

第 6 章 コンセンサス予測¹

6-1 コンセンサス予測とは

複数の予測結果の平均による予測を、一般的にコンセンサス予測と呼ぶ。平均のとり方は、コンセンサス予測を構成する各予測結果に均等な重みを置く単純平均、予測精度が相対的に高い予測結果に大きい重みをつける重み付き平均がある（例えば、山口 2013）。コンセンサス予測は熱帯低気圧の進路予報の分野で幅広く利用されており（Elliott and Yamaguchi 2014）、コンセンサス予測の方が各予測よりも統計的に精度が高くなることが多くの先行研究で示されている（例えば、Goerss 2000; Goerss *et al.* 2004; Komori *et al.* 2007; Sampson *et al.* 2007）。強度予報の分野では、NHC や米国合同台風警報センター（Joint Typhoon Warning Center、JTWC）において、数値予測の結果や統計・力学手法による強度予測の結果など複数の予測結果に基づくコンセンサス予測が発表予報に採用されている（DeMaria *et al.* 2014）。

本章では、現在気象庁の強度予報で使用されている気象庁全球モデル（JMA 2013）による予測結果と統計手法による強度予測である SHIFOR（付録 1）、さらに台風予報・解析技術高度化プロジェクトのもと新たに開発した SHIPS（2 章）、LGEM（3 章）、CHIPS（5 章）による 5 つの予測結果を用いたコンセンサス予測の精度検証結果を示す。また、この 5 つの予測結果に、気象庁非静力学モデル（NHM、Saito *et al.* 2006）による台風強度予測を加えた、全 6 つの予測結果でコンセンサス予測を作成した場合の精度検証結果も示す。

6-2 使用データ

SHIPS、LGEM、CHIPS の予測データは、それぞれ 2 章、3 章、5 章で用いられている 2013 年～2015 年のデータを使用した。気象庁全球モデル（以下、GSM）と SHIFOR の予測データは、気象庁の現業強度予報で使用されている 2013 年～2015 年のデータを使用した。気象庁非静力学モデルによる予測データに関しては、Ito *et al.* (2018) が 2012 年～2014 年の台風を対象として北西太平洋域全域で数値実験を行った際のデータを使用した。数値実験の設定等については Ito *et al.* (2018) を参照されたい。検証には、気象庁ベストトラックを使用した。

6-3 気象庁版コンセンサス予測の構築

本章では、コンセンサス予測を構成する各予測結果に均等な重みを置く単純平均に加え、重み付き平均によるコンセンサス予測の検証を行う。また、予測データをそのまま使う手法に加え、初期時刻によるバイアスを補正した予測（傾向予測）によるコンセンサス予測の検証も行う。本章のコンセンサス予測は、中心気圧（Pmin）を対象としたコンセンサス予測であり、6-4 節の精度検証では中心気圧の予測精度に関して検証が行われる。

6-4 精度検証

まず始めに、個別の予測結果の検証結果を示す。これらは各章で詳細に記述されているが、本章ではそれらをまとめて共通のサンプルで検証を行った結果を示す。このように検証対象の全ガイダンスで検証サンプルを揃えて検証を行う手法を、共通サンプル検証という。図 6.1 は、GSM、SHIPS、LGEM、CHIPS、SHIFOR による強度予測の平方根平均二乗誤差（RMSE）、平均絶対誤差（MAE）、系統誤差（BIAS）で、検証対象は 2013～2015 年の全台風、全初期時刻（00、06、12、18 UTC）の予測である。GSM は、12UTC 初期時刻の予報時間は 11 日であるが、00、06、18UTC 初期時刻は予測時間が 84 時間であるため、

