

應用貝氏模型平均法整合ENSO海溫統計預報模型

陳翠玲¹ 朱寶信² 陳孟詩³

中央氣象局科技中心¹ 夏威夷大學² 中央氣象局第三組第二科³

摘要

本文是中央氣象局104年度「災害性天氣監測與預報作業建置計畫」的第二部分。綜合預報是一個決定預報品質的數值過程，可以被用來量化預報值與觀測值間的關係。理論上，結合兩個以上的預報模式能夠產生高於單一模式的預報輸出 (Leith, 1974)，可以得到較為客觀之結果。近來氣候模式預報上長足進步，因此擁有不只單一模式的ENSO預報，然而因為有過多資訊，特別是有可能在不同模型上產生互相矛盾的結果，使得應用端的使用者無所適從。結果這種矛盾的現象方法之一：則是發展客觀方法來整合不同模型之間的預報結果。本計畫受惠於氣象局局內不只一組ENSO海溫統計預報模型的條件，期望使用綜合預報在綜多預報方法中，能夠更有效率與客觀的發法來整合資訊，而非主觀性較強的人工判斷。

本文採用貝氏模型平均法 (Bayesian Model Average, BMA) 應用於局內的三個ENSO預報方法：CCA、CA與CLIPER模式。選擇使用BMA的原因，是希望利用貝氏方法：利用歷史資料來模擬各模式的預報結果，進而推演量化各模式的預報機率來進行資料整合。本實驗的資料來自上述3個統計預報模式自1980年到2014年的4個海溫指標：N4、N3.4、N3、N1+2的歷史實際觀測資料來進行預報結果的探討。最後綜合預報結果的會與系集平均權重法 (Equally Weighted Ensemble, EWE) 來進行方法的比較。

關鍵字：貝氏模型平均法 (BMA)、綜合預報 (Consolidation Forecast)、ENSO

一、前言

近年知識的增加與電腦計算的增進，多數氣候預報都具備多於一組預報模式的能力。然而，陳列更多的預報模式所提供的資訊不但不一定能夠增加預報能力，還可能因為各模式不同的特性產生互相矛盾的預報結果，使得在應用端無所適從。所以預報仍然有一定程度是倚賴人工主觀來預報資訊的判斷，因此設法客觀的萃取不同的資訊並且進行資料整合是一個重要的研究方向，而綜合預報 (Consolidation Forecast) 則是回應此一問題產生的解決方案。

綜合預報是一個決定預報品值的數值過程。理論上，彼此假設與決定因子互相獨立的預報模式中，結合兩個或更多的預報模式可以產生較單一模式輸出高於平均得預報 (Leith, 1974)，進而取得更為客觀的預報。這樣的概念已經在氣候預報上有相當多的應用；例如，使用複迴歸分析來決定各模式的低階譜模式 (low-order spectral model) 係數來

進行資料整合 (Krishnamurti et al., 1999、2000)、或是使用簡易的平均權重的系集預報與藉由線性迴歸所得的係數來建立各個模式的權重 (Peng et al., 2002、Unger et al., 2008)、或是在不同模式不同初始條件下使用不同模式進行預測，來增加預測條件的寬度，增加比單一模式預測更可靠的機會 (O'Lenic et al., 2008)、或是將模型相對的預報值與事後校驗進行兩者間不確定性的量化 (Klopper and Landman, 2003)。

另一相當受到歡迎的貝氏 (Bayes) 方法，是近來統計資料後處理 (Statistical post processing) 領域中應用相當頻繁的方法之一。在本研究中即使用的貝氏模型平均 (Bayesian Model Average, BMA, Hoeting et al., 1999)，來預報海溫指標；其應用的範例包括，Raftery et al. (2005) 估計美國西部地面溫度與海平面氣壓、Wang et al. (2012) 使用於澳洲季雨量的網格化估計，另外還有臺灣測站溫度推估 (張宇軒 等 2011、馮智勇 等 2014)、風速與風向 (Liu and Xie, 2014) 的機率預報等。

