



## Penerapan Metode *Bayesian Model Averaging* untuk Kalibrasi Prediksi *Ensemble* Curah Hujan Bulanan di Pulau Jawa

Implementation of Bayesian Model Averaging Method to Calibrate Monthly Rainfall Ensemble Prediction over Java Island

**Robi Muharsyah<sup>1,2\*</sup>, Tri Wahyu Hadi<sup>2</sup>, Spto Wahyu Indratno<sup>3</sup>**

<sup>1</sup> Badan Meteorologi Klimatologi dan Geofisika, Jl. Angkasa I No.2 Jakarta 10720

<sup>2</sup> Program Studi Sains Kebumihan, FITB, Institut Teknologi Bandung, Jl. Ganesha No.10, Bandung 40132

<sup>3</sup> Program Studi Matematika, FMIPA, Institut Teknologi Bandung, Jl. Ganesha No.10, Bandung 40132

### ARTICLE INFO

#### Article history:

Received 30 October 2019

Received in revised from 29 March 2020

Accepted 6 April 2020

doi: 10.29244/j.agromet.34.1.20-33

#### Keywords:

BMA

ECMWF System 4

Probabilistic forecast

Reliability

Skill

### ABSTRACT

Bayesian Model Averaging (BMA) is a statistical post-processing method for producing probabilistic forecasts from an ensemble prediction in the form of predictive Probability Density Function (PDF). BMA is commonly used to calibrate Ensemble Prediction System (EPS) in a shorter-range forecast. Here, we applied the BMA for a longer forecast at a seasonal interval. This study aimed to develop the implementation of the BMA method to calibrate the seasonal forecast (long range) of monthly rainfall from the RAW output of the EPS European Center for Medium-Range Weather Forecasts (ECMWF) system 4 model (ECS4). This model was calibrated with observational data from 26 stations over Java Island in 1981-2018. BMA predictive PDF was generated with a gamma distribution, which was obtained based on two training schemes, namely sequential (BMA-JTS) and conditional (BMA-JTC) training windows. Generally, both of BMA-JTS and BMA-JTC were able to produce better distribution characteristics of ensemble prediction than that of RAW model ECS4. Both BMA methods showed a good performance as indicated by a high accuracy, small bias, and small uncertainty to the observed rainfall. Our findings revealed that BMA-JTC was able to improve the quality of probabilistic forecasts of below and above normal events. The improvement was shown in most stations over Java Island, in which the model was a good skill forecast based on Brier Skill Score (BSS).

## PENDAHULUAN

Prediksi musim merupakan bagian dari layanan informasi iklim yang dikeluarkan oleh berbagai badan meteorologi di dunia. Saat ini, prediksi musim yang mutakhir berupa prediksi berpeluang atau disebut probabilistik (Eide et al., 2017; Fang and Li, 2015; Han et al., 2016; Olson et al., 2016; Seong et al., 2017; Sillmann et al., 2017). Sebagai contoh prediksi probabilistik curah hujan bulanan yang memberikan informasi kepada pengguna bahwa pada beberapa bulan ke depan akan terjadi sejumlah curah hujan dengan peluang tertentu di suatu wilayah. Bagi pengguna, prediksi probabilistik memberikan informasi tingkat resiko sehingga mereka dapat mengambil keputusan yang tepat terkait pemahaman mereka

terhadap ketidakpastian dalam prediksi tersebut (Alessandrini et al., 2015). Namun demikian, kemanfaatan prediksi probabilistik bergantung pada kualitas dari model yang digunakan. Kualitas tersebut dinilai menggunakan ukuran kuantitatif, seperti: bias, asosiasi, akurasi, kehandalan, reliabilitas, resolusi, ketajaman, diskriminasi dan ketidakpastian (Vogel et al., 2018). Di Indonesia, Badan Meteorologi Klimatologi dan Geofisika (BMKG) telah memberikan layanan prediksi probabilistik curah hujan bulanan berdasarkan keluaran model ECMWF System 4. Namun demikian, kualitas prediksi tersebut belum dikaji menggunakan ukuran-ukuran seperti yang telah disebutkan.

Dalam model prediksi dinamis cuaca dan iklim, ketidakpastian akan selalu ada karena sifat alam yang *chaos* (kacau) dan saling berhubungan di dalam sistem

\* Corresponding author: [robi.muhsyah@gmail.com](mailto:robi.muhsyah@gmail.com)

iklim yang tidak linier (Wang *et al.*, 2017). Salah satu sumber ketidakpastian tersebut adalah keterbatasan informasi mengenai kondisi awal sistem iklim (Monier *et al.*, 2015). Namun hal ini dapat diatasi dengan membuat beberapa prediksi tunggal dengan nilai kondisi awal yang berbeda, atau dikenal sebagai prediksi *ensemble* (Baran and Möller, 2015; Bauer *et al.*, 2015; Schefzik, 2016). Prediksi *ensemble* dapat memberikan prediksi yang baik dan buruk (Bauer *et al.*, 2015). Prediksi *ensemble* yang buruk mengandung bias dan kesalahan dispersi karena belum terkoreksi atau terkalibrasi (Baran and Möller, 2017; Raftery *et al.*, 2005; Wang *et al.*, 2018). Bias disebabkan oleh sifat umum dari keluaran model prediksi dinamis cuaca dan iklim akibatnya nilai prediksi berbeda dari observasi (L'Heureux *et al.*, 2016). Di sisi lain, kesalahan dispersi disebabkan oleh prediksi *ensemble* yang *over confidence*, ditandai oleh sebaran anggota *ensemble* yang berkumpul pada satu nilai tertentu dan jauh dari observasi. Kesalahan dispersi seperti ini disebut juga *underdispersive* yang menyebabkan prediksi probabilistik menjadi tidak *reliable* dan kehandalannya buruk (Berner *et al.*, 2015; Jankov *et al.*, 2017; Kuswanto *et al.*, 2019; Siuta *et al.*, 2017; Späth *et al.*, 2015; Žagar *et al.*, 2015).

