

**PHÂN TÍCH, ĐÁNH GIÁ MỘT SỐ PHƯƠNG PHÁP THỐNG KÊ
HIỆU CHỈNH SAI SỐ TỪ MÔ HÌNH MƯA NGÀY VỀ TRẠM MƯA -
ỨNG DỤNG CHO CÁC TRẠM MƯA THUỘC TỈNH BÌNH ĐỊNH**

Ngô Lê An¹, Lê Thị Hải Yến¹, Ngô Lê Long¹, Nguyễn Thị Thu Hà¹

Tóm tắt: Các mô hình khí hậu toàn cầu hoặc khu vực đang được sử dụng rộng rãi để mô phỏng các thông tin khí tượng trên một phạm vi không gian dù kết quả mô phỏng lượng mưa từ các mô hình này vẫn còn các sai số. Để sử dụng hiệu quả số liệu, có nhiều kỹ thuật hiệu chỉnh sai số thống kê đã được nghiên cứu ứng dụng. Trong bài báo này tám phương pháp hiệu chỉnh đại diện cho ba nhóm biến đổi dựa vào phân bố xác suất lý thuyết, biến đổi có tham số và biến đổi phi tham số được đưa vào đánh giá. Chỉ tiêu bình quân sai số tuyệt đối (MAE) được sử dụng để xếp hạng các phương pháp được tính toán từ phương pháp đánh giá chéo (cross-validation). Kết quả cho thấy nhóm biến đổi phi tham số cho hiệu quả hiệu chỉnh sai số tốt nhất cả về phân bố cường độ mưa lẫn số ngày có mưa, trong khi nhóm biến đổi dựa trên phân bố xác suất lý thuyết kém hiệu quả nhất. Nghiên cứu cũng chỉ ra rằng tại vị trí các trạm đo mưa có lượng mưa ngày lớn sẽ cho kết quả tính toán sai số nhiều nhất do khả năng ngoại suy các giá trị cực trị của các phương pháp hiệu chỉnh vẫn còn hạn chế.

Từ khóa: Thống kê hiệu chỉnh sai số, đánh giá chéo, mô hình mưa ngày.

1. ĐẶT VẤN ĐỀ

Các mô hình khí hậu toàn cầu (Global Climate Model – GCM) hoặc khu vực (Regional Climate Model – RCM) hiện đang được sử dụng rộng rãi để mô phỏng các thông tin khí tượng trên một phạm vi không gian. Nhiều nghiên cứu đã chỉ ra sai số trong kết quả mô phỏng lượng mưa từ các mô hình. Các sai số này xuất hiện thường là do những giả thiết làm đơn giản hoá cũng như những giới hạn trong mô hình số khi mô phỏng quá trình hình thành mưa (Rauscher và nnk, 2010). Do vậy, để sử dụng hiệu quả số liệu mưa từ mô hình, cần xử lý sai số từ các kết quả mô phỏng. Hiện nay, các kỹ thuật xử lý khác nhau được nghiên cứu nhiều nhằm đưa ra các kết quả được hiệu chỉnh phù hợp nhất với

các số liệu quan trắc tại các trạm đo mưa. Về cơ bản, kỹ thuật hiệu chỉnh sai số dựa trên việc hiệu chỉnh các đặc trưng thống kê chuỗi số liệu tính toán để phù hợp với các đặc trưng thống kê của chuỗi quan trắc. Các đặc trưng này có thể bao gồm trung bình, độ lệch chuẩn, hàm phân bố... (Maraun Douglas, 2016). Hai nhóm hiệu chỉnh chính thường được sử dụng là: i) Thay đổi tỷ lệ; và ii) Hiệu chỉnh Phân vị. Nhóm thay đổi tỷ lệ là nhóm phương pháp đơn giản được tính toán hiệu chỉnh dựa trên chênh lệch giữa giá trị thực đo và tính toán. Nhóm hiệu chỉnh phân vị thì cố gắng hiệu chỉnh sự biến thiên của phân bố từ mô hình sao cho phù hợp với sự biến thiên của số liệu thực đo. Nhìn chung, các kỹ thuật này vẫn còn tồn tại việc phải giả thiết mối quan hệ giữa các biến độc lập và phụ thuộc là không đổi trong tương lai (Wilby và Wigley, 1997).