本研究利用氣象局內的ENSO統計預報-CCA、CA、CLIPER模式-來建置BMA綜合預報方法來預報ENSO指標，並且使用EWE來比較兩個方法的預報能力。本文的規畫如下，在第2節簡介BMA方法後，接下來會簡單呈現利用BMA與EWE來跑不同條件的模擬資料的模擬結果，最後會將氣象局內近年來的ENSO統計預報資料進行BMA與EWE的實驗過程與預報結果進行比較，最後進行方法評析與研究遠景的建議。

二、BMA介紹

BMA的概念是利用貝氏定率的條件機率來給予每一個模式的不確定性量化的過程。其模式可以表示如下：

$$p(y) = \sum_{k=1}^K p(y|M_k) \cdot p(M_k|y^T)$$

，其中 y 是預報值、 M_k 為單一模式 $k = 1, \dots, K$ 、 y^T 為過往歷史觀測資料。 $p(y|M_k)$ 是基於模式 M_k 所得的條件機率函數； $p(M_k|y^T)$ 為藉由訓練資料而得 M_k 的後驗機率密度函數，並假設 $\sum_{k=1}^K p(M_k|y^T) = 1$ ，來使用微單一預報模式的權重。 $p(y)$ 可以被詮釋為每一個模式的權重總和，因此BMA也可以視為權重平均的條件機率函數。實際應用的假設命令如下。

命令：1. y 是預報值， y^T 則為訓練期歷史觀測資料；2. 總共 $k = 1, \dots, K$ 預報模式 M_k ；3. f_k 為預報模式 M_k 的預報值；4. $g_k(y|f_k)$ 為 y 在為 f_k 條件上的機率分布。所以BMA預報模式可以表示如下：

$$p(y|f_1, \dots, f_K) = \sum_{k=1}^K w_k \cdot g_k(y|f_k)$$

，而 $w_k = g(f_k|y^T)$ 是為第 k 個預報模式在資料的訓練期間為最佳預報模式的後驗機率；而且， $\sum_{k=1}^K w_k = 1$ 。

因為進行分析的預報實驗目標式NINO Index值 (N4、N3.4)，所以假設預報值為常態分配，所以條件機率函數平均數為 $a_k + b_k f_k$ ，變異量為 σ^2 ，所以

$$y|f_k \sim N(a_k + b_k f_k, \sigma^2)$$

。所以BMA預報模式的期望值，亦即預報值 y 的條件期望值，即為

$$E(y|f_1, \dots, f_K) = \sum_{k=1}^K w_k (a_k + b_k f_k)$$

。在這裡需要被估計的統計值為： a_k 、 b_k 、 w_k 與 σ^2 。

估計BMA模型的參數，首先估計 a_k 與 b_k 。假設 y_t 代表在時間 t 的NINO指標觀測值，

f_{kt} 代表系集內第 k 個預報模式在時間 t 的預報值。因為 $y_t|f_{kt}$ 的常態假設，以 y_t 與 f_{kt} 進行迴歸取得的 a_k 與 b_k 。取得的 a_k 與 b_k 還可以視為系集模式的偏估矯正。

關於估計 w_k 與 σ^2 ，使用的方法是利用歷史資料藉由對數最大似然估計 (LMLE) 估計，此方法的原理是將找到的估計參數，使得機率密度函數最佳化。BMA模型的LMLE可以表示如下，

$$\ell(w_1, \dots, w_K, \sigma) = \sum_t \log \left(\sum_{k=1}^K w_k \cdot g_k(y_t|f_{kt}) \right)$$

因為 $\ell(w_1, \dots, w_K, \sigma)$ 複雜的代數項，需要使用數值演算法來求解，因此在此使用EM演算法來估計 w_k 與 σ^2 。EM是為兩個階段—E(Expectation)階段與M(Maximization)階段的迭代演算法。起先先給予初始參數 θ 的估計向量， $\theta^{(0)} = (w_1^{(0)}, \dots, w_K^{(0)}, \sigma^{(0)})$ ，首先E階段，先定義 z_{kt} 為0或1，而在BMA中， $\hat{z}_{kt}^{(j)} = \frac{w_k g(y_t|f_{kt}, \sigma^{(j-1)})}{\sum_{i=1}^K w_i g(y_t|f_{it}, \sigma^{(j-1)})}$ ，是為第 j 階段的運算索引。而 $g(y_t|f_{kt}, \sigma^{(j-1)})$ 為 y_t 的常態分布的機率密度函數，其中平均數為 $a_k + b_k \cdot f_{kt}$ ，變異量為 $\sigma^{(j-1)}$ 。接下來是M階段，是從E階段所估計出的 $\hat{z}_{kt}^{(j)}$ 來估計 w_k 與 σ^2 ，而