Masalah bias dan kesalahan dispersi dapat diatasi dengan *post processing* statistik yang bertujuan untuk memperbaiki kualitas prediksi (Guan *et al.*, 2015; Mendoza *et al.*, 2015). *Post Processing* dibedakan menjadi dua: koreksi bias (Khajehi *et al.*, 2018; Maraun *et al.*, 2017) dan kalibrasi *ensemble* (Gneiting, 2014; Schepen *et al.*, 2016; Taillardat *et al.*, 2016; Yang *et al.*, 2017; Zhu and Luo, 2015). Koreksi bias cukup efektif digunakan dalam mengurangi bias jika terdapat korelasi signifikan antara *raw* model dengan observasi (Bazile *et al.*, 2017; Manzanos *et al.*, 2018; Zhao *et al.*, 2017), sedangkan kesalahan dispersi dan masalah reliabilitas pada prediksi *ensemble* dapat diatasi oleh metode kalibrasi *Bayesian Model Averaging* (BMA) (Baran and Möller, 2015; Kontis *et al.*, 2017; Raftery *et al.*, 2005).

BMA awalnya dimanfaatkan untuk prediksi jangka pendek. Raftery *et al.* (2005) dan Sloughter *et al.* (2007) menggunakan BMA untuk mengkalibrasi prediksi suhu permukaan (T2m) dan curah hujan 48-jam kedepan dari keluaran multi model prediksi skala meso University of Washington MM5 di Pacific Northwest Amerika Serikat. Keduanya menunjukkan bahwa BMA mampu mengatasi permasalahan *under dispersive* pada kasus *riil* dan simulasi serta dapat menghasilkan prediksi probabilistik yang lebih reliabel dari *raw* model. Sebagai contoh, bentuk kurva *Reliability Diagram* untuk prediksi probabilistik kejadian hujan atau tidak hujan

(*probability of precipitation*) hasil BMA berada pada kategori sempurna (Sloughter *et al.*, 2007).

BMA juga digunakan untuk prediksi jangka pendek dan menengah dari Sistem Prediksi *Ensemble* (SPE), seperti kajian dari Wilson *et al.* (2007) menggunakan BMA untuk mengkalibrasi prediksi T2m untuk 1-10 hari ke depan dari 18 anggota SPE Meteorology Service of Canada. Kemudian Han *et al.* (2018) mengkalibrasi kecepatan angin 3-10 hari kedepan dari 51 anggota SPE ECMWF. Selanjutnya, Kolachian and Saghafian (2019) mengkalibrasi prediksi curah hujan dari 11 anggota SPE sub-seasonal to seasonal (S2S) ECMWF. Ketiga kajian tersebut menyimpulkan bahwa kemampuan BMA dipengaruhi oleh jenis distribusi variabel cuaca yang dikalibrasi dan penggunaan metode koreksi bias. Kemudian pendapat dari Gneiting (2014) dan Wilson *et al.* (2007) menyarankan penggunaan prediksi masa lalu (*reforecast*) dapat mempengaruhi hasil kalibrasi BMA. Hal ini disebabkan karena BMA dapat diterapkan dengan dua skema jendela *Training Conditional* (JTC) dan jendela *Training Sequential* (JTS) (Erickson *et al.*, 2012).

Berdasarkan uraian di atas, metode BMA baru digunakan untuk mengkalibrasi prediksi *ensemble* jangka pendek hingga menengah. Oleh karena itu penelitian ini difokuskan pada penerapan BMA untuk mengkalibrasi prediksi musim (jangka panjang) berupa prediksi curah hujan bulanan menggunakan keluaran model SPE ECMWF System 4 (selanjutnya: model ECS4). Penelitian ini bertujuan untuk: (i) mengembangkan penerapan metode BMA dengan memanfaatkan ketersediaan *reforecast* model ECS4 sehingga dapat dilakukan konfigurasi jenis jendela training BMA: JTS dan JTC; (ii) memperbaiki bias, kesalahan dispersi dan meningkatkan reliabilitas, resolusi serta kehandalan prediksi probabilistik curah hujan bulanan dari model ECS4 dengan metode BMA, sehingga diperoleh kualitas prediksi musim yang lebih baik untuk wilayah Indonesia, secara khusus di Pulau Jawa.

## BAHAN DAN METODE

### Data Observasi dan Model

Data yang digunakan berupa jumlah curah hujan dalam satu bulan. Data observasi diambil dari pengamatan pada 26 stasiun BMKG di P. Jawa (Tabel 1). Semua stasiun mempunyai data kosong kurang dari 20% dari periode 1981-2018 (456 bulan). Data kosong diisi dengan nilai klimatologi bulan yang bersesuaian. Selanjutnya, data model yang dikalibrasi adalah curah hujan keluaran model ECS4.

**Tabel 1** Karakteristik 26 stasiun BMKG di Pulau Jawa yang disusun berdasarkan jumlah persentase nilai curah hujan nol berdasarkan data curah hujan bulanan (Jan 1981 – Des 2018)

No	No STA	Name STA	Code STA	LON	LAT	Ketinggian (