¹ Trường Đại học Thủy lợi.

Một số nghiên cứu điển hình liên quan đến các kỹ thuật này có thể kể đến: Wilby và nnk (2002) đã giới thiệu mô hình chi tiết hóa thống kê (Statistical Down Scaling Model -SDSM) nhằm chi tiết hoá lượng mưa từ mô hình toàn cầu về từng trạm bằng cách xây dựng quan hệ tương quan giữa lượng mưa và các đặc trưng khí hậu của vùng trong mô hình. Công cụ có ưu điểm tính toán nhanh chóng, không đòi hỏi nhiều tài nguyên máy tính nhưng nhược điểm cần có nhiều dữ liệu về các yếu tố khí tượng khác. Lenderink và nnk (2007), Schimidli và nnk (2006) sử dụng phương pháp thay đổi tỷ lệ kết hợp với việc hiệu chỉnh số ngày mưa. Ines và Hansen (2006) đã nghiên cứu hiệu chỉnh sai số lượng mưa ngày từ mô hình mưa toàn cầu. Các tác giả nhận thấy, kết quả mô phỏng thường cho quá nhiều trận mưa xảy ra kèm theo cường độ mưa khá nhỏ nếu so với lượng mưa tại các trạm đo nằm trong các ô lưới mô phỏng tương ứng, từ đó đã đề xuất phương pháp hiệu chỉnh cả phân bố tần suất và cường độ mưa cho lượng mưa ngày từ mô hình mưa toàn cầu. Piani và nnk (2010) đã đề xuất một hàm biến đổi, hiệu chỉnh hàm phân bố lượng mưa mô phỏng lũy tích về hàm phân bố xác suất mưa thực đo, sử dụng hàm phân bố Gamma để mô phỏng phân bố lượng mưa ngày. Theo Haerter và nnk (2015), nhóm phương pháp phân vị nhằm điều chỉnh hàm phân bố xác suất của kết quả đầu ra của mô hình khí hậu với dữ liệu khí hậu quan trắc thực tế được ứng dụng nhiều nhất trong nhóm các phương pháp hiệu chỉnh thống kê. Bài báo sẽ nghiên cứu đánh giá một số phương pháp thống kê hiệu chỉnh phân vị thường dùng từ mô hình mưa ngày với dữ liệu tại trạm đo mưa, từ đó phân tích và khuyến nghị lựa chọn phương pháp phù hợp. Vùng nghiên cứu là các lưu vực sông thuộc tỉnh Bình Định do đây là một trong những vùng dự kiến có biến động về mưa lớn và có mật độ trạm đo mưa khá cao (Bộ Tài nguyên và Môi trường, 2016).

2. PHƯƠNG PHÁP NGHIÊN CỨU VÀ DỮ LIỆU

2.1. Phương pháp nghiên cứu

Theo Piani và nnk (2010), hàm biến đổi thống kê có dạng:

$$P_o = F_o^{-1}(F_m(P_m)) \quad (1)$$

Với P_o , P_m lần lượt là lượng mưa thực đo và lượng mưa tính toán từ mô hình. Hàm F_m ở đây là hàm phân bố lũy tích của lượng mưa tính toán. F_o^{-1} là hàm nghịch đảo của hàm phân bố lũy tích tương ứng với lượng mưa thực đo.

Dựa vào các phương pháp xác định hàm biến đổi khác nhau, phương pháp thống kê hiệu chỉnh sai số được đánh giá trong nghiên cứu này được chia thành ba nhóm như sau:

- Nhóm biến đổi dựa vào phân bố xác suất lý thuyết: các phân bố Gamma, Weibull, Log-Normal, hàm mũ (Exponential). Thông số của các hàm xác suất được tính toán theo phương pháp thích hợp tối đa cho cả giá trị P_m và P_o .