$$w_k^{(j)} = \frac{1}{n} \cdot \sum_t \hat{z}_{kt}^{(j)}$$

$$\sigma^{2(j)} = \frac{1}{n} \sum_t \sum_{k=1}^K \hat{z}_{kt}^{(j)} (y_t - f_{kt})^2$$

其中， n 為訓練期的觀測資料長度。在E、M階段的交替中，直到LMLE收斂為止，意即 $|\ell^{(j)} - \ell^{(j-1)}| < \epsilon$ 。

另外簡述EWE，表示如下：

$$F = \sum_{k=1}^K w_k f_{M_k}$$

，其中 $w_k = \frac{1}{K}$ 賦予第 k 個預報模式 f_{M_k} 的權重，亦即將所有的模式賦予相通權重來取得EWE的系集預報值。

三、模擬資料實驗

在此進行模擬實驗來驗證方法的能力與模擬後的參考性。在此使用NINO3.4 值來進行BMA方法與平均權重的模擬比較。

因為假設海溫指標常態分布，模擬資料皆以NINO3.4為母體的季平均、標準差，如所表 1示，為常態隨機機制生成。在每一組模擬實驗，都進行5組不同的機率分布資料代表5組模式並使用BMA與EWE作系集預報結果；在這樣的機制下，每一組實驗都會進行1000

次的重覆模擬，來比較預報結果的絕對平均誤差值（MAE），最後統計MAE的分布來比較不同模式與BMA、EWE之間的結果。在這裡呈現兩組模擬資料¹中，一組代表模型估計結果有偏估，如圖 1所示、另一組將每一組模擬資料所代表模型估計偏估外，模型間估計能力也有所差異為代表來進行模擬，如圖 2所示。其結果簡述如下。

第一組模擬資料實驗的結果，如圖 3、表 2所示，雖然BMA所得MAE分布的變異量較EWE大，然而其MAE均值較EW低。第二組模擬資料實驗的結果，如圖 4、表 3所示，則是BMA得到較小的MAE均值，而MAE分布上BMA也較EME優。兩例中的模擬資料看出雖然EWE的MAE分布變異量較小，但是在模式彼此之間差異相當大的情況下，BMA與EWE的MAE差別已經有相當程度的接近。

大略的可以看出當模式之間有相當大的差異時，BMA的優勢將會比EWE好。然而，EWE仍然是一個可以參考的指標之一。接下來，則是實際應用統計ENSO預報方法來進行BMA、EWE的綜合預報比較。

四、實際ENSO預報應用實驗

本研究挑選N4與N3.4海溫指標觀測資料進行綜合預報BMA與EWE的實驗，另外的N3與N1+2海溫指標的預報實驗結果可以參考文獻[13]。

$$MAE_{\text{model}} = \frac{\sum_{t=1}^T (x_{\text{model}}(t) - x_{\text{obs}}(t))}{T}$$

方程式 1 絕對誤差平均(MAE)

$$\text{Corr}(x_{\text{model}}, x_{\text{obs}}) = \frac{\text{COV}(x_{\text{model}}, x_{\text{obs}})}{\sigma_{x_{\text{model}}} \sigma_{x_{\text{obs}}}}$$

方程式 2 相關係數(CORR)

$$RMSE_{\text{model}} = \sqrt{\frac{\sum_{t=1}^T (x_{\text{model}}(t) - x_{\text{obs}}(t))^2}{T}}$$

方程式 3 絕對均方根誤差(RMSE)