¹ 山口宗彦

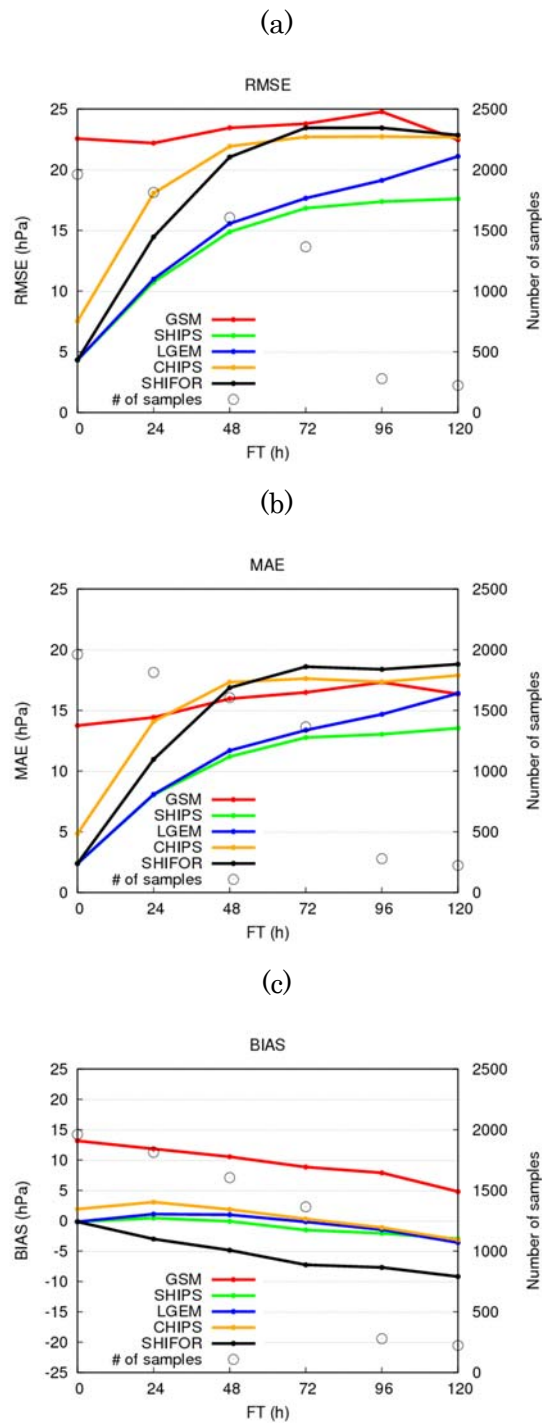


図 6.1 GSM(赤線), SHIPS(緑線), LGEM(青線), CHIPS(橙線), SHIFOR(黒線) による強度予測の (a) 平方根平均二乗誤差 (RMSE) , (b) 平均絶対誤差 (MAE), (c) 系統誤差 (BIAS)。単位は hPa で左 y 軸に対応する。黒丸は検証の事例数で右 y 軸に対応する。検証対象は 2013 ～ 2015 年の全台風、全初期時刻の予測 (00, 06, 18UTC 初期時刻は予測時間 3 日まで、12UTC 初期時刻は 5 日まで)。

検証の事例数は 4、5 日予測で少なくなる。SHIPS は、RMSE、MAE、BIAS のどの検証においても、既存の GSM や SHIFOR よりも誤差が小さい。RMSE、MAE は、3 日予測までは誤差が増大し、それ以降ほぼ変化しないという特徴を持つ。BIAS は、予報期間を通じてほぼゼロに近い値であるが、予測前半(1、2 日予測)は中心気圧を浅く、予測後半(4、5 日予測)は中心気圧を深く予測する傾向がある。LGEM は、SHIPS とほぼ同じ特徴を持っているが、予測時間 4、5 日では SHIPS よりも RMSE、MAE が大きい。RMSE は、既存の GSM や SHIFOR よりも誤差が小さいが、MAE で評価すると LGEM と GSM は 5 日予測でほぼ同じ値となっている。CHIPS は、既存の GSM や SHIFOR と比較すると BIAS は小さい。RMSE、MAE では目立った改善は見られないものの、少なくとも GSM と SHIFOR と同程度の精度を持っている。CHIPS は台風強度の傾向予測は得意であり、また SHIPS や LGEM が不得意とする急発達への予測を得意とする。また、コンセンサス予測の観点からは、独立した予測結果を複数用いることが精度向上に寄与すると期待される。

表 6.1 は、精度評価のベースラインとなる SHIFOR の RMSE、MAE に対する各予測の改善率(%)を示す。RMSE の改善率と MAE の改善率を比較すると、概して後者の方が大きい傾向にある。これは、RMSE は大外しをより大きな誤差として扱うためであると考えられる。予報時間別に見た最大改善率は RMSE で 22.9 ~ 29.2 %、MAE では 26.6 ~ 33.6 % であり、全予報時間で SHIPS が最大の改善率を与える。