- Nhóm biến đổi có tham số:

- o Tỷ lệ (QQ_Scale): $\bar{P}_o = bP_m$ (với b là tham số)

- o Tương quan tuyến tính (QQ_Linear): $\bar{P}_o = a + bP_m$ (với a và b là tham số)

- o Tương quan lũy thừa (QQ_Power): $\bar{P}_o = bP_m^c$ (với b và c là tham số)

Trong đó \bar{P}_o là giá trị tính toán được hiệu chỉnh. Các tham số sử dụng trong nhóm biến đổi này được tìm kiếm bằng cách so sánh đường tần suất lũy tích kinh nghiệm với hàm tính toán sử dụng đánh giá sai số bình phương tối thiểu.

- Nhóm biến đổi phi tham số (EQ_Linear): hàm phân vị thực nghiệm với các giá trị phân vị được nội và ngoại suy theo hồi quy tuyến tính.

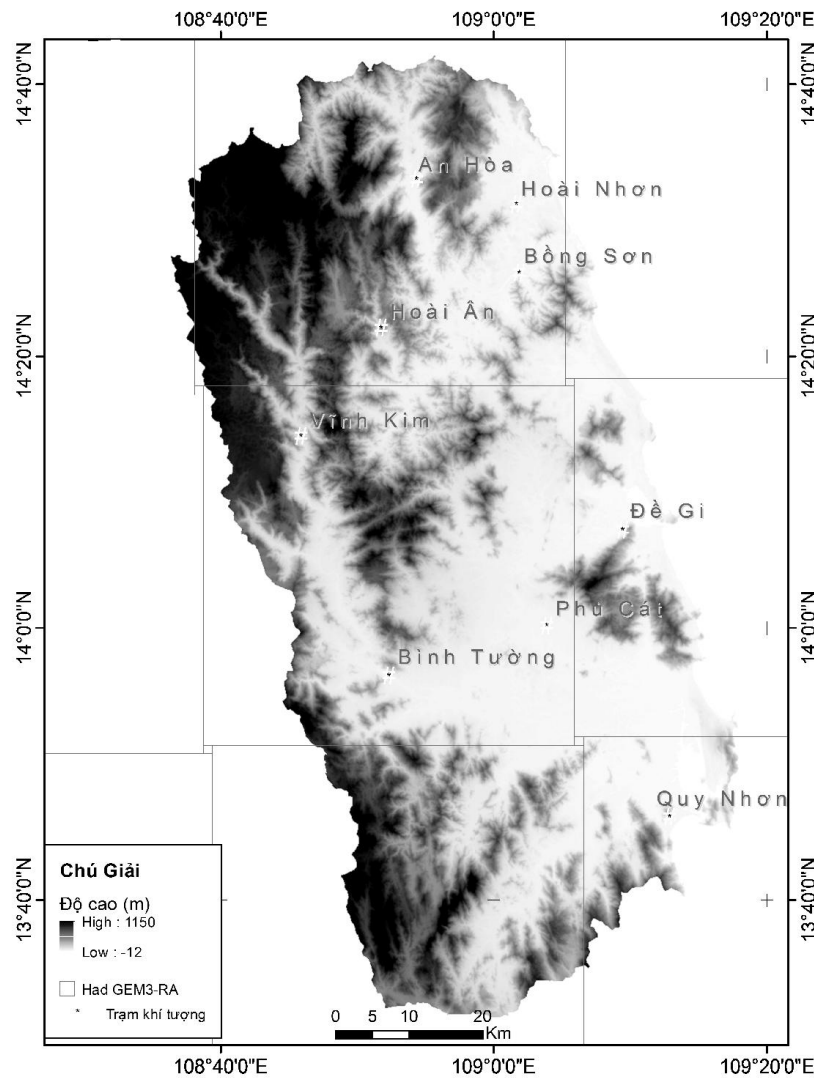
2.2. Dữ liệu

Dữ liệu mưa mô hình được lấy từ mô hình khí hậu địa phương HadGEM3-RA (RCM) có kích thước mô phỏng nhỏ hơn là $0,44^\circ \times 0,44$ (xấp xỉ 50km) với các biên đầu vào từ mô hình HadGEM2-AO của Anh có kích thước lưới là $1,875^\circ \times 1,25^\circ$ cho các vùng Đông Á, Ấn Độ và