實驗資料時間的範疇自 1980年1月至 2014年12月的CCA、CA、CLIPER三個ENSO統計模式實際歷史預報資料。另外，預報分為4個前置時間（Lead Time, LT），其中4個LT的間隔皆為3個月，並利用MAE（方程式 1）、CORR（方程式 2）、RMSE（方程式 3）來測量個模式預報的表現能力。

BMA在估計參數過程中需要歷史資料「訓練」，原本氣象局使用的統計預報模式則是改原本使用50年長度的交叉驗證（cross validation）到30年，使得預報資料可以從2000年提前到1980年來增加預報資料長度，進行目標月份的ENSO預報。並在BMA應用上，每一個模式權重決定ENSO觀測值根據該模式的預報值條件機率，如第2節表示的 $p(y)$ ；模式權重為 $w_k = p(M_k|y^T)$ 來決定預報值 $p(y)$ 。以下為LT=1到LT=4的實驗結果的呈現。

如果實驗結果來看，從LT=1到4，CLIPER持續法的預測似乎都是最好的，CCA是最差的。如果分開LT討論來觀察，LT=1結果如下，參考表 4，BMA較EWE方法為佳，然而BMA類似系集權重平均的結果，使得BMA表現較CLIPER差。圖 5泰勒圖則可以從CORR、RMSE看到NINO3.4BMA的表現較接近CLIPER。

LT=2，參考表 5，雖然3個統計預報模式中CLIPER表現最好，但其MAE與RMSE都是LT=1的兩倍，其餘的模式表現差異都不大。如圖 6，在系集模式表現上，BMA與EWE差異已經不大。

LT=3，各個模式的預報能力皆較LT=2降低，如所表 6示。而圖 7可以看出BMA、CLIPER、EWE的表現在LT拉長後，表現越來越接近彼此。

LT=4，因為LT拉長後，各個模式的預報能力下降，預報得到的資訊也下降。LT=4與LT=3的差異已經不大，這些都可以從表 7圖 8中觀察到。

基本上，LT=1、2是預報資訊最充足的時間點，BMA方法會較EWE來得好；在LT拉長後，EWE會比BMA好，大致上BMA、EWE、CLIPER預報能力的指標是越來越接近的。

五、結論與建議

以不同的NINO指標觀察各模式在預報的LT增加表現，順序由CORR，再到MAE與RMSE-- 這些指標對於模式預報比較並探討預報能力優劣可能的原因。

參考CORR，如圖 9所示，LT增加，CORR降低。在系集模式表現上，N4指標預報則是當LT=1、2，BMA較EWE佳，而當LT=3之後，EWE高於BMA；N3.4指標預報下，BMA可能在統計容忍範圍之內上，是較EWE有競爭力。參考MAE，如圖 10所示，LT增加下，MAE皆為增加。然而在LT=1、2，BMA較EWE佳；LT=3、

¹在文獻[13]中有另外 3 組模擬資料分析結果可供參考

4, 兩個系集預報表現則顛倒過來。而LT越短, BMA小於EWE的MAE值就更為明顯。參考RMSE, 如圖 11所示, 趨勢與MAE類似。而且在LT增加後, CLIPER與BMA與EWE的預報能力接近。

模擬實驗與ENSO預報應用實驗的結果, 可以觀察出BMA與EWE的差異。探討BMA預報能力, 首先當資訊較為明朗時, BMA較EWE有優勢。以此推論, BMA方法可能最好的應用條件在有更多模式同步使用, 並且各模式之間的差異極大的情況下, 如模擬實驗第2組(模式間預報上有偏估, 並且變異量差異大)所呈現的。第二, LT較短時, BMA模式 $a_k + b_k f_k$ 的假設, 會因為觀測值與預報值的CORR不高產生問題, 如CCA模式所產生的預報值與觀測值的低相關性, 它並未能在預報中提供該給予的資訊給BMA。因著BMA強烈的倚賴模式在觀測條件機率密度函數是否能精確描述觀測值的分布, 其預報能力會受到相當的影響。第三, 本實驗有一些ENSO預報上的困難點並未排除, 例如春季預報障礙, 以及聖嬰反聖嬰年的海溫變化, 這些問題可能可以獨立出來進行階層的分析, 讓這樣的資訊混和都是可能讓BMA方法的應用產生窒礙的原因之一。另外, EWE與BMA預報能力旗鼓相當, 主要是因為CLIPER優異的預報能力。

