。

（高藪）

以上