Tây Thái Bình Dương. Mô hình này đã mô phỏng lại chuỗi mưa ngày trong quá khứ từ năm 1960 đến 2005. Trong nghiên cứu này, chuỗi dữ liệu thực đo giai đoạn quá khứ từ năm 1981 đến 2005 được sử dụng để so sánh đánh giá với chuỗi dữ liệu mô phỏng từ mô hình. Dữ liệu thực đo được lấy từ 9 trạm đo mưa được trình bày ở bảng 1. Dữ liệu mưa từ mô hình sử dụng để so sánh với từng trạm đo sẽ được lấy từ ô lưới chứa trạm đo mưa đó.

Bảng 1. Các trạm đo mưa trong nghiên cứu

TT	Tên trạm	Kinh độ	Vĩ độ
1	An Hoà	108.9080	14.5480
2	Bồng Sơn	109.0333	14.4333
3	Đề Gi	109.1597	14.1180
4	Hoài Ân	108.8640	14.3655
5	Phù Cát	109.0667	14.0000
6	Vĩnh Kim	108.7667	14.2333
7	Hoài Nhơn	109.0300	14.5168
8	Quy Nhơn	109.2167	13.7667
9	Bình Tường	108.8743	13.9395



Hình 1. Sơ đồ vị trí các trạm nghiên cứu và ô lưới mô hình HadGEM3-RA cho khu vực tỉnh Bình Định

2.3 Chỉ tiêu đánh giá

Để đánh giá khả năng hiệu chỉnh sai số giữa các phương pháp, nghiên cứu này sử dụng chỉ

số bình quân sai số tuyệt đối (Mean Absolute Error – MAE) được tính trong từng khoảng không gian xác suất giữa phân bố kinh nghiệm

và tính toán theo đề xuất của Gudmundsson và nnk (2012). Các giá trị MAE0.1, MAE0.2, ..., MAE1.0 tương ứng là bình quân sai số tuyệt đối trong khoảng xác suất từ 0 đến 0,1; 0,1 đến 0,2;...; 0,9 đến 1,0. Các chỉ số này được xác định bằng kỹ thuật đánh giá chéo (Cross-validation). Theo đó, chuỗi số được chia thành 10 mẫu khác nhau theo thời gian. Mô hình sẽ tính toán hiệu chỉnh trên 9 mẫu và kiểm định ở mẫu còn lại. Tính toán chỉ số MAE và lặp lại lần lượt các bước hiệu chỉnh kiểm định này cho từng mẫu. Chỉ số MAE cuối cùng chính là chỉ số tính trung bình trong các bước kiểm định. Phương pháp hiệu chỉnh nào cho chỉ số MAE nhỏ hơn sẽ có nghĩa sai lệch giữa tính toán và thực đo ít hơn, hay nói cách khác là phương pháp đó tốt hơn.

Ngoài ra, số ngày có mưa cũng là một chỉ tiêu đánh giá trong nghiên cứu này. Ngày có mưa được coi là ngày có lượng mưa lớn hơn hoặc bằng 0.1mm.

3. KẾT QUẢ NGHIÊN CỨU

Chỉ số MAE được tính toán cho 9 trạm nghiên cứu theo các phương pháp khác nhau. Kết quả tính toán được trình bày ở Hình 2 và Hình 3.