最原始的BMA方法是為要幫助量化模式與生俱來的不確定性, 在此實驗中可以看出BMA在這一方面的特質。在未來綜合預報的研究上, 能夠加入動力模式的預報資料與統計模式一起進行資訊的整合應該是未來實際操作的重點。另外, 在非常態分布的氣候資料, 如雨量上的應用, 雖然會與一些常態分布的變數有非常不一樣的數值方法應用, 但是理念相近之下, 也是在近期可以馬上進行操作的研究方向。

六、參考文獻

1. Leith, C.E., 1974: Theoretical skill on Monte Carlo Forecasts. *Mon. Wea. Rev.*, V.102, N6, 409-418
2. Krishnamuri, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang and co-authors, 1999: Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* 285, 1548-1550
3. Krishnamuri, T. N., C. M. Kishtawal, D. W. Shin, C. E. Williford, 2000: Multi-model

superensemble forecasts for weather and seasonal climate. *J. Climate*, vol. 13, 4196-4216

4. Peng, P., A. Kumar, H. M. van den Dool, and A. G. Barnston, 2002: An analysis of multimodel ensemble prediction for seasonal climate anomalies. *J. Geophys. Res.*, 107(D23), 4710
5. Unger, D.A., H. van den Dool, E.O' Lenic, and D. Collins, 2008: "Ensemble regression." *Mon. Wea. Rev.* 137, no.7, 2365-2379
6. O' Lenic, E.A., D. A. Unger, M. S. Halpert and K. S. Pelman, 2008: Developments in operational long-range climate prediction at CPC. *Wea. Forecasting*, vol. 23, 496-515
7. Klopper, E. and W.A. Landman, 2003: A simple approach for combining seasonal forecasts for southern Africa. *Met. Apps*, 10:319-327
8. Raftery, A.E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, 133, 1155-1174
9. Hoeting, J.A., A.E. Raftery and D. Madigan, 2002: A Method for Simultaneous Variable and Transformation Selection in Linear Regression. *Journal of Computational and Graphical Statistics* 11 (485-507)
10. 張語軒、張庭槐與吳蕙如, 2011: "貝式模型平均於溫度機率預報之應用"。 *氣象學報*, 49(1), 19-38
11. 馮智勇、劉家豪、林佑蓉、蔡雅婷、陳雲蘭, 2014: "應用BMA發展多模式MOS策略溫度機率預報", 交通部氣象局
12. Liu, J. and Z. Xie, 2014: BMA probabilistic quantitative precipitation forecasting over the Huaihe Basin using TIGGE multi-model ensemble forecasts. *Mon. Wea. Rev.*, 142, 1542-1555
13. 104年度氣候模式及應用作業委託案—災害性天氣監測與預報作業建置計畫--綜合預報方法ENSO統計預報模式應用期末報告