次に、コンセンサス予測の検証結果を示す。本調査では、1) 単純平均、2) 傾向予測を使った単純平均、3) 重み付き平均、4) 傾向予測を使った重み付き平均、の全 4 種類のコンセンサス予測を作成した。「単純平均」は、GSM、SHIPS、LGEM、CHIPS、SHIFOR の全 5 つの予測結果に均等に 0.2 の重みを付けた予測結果である。「傾向予測」とは、予報初期時刻におけるベストトラックと予報時間 0 時間における予測値との差(初期バイアス)を、その後の予測値に足したものを予測値とするものである。「重み付き平均」および「傾向予測を使った重み付き平均」は、RMSE、MAE の改善率がそれぞれ最大となるように各予測結果の重みを予報時間ごとに変えたものである。表 6.2 は、SHIFOR の RMSE、MAE に対する各コンセンサス予測の改善率(%)を示す。表 6.1 同様、MAE の改善率の方が RMSE の改善率よりも大きい傾向がある。予報時間別に見た最大改善率は RMSE で 26.2 ~ 33.9 %、MAE では 30.1 ~ 36.0 % であり、個々の予測結果が与える最大の改善率よりも大きくなる。1、2 日先は、「傾向予測を使った重み付き平均」、3 日先以降は「重み付き平均」が最大の改善率を与える。表 6.3 は、最大の改善率を与えるコンセンサス予測の各強度予測の重みを示す。図 6.1、表 6.1 の結果が示すとおり、SHIPS の重みが相対的に大きい傾向がある。CHIPS や SHIFOR の重みは予測時間によってはゼロでは無く、改

表 6.1 各強度予測 (GSM, SHIPS, LGEM, CHIPS) の SHIFOR に対する改善率(%). (a) は RMSE, (b) は MAE の改善率で、各予測時間で最も精度の良い強度予測の改善率を太字とした。検証のサンプルは図 6.1 と同じ。

(a)					
改善率(%)	1 日予測	2 日予測	3 日予測	4 日予測	5 日予測
GSM	-53.4	-11.4	-1.4	-5.6	+1.7
SHIPS	+25.7	+29.2	+28.2	+25.8	+22.9
LGEM	+23.9	+26.0	+24.7	+18.4	+7.6
CHIPS	-24.8	-4.1	+3.1	+3.0	+0.7

(b)					
改善率(%)	1 日予測	2 日予測	3 日予測	4 日予測	5 日予測
GSM	-31.5	+5.3	+11.3	+5.8	+13.0
SHIPS	+26.6	+33.6	+31.3	+29.1	+28.0
LGEM	+26.2	+30.7	+28.1	+20.1	+12.7
CHIPS	-28.5	-2.6	+5.2	+5.6	+4.8

善率の向上に寄与していることが分かる。GSM は、予測時間後半で重みが大きくなる傾向がある。これは、力学的な手法に基づいている予測結果の影響が予測時間とともに増大し、一方で統計的な手法に依存している SHIPS、LGEM の影響が予測時間とともに減少していることを示しているのかも知れない。

最後に、これまでに登場した全 5 つの予測結果に、気象庁非静力学モデル (NHM、Saito *et al.* 2006) を加えて全 6 つの予測結果でコンセンサス予測を作成した場合の改善率を示す。Ito *et al.* (2018) は、2012 ～ 2014 年の全台風、12UTC 全初期時刻の予測 (3 日予測) を NHM を使用して行い、その強度予測誤差が GSM よりも小さいこと、さらに米国のハリケーン専用の領域モデル (Hurricane Weather Research and Forecasting Model; HWRF) と同程度の予測精度を持つことを示した。表 6.4 は、SHIFOR の

表 6.2 各コンセンサス予測 (単純平均、単純平均+傾向予測、重み付き平均、重み付き平均+傾向予測) の SHIFOR に対する改善率 (%)。 (a) は RMSE, (b) は MAE の改善率で、各予測時間で最も改善率の高い強度予測の値を太字とした。検証のサンプルは図 6.1 と同じ。

(a)

改善率(%)	1 日予測	2 日予測	3 日予測	4 日予測	5 日予測
単純平均	+18.8	+25.8	+27.8	+26.1	+24.5
単純平均 + 傾向予測	+30.5	+26.4	+24.2	+20.9	+17.0
重み付き平均	+27.4	+30.1	+30.2	+27.7	+26.2
重み付き平均 + 傾向予測	+33.9	+31.6	+29.1	+26.2	+21.1