Trong Hình 2, đường nét liền ký hiệu SIM thể hiện sai số của mô hình trước khi hiệu chỉnh, các điểm chấm tương ứng là các phương pháp hiệu chỉnh khác nhau. Kết quả cho thấy nhìn chung các phương pháp hiệu chỉnh có tác dụng làm giảm sai số kể cả tính trung bình hay theo phân bố từng khoảng. Các sai số lớn chủ yếu xuất hiện ở phần nửa trên của không gian xác suất, phần nửa dưới ít sai số do chủ yếu là những ngày không có mưa. Các dạng hàm phân bố như hàm mũ (Exponential), Gamma,

LogNormal kém hiệu quả hơn đối với khoảng tần suất cực đại (MAE1.0). Trong khi các hàm biến đổi có tham số như Tỷ lệ, Tuyến tính và Luỹ thừa thì lại kém hiệu quả ở các khoảng phân bố gần cực đại (MAE0.7 đến MAE0.9) cho một số trạm. Đặc biệt ở MAE0.9 thì các hàm này không có hiệu quả khi hiệu chỉnh sai số nếu xét trung bình của cả 9 trạm trên toàn khu vực. Nhóm biến đổi phi tham số cho kết quả hiệu chỉnh tốt nhất khi ở tất cả các khoảng phân bố trên cả 9 trạm đều cho kết quả bình quân sai số tuyệt đối thấp hơn so với số liệu mưa mô hình chưa được hiệu chỉnh. Lợi thế của phương pháp biến đổi phi tham số là tính linh hoạt trong việc mô tả phân phối xác suất khi nó không phụ thuộc vào bất cứ hàm tính toán nào.