	DJF	FMA	AMJ	ASO	OND
--	-----	-----	-----	-----	-----

μ	-.19	-.19	-.18	-.2	-.17
σ	.96	.67	.54	.80	.99

表 1 參考 1950 年 1 月至 2015 年 9 月的 NINO3.4 指標的季平均與標準差所生成的模擬資料。

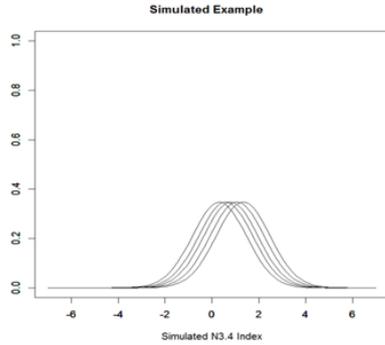


圖 1 代表偏估與相同變異量的估計模式第一組模擬資料的密度函數。

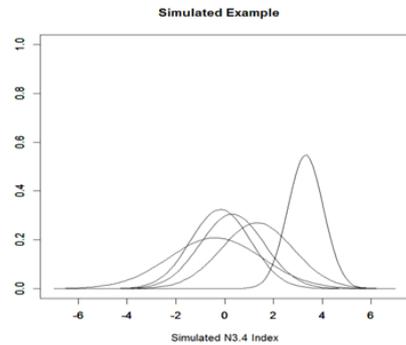


圖 2 代表偏估與不同變異量的估計模式第二組模擬資料的密度函數。

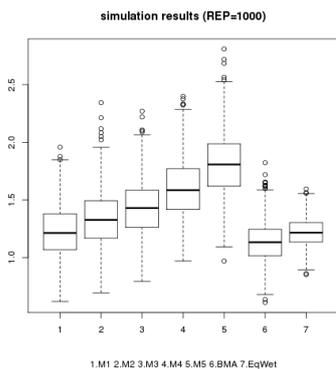


圖 3 第一組模擬資料 1000 次實驗所生成的 1000 組 EMA 箱型圖。

M1	M2	M3	M4	M5	BAM	EW
1.2	1.3	1.4	1.6	1.8	1.14	1.22
3	4	3	0	1		

表 2 第一組模擬資料 1000 次實驗得到 1000

組 EMA 平均值。

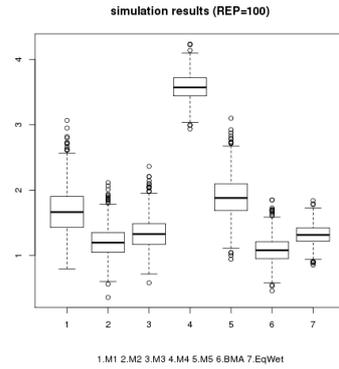


圖 4 第二組模擬資料 1000 次實驗所生成的 1000 組 EMA 箱型圖。

M1	M2	M3	M4	M5	BAM	EW
1.6	1.2	1.3	3.5	1.9	1.09	1.32
7	1	4	7	0		

表 3 第二組模擬資料 1000 次實驗得到 1000 組 EMA 平均值。

N4(3.4)	CCA	CA	CLIPER	BMA	EW
MAE	.56(.74)	.32(.38)	.16(.19)	.23(.24)	.31(.37)
COR	.13(.11)	.76(.80)	.95(.95)	.85(.91)	.82(.83)
RMSE	.71(.96)	.4(.48)	.19(.24)	.32(.33)	.38(.47)

表 4 LT=1，ENSO 指標對應個預報模式的 MAE、COR 與 RMSE 表。

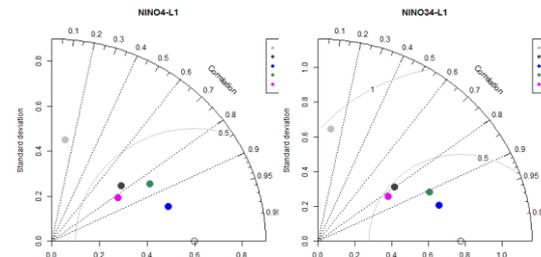


圖 5 LT=1 五個模式的泰勒圖，左邊與右邊分別為 ENSO 指標 N4、N3.4；灰色、黑色、藍色、綠色、桃色分別代表 CCA、CA、CLIPER、BMA、EW 模式。

N4(3.4)	CCA	CA	CLIPER	BMA	EW
MAE	.56(.75)	.39(.48)	.29(.37)	.35(.44)	.37(.45)
COR	.13(.12)	.62(.65)	.80(.80)	.71(.71)	.69(.70)
RMSE	.70(.96)	.48(.60)	.37(.47)	.43(.56)	.45(.57)

表 5 LT=2 各模式表現，細節參考表 4。

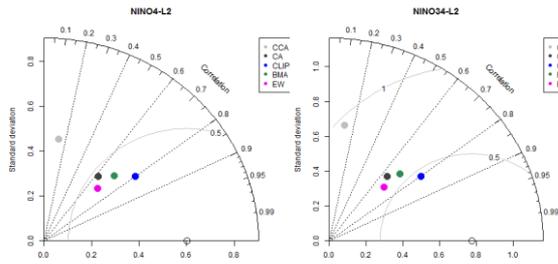


圖 6 LT=2 五個模式的泰勒圖，細節參考圖 5。

N4(3,4)	CCA	CA	CLIPER	BMA	EW
MAE	.55(.78)	.45(.57)	.40(.52)	.44(.55)	.42(.54)
COR	.16(.10)	.49(.47)	.61(.59)	.50(.51)	.56(.52)
RMSE	.71(1.0)	.55(.73)	.50(.65)	.54(.70)	.51(.68)