(b)

改善率(%)	1 日予測	2 日予測	3 日予測	4 日予測	5 日予測
単純平均	+21.8	+30.4	+30.5	+27.5	+29.1
単純平均 + 傾向予測	+32.7	+29.5	+25.4	+20.4	+20.8
重み付き平均	+29.0	+35.3	+34.8	+30.1	+32.6
重み付き平均 + 傾向予測	+36.0	+35.4	+31.2	+27.3	+26.8

表 6.3 表 6.2 で最も改善率の高い強度予測を与える各強度予測 (GSM, SHIPS, LGEM, CHIPS, SHIFOR) の重み。 (a) は RMSE, (b) は MAE。検証のサンプルは図 6.1 と同じ。

(a)

重み	1 日予測	2 日予測	3 日予測	4 日予測	5 日予測
GSM	0.1	0.1	0.2	0.2	0.3
SHIPS	0.5	0.6	0.5	0.5	0.4
LGEM	0.3	0.2	0.2	0.2	0.2
CHIPS	0.1	0.1	0.1	0.1	0.0
SHIFOR	0.0	0.0	0.0	0.0	0.1

(b)

重み	1 日予測	2 日予測	3 日予測	4 日予測	5 日予測
GSM	0.2	0.1	0.3	0.2	0.3
SHIPS	0.4	0.6	0.5	0.5	0.4
LGEM	0.3	0.2	0.2	0.2	0.2
CHIPS	0.1	0.1	0.0	0.1	0.0
SHIFOR	0.0	0.0	0.0	0.0	0.1

RMSE、MAE に対する各コンセンサス予測の改善率 (%) で、NHM を使用した場合と使用しない場合の改善率をそれぞれ示す。検証対象は 2013 ～ 2014 年の全台風で、12UTC 初期時刻の予測で予測時間 3 日である。RMSE、MAE とともに、どの予測時間でも NHM を使用した方が改善率は大きい。例えば、3 日予測では、RMSE は 35.7 から 39.8 % へ、MAE は 42.1 から 46.8% へと、4 % 程度改善率が上昇しており、NHM を使用することにより予測精度が向上することが分かった。

6-5 まとめと課題

本章では、気象庁の強度予報で使用されている GSM、SHIFOR に、「台風予報・解析技術高度化プロジェクトチーム」で開発を行った SHIPS、LGEM、CHIPS を加え、全 5 つのガイダンスを用いてコンセンサス予測による強度予測の有効性を調査した。SHIFOR からの改善率で見ると、個別のガイダンスの精度検証では、予報時間 1 ～ 5 日の全予報時間で SHIPS が最大の改善率を与えた。これに対して、重み付き平均、または傾向予報を用いた重み付き平均によるコンセンサス予測は SHIPS よりも改善率が大きくなることが分かった。さらに、上記 5 つのガイダンスに NHM による強度予測を加えてコンセンサス予測を作成すると、さらに改善率が大きくなることが分かった。

現業予報の観点からは、重み付き平均という手法は扱いづらいかも知れない。本調査では、単純平均、または傾向予報を用いた単純平均によるコンセンサス予測の検証を行ったが、SHIPS 単独の予測に対す

表 6.4 NHM による強度予測を加えない場合と加えた場合の各コンセンサス予測 (単純平均、単純平均 + 傾向予測、重み付き平均、重み付き平均 + 傾向予測) の SHIFOR に対する改善率 (%)。 (a) は RMSE, (b) は MAE の改善率で、各予測時間で最も改善率の高い強度予測の値を太字とした。検証対象は 2013 ～ 2014 年の全台風、12UTC 初期時刻の予測 (予測時間 3 日まで)。

(a)

改善率(%)	1 日予測		2 日予測		3 日予測	
	NHM なし	NHM あり	NHM なし	NHM あり	NHM なし	NHM あり
単純平均	+18.9	+12.6	+25.0	+25.3	+31.6	+34.0
単純平均 + 傾向予報	+31.7	+30.5	+27.0	+24.2	+29.3	+24.4
重み付き平 均	+30.1	+21.3	+30.6	+31.9	+35.1	+39.8
重み付き平 均 + 傾向予報	+35.0	+36.1	+31.8	+31.2	+35.7	+31.3