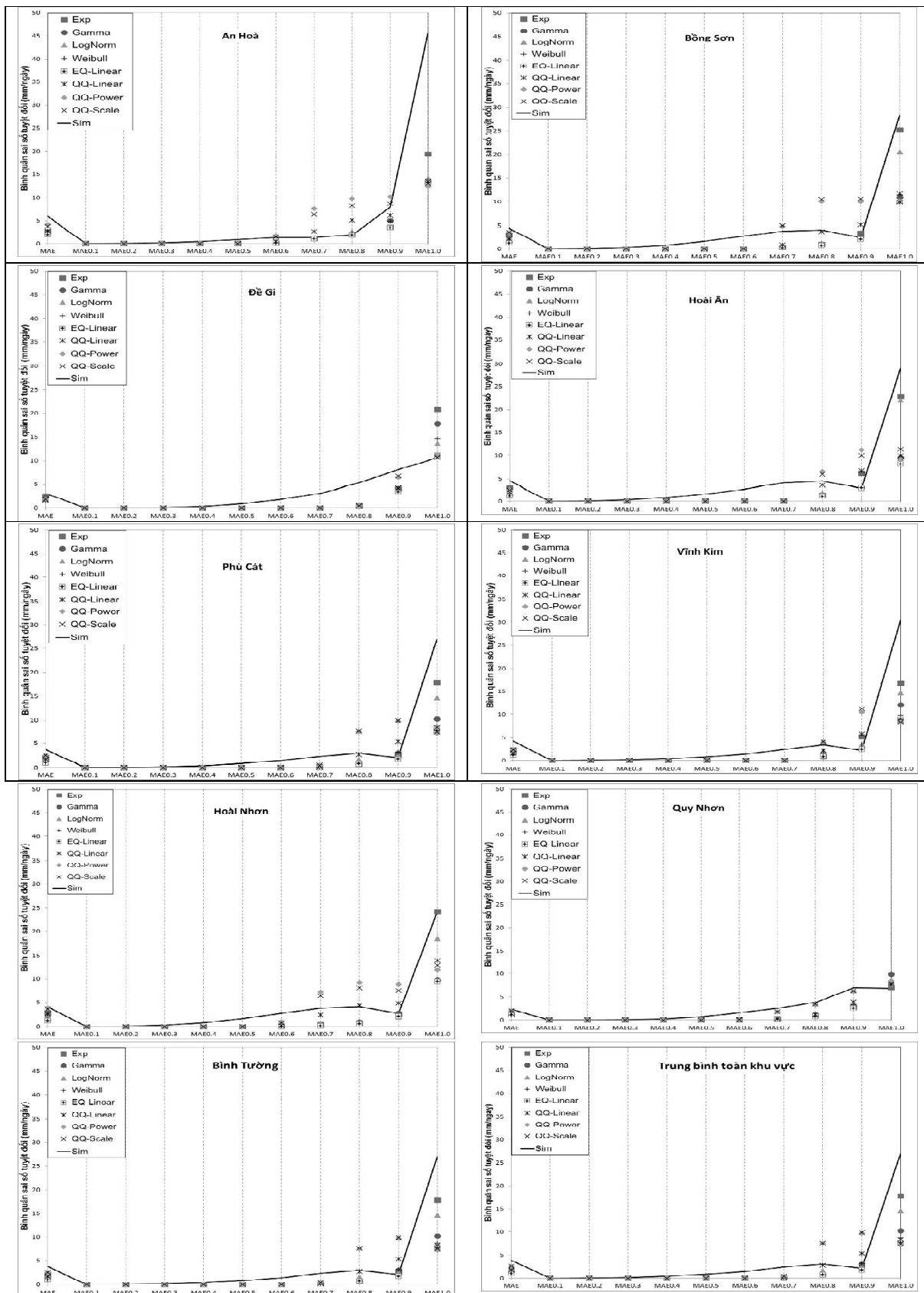
Hình 3 thể hiện phân bố của các bình quân sai số tuyệt đối theo không gian ứng với các dạng hàm hiệu chỉnh khác nhau. Nhìn chung, các trạm đo mưa phía Đông khu vực giáp biên như Bồng Sơn, Đề Gi, Hoài Nhơn thường có giá trị MAE cao hơn so với các trạm đo mưa nằm ở khu vực phía Tây trên địa phận lưu vực sông Kone. Kể cả khi chuẩn hoá giá trị MAE bằng cách chia cho lượng mưa trung bình thì các trạm này cũng cho giá trị cao hơn. Điều này có thể giải thích là do các trạm này gần biển... có chế độ mưa phức tạp hơn với lượng mưa cực trị lớn, biến động mạnh nên khó hiệu chỉnh hơn. Khi tính toán cho các trường hợp cực trị, các phương pháp tính toán cần phải ngoại suy dựa trên bộ số liệu hiệu chỉnh. Nếu các phương pháp hiệu chỉnh trên bộ số liệu này mô tả xu thế không phù hợp sẽ dẫn đến kết quả hiệu chỉnh trên bộ số liệu kiểm định sẽ có sai số lớn.