表 6 LT=3 各模式表現，細節參考表 4

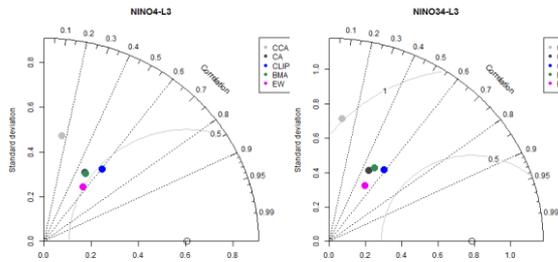


圖 7 LT=3 五個模式的泰勒圖，細節參考圖 5。

N4(3,4)	CCA	CA	CLIPER	BMA	EW
MAE	.56(.80)	.49(.64)	.44(.59)	.47(.61)	.45(.60)
COR	.14(.05)	.37(.30)	.50(.45)	.42(.39)	.45(.36)
RMSE	.72(1.0)	.60(.81)	.56(.75)	.58(.77)	.55(.76)

表 7 LT=4 各模式表現，細節參考表 4

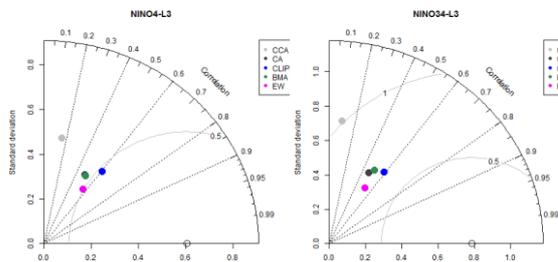


圖 8 LT=4 五個模式的泰勒圖，細節參考圖 5。

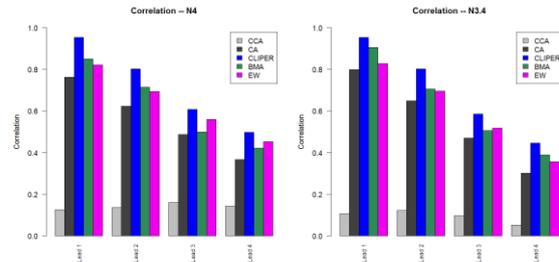


圖 9 CORR 比較模式：左 N4 右 N3.4 在 LT=1 到 4 模式之間的優劣表現。(各模式代表的顏色請參考圖 5)

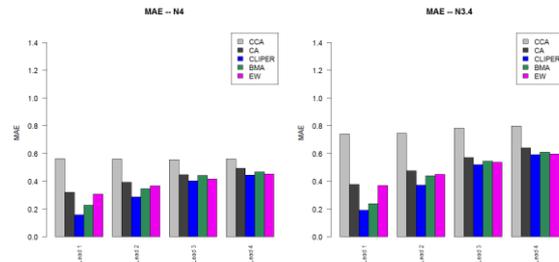


圖 10 MAE 比較模式：左 N4 右 N3.4 在 LT=1 到 4 模式之間的優劣表現。(各模式代表的顏色請參考圖 5)

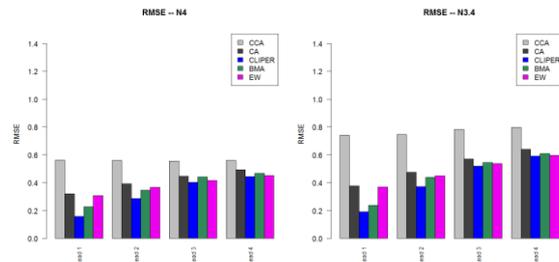


圖 11 RMSE 比較模式：左 N4 右 N3.4 在 LT=1 到 4 模式之間的優劣表現。(各模式代表的顏色請參考圖 5)