(b)

改善率(%)	1 日予測		2 日予測		3 日予測	
	NHM なし	NHM あり	NHM なし	NHM あり	NHM なし	NHM あり
単純平均	+19.9	+14.9	+33.0	+34.1	+37.5	+40.5
単純平均 + 傾向予報	+31.4	+29.9	+33.0	+29.7	+33.7	+28.1
重み付き平 均	+29.3	+22.1	+37.4	+39.1	+42.1	+46.8
重み付き平 均 + 傾向予報	+32.8	+34.1	+36.9	+36.7	+41.9	+36.1

るメリットは大きくなかった。コンセンサス予測では、同程度の精度の予測を組み合わせることで精度の改善が見込め、精度が相対的に悪い予測を加えると逆にコンセンサス予測の精度が悪化する可能性がある（例えば、Elsberry and Carr 2000）。実際のコンセンサス予測の運用では、そのような観点からコンセンサスを構成する予測の選択（たとえば、SHIPS と NHM の 2 つによるコンセンサス予測など）を慎重に行う必要があるだろう。

参考文献

- DeMaria, M., C.R. Sampson, J.A. Knaff, and K.D. Musgrave, 2014: Is Tropical Cyclone Intensity Guidance Improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, <https://doi.org/10.1175/BAMS-D-12-00240.1>
- Elliott, G., and M. Yamaguchi, 2014: Advances in Forecasting Motion, *WMO 8th International Workshop on Tropical Cyclones (IWTC-8)*, 44pp. [Available online at http://www.wmo.int/pages/prog/arep/wwrp/new/documents/Topic1_AdvancesinForecastingMotion.pdf]
- Elsberry, R. L., and L. E. Carr III, 2000: Consensus of dynamical tropical cyclone track forecasts - Errors versus spread, *Mon. Wea. Rev.*, **128**, 4131–4138, [https://doi.org/10.1175/1520-0493\(2000\)129<4131:CODTCT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)129<4131:CODTCT>2.0.CO;2)
- Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models, *Mon. Wea. Rev.*, **128**, 1187–1193, [https://doi.org/10.1175/1520-0493\(2000\)128<1187:TCTFUA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1187:TCTFUA>2.0.CO;2)
- Goerss, J. S., C. R. Sampson, and J. M. Gross, 2004: A history of western North Pacific tropical cyclone track forecast skill, *Wea. Forecasting*, **19**, 633–638, [https://doi.org/10.1175/1520-0434\(2004\)019<0633:AHOWNP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0633:AHOWNP>2.0.CO;2)
- Ito, K., M. Sawada, and M. Yamaguchi, 2018: Tropical cyclone forecasts for the Western North Pacific with high-resolution atmosphere and coupled atmosphere-ocean models, *Papers in Meteorology and Geophysics*, **67**, 15–34, <https://doi.org/10.2467/mripapers.67.15>
- Japan Meteorological Agency, 2013: Outline of the operational numerical weather prediction at the Japan Meteorological Agency. Appendix to WMO Technical Progress Report on the Global Data-processing and Forecasting System and Numerical Weather Prediction Research. Japan Meteorological Agency, Tokyo, Japan. [Available online at http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2013-nwp/pdf/outline2013_all.pdf]
- Komori, T., M. Yamaguchi, R. Sakai, and Y. Takeuchi, 2007: WGNE intercomparison of tropical cyclone forecasts with operational global models: Quindecennial report, *Science Highlights*, WCRP, 4pp.
- Saito, K., T. Fujita, Y. Yamada, J. Ishida, Y. Kumagai, K. Aranami, S. Ohmori, R. Nagasawa, S. Kumagai, C. Muroi, T. Kato, H. Eito, and Y. Yamazaki, 2006: The Operational JMA Nonhydrostatic Mesoscale Model. *Mon. Wea. Rev.*, **134**, 1266–1298, <https://doi.org/10.1175/MWR3120.1>
- Sampson, C. R., J. A. Knaff, and E. M. Fukada, 2007: Operational evaluation of a selective consensus in the western north pacific basin, *Wea. Forecasting*, **22**, 671–675, <https://doi.org/10.1175/WAF991.1>
- 山口宗彦, 2013: 台風の進路予報, 気象研究ノート 台風研究の最前線 (下), **227**, 15–35.