Bảng 2. Chênh lệch số ngày có mưa sau khi hiệu chỉnh so với thực tế (%)

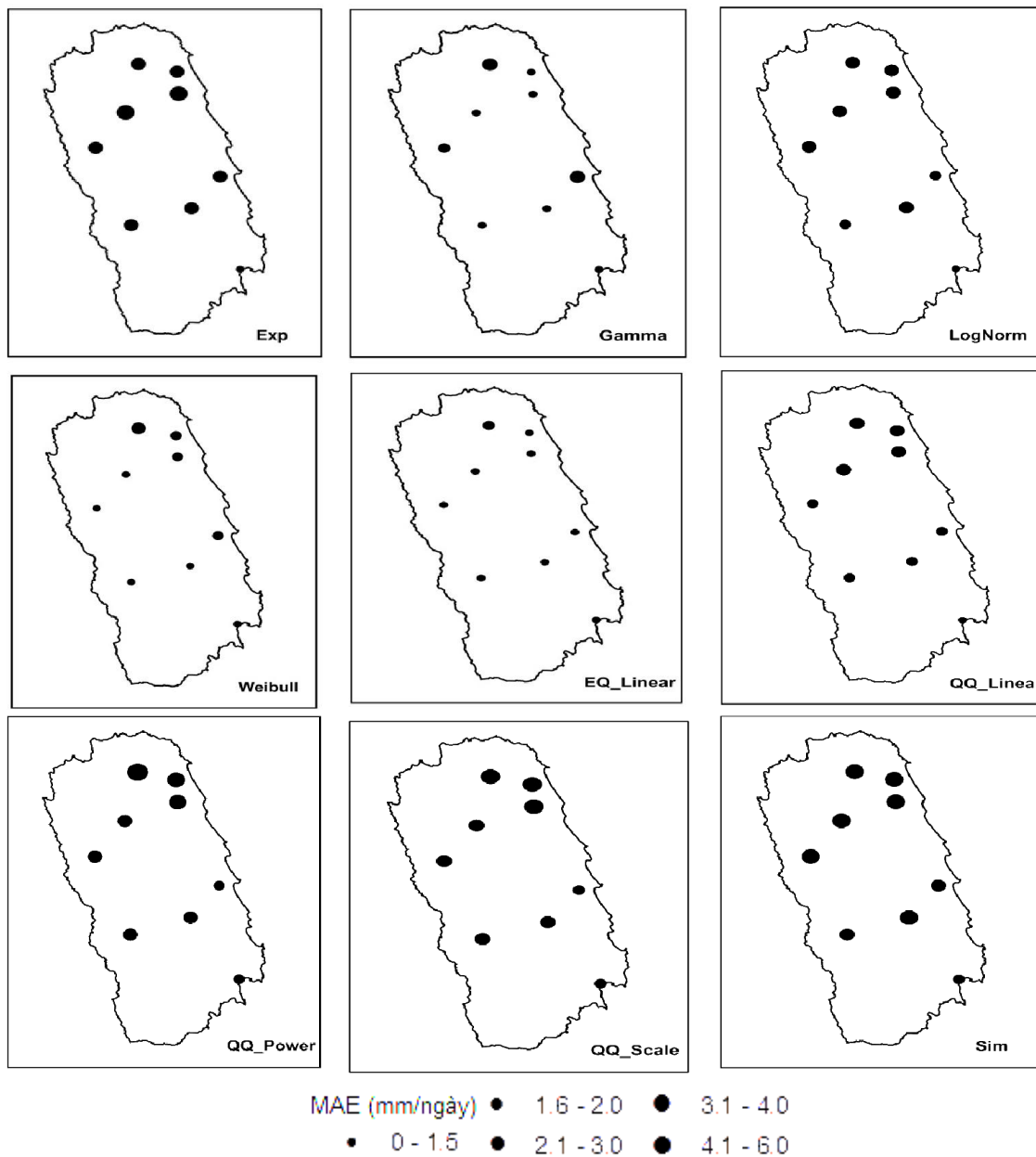
Exp	Gamma	LogNorm	Weibull	EQ-Linear	QQ-Linear	QQ-Power	QQ-Scale	Không hiệu chỉnh
6.97	15.01	11.94	6.15	0.83	3.55	0.67	0.67	280

Bảng 2 thể hiện chênh lệch số ngày có mưa theo các phương pháp hiệu chỉnh so với số liệu thực tế. Trước khi hiệu chỉnh, kết quả mô phỏng từ mô hình khí hậu cho số ngày có mưa rất nhiều (gấp 2,8 lần so với thực tế). Sau khi sử dụng các phương pháp hiệu chỉnh, nhìn chung, số ngày có mưa trong kết quả của mô

hình đã được điều chỉnh về xấp xỉ với thực tế. Phương pháp hiệu chỉnh phi tham số vẫn cho sai lệch ít nhất cùng với nhóm hiệu chỉnh có tham số là tương quan tuyến tính và tương quan luỹ thừa, thể hiện tính hiệu quả của phương pháp này trong việc hiệu chỉnh sai số của các mô hình mưa.



Hình 2. Bình quân sai số tuyệt đối của các trạm theo các phương pháp



Hình 3. Bình quân sai số tuyệt đối (MAE) theo các hàm hiệu chỉnh khác nhau tại các trạm đo mưa thuộc tỉnh Bình Định

4. KẾT LUẬN

Bài báo đã nghiên cứu sử dụng 8 phương pháp hiệu chỉnh sai số chia thành 3 nhóm: hàm phân bố xác suất, biến đổi có tham số và biến đổi phi tham số cho 9 trạm đo mưa thuộc tỉnh Bình Định từ mô hình mưa HadGEM3-RA. Báo cáo sử dụng phương pháp Cross-validation để đánh giá hiệu quả các phương pháp này. Kết quả nghiên cứu cho thấy, nhóm phương pháp biến đổi phi tham số có hiệu quả cao nhất cả về hiệu chỉnh cường độ mưa lẫn

số ngày có mưa. Các phương pháp thuộc nhóm hàm phân bố xác suất cho hiệu quả thấp nhất vì các hàm phân bố khó mô tả chính xác biến động mưa ngày. Các hàm phân phối xác suất có nhiều thông số hơn cho kết quả tốt hơn các hàm ít thông số vì nó dễ mô tả tần suất kinh nghiệm chính xác hơn. Điều này càng thể hiện rõ khi xét trên không gian các trạm đo, các sai số lớn thường ở các trạm đo mưa gần biển nơi có lượng mưa ngày lớn kèm theo sự biến động của nó cao theo thời gian.

TÀI LIỆU THAM KHẢO

- Bộ Tài nguyên và Môi trường (2016). *Kịch bản Biến đổi khí hậu và nước biển dâng cho Việt Nam*.
- Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: *Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods*, *Hydrol. Earth Syst. Sci.*, 16, 3383-3390, doi:10.5194/hess-16-3383-2012, 2012.
- Haerter, J. O., B. Eggert, C. Moseley, C. Piani and P. Berg (2015), *Statistical precipitation bias correction of gridded model data using point measurements*, *Geophys. Res. Lett.*, 42, 1919 - 1929, doi:10.1002/2015GL063188.
- Ines AVM, Hansen JW (2006). *Bias correction of daily GCM rainfall for crop simulation studies*. *Agric For Meteorol* 138:44-53.
- Lenderink, G., Buishand, A., and van Deursen, W. (2007): *Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach*, *Hydrol. Earth Syst. Sci.*, 11, 1145–1159, doi:10.5194/hess-11-1145-2007.
- Maraun, D. (2016). *Bias Correcting Climate Change Simulations - a Critical Review*. *Current Climate Change Reports*, 2, 211-220.
- Rauscher, S., Coppola, E., Piani, C., and Giorgi, F., (2010). *Resolution effects on regional climate model simulations of seasonal precipitation over Europe*, *Clim. Dynam.*, 35, 685–711, doi:10.1007/s00382-009-0607-7, 2010.
- Schmidli, J., Frei, C., and Vidale, P. L. (2006): *Downscaling from GC precipitation: A benchmark for dynamical and statistical downscaling methods*, *Int. J. Climatol.*, 26, 679–689, doi:10.1002/joc.1287.
- Wilby, R.L., Wigley, T.M.L, (1997). *Downscaling general circulation model output: a review of methods and limitations*. *Progress in Physical Geography* 214, 530-548.
- Wilby, R.L., Dawson, C.W. and Barrow E.M., (2002). *SDSM - a decision support tool for the assessment of regional climate change impacts*, *Environmental Modelling & Software*, 17, 147-159.

Abstract:

SOME STATISTICAL TRANSFORMATION METHODS FOR BIAS CORRECTION OF DAILY PRECIPITATION FROM METEOROLOGICAL MODELS TO THE STATION SCALE – A CASE STUDY IN BINH DINH PROVINCE

Global Climate Model and Regional Climate Model are widely used to simulate regional climate despite large errors of models. Some bias correction techniques are applied to get more accuracy results. This research studies some common bias correction methods for daily precipitation to the station scale, a case study in Binh Dinh province. Eight methods of three Transformation groups: distribution derived transformations, parametric transformations, nonparametric transformations are selected for review. Mean absolute error index estimated from Cross-validation technique is used for ranking the methods. The nonparametric transformations is ranked as the best method in reducing biases of both of precipitation intensities and wet days. Distribution derived transformations is less effective. Large errors are occurred in the stations which have extreme precipitation due to the limitations of extrapolation of these techniques.

Keywords: Bias correction, cross-validation, Binh Dinh province, daily precipitation model...

BBT nhận bài: 16/3/2017

Phản biện xong: 09/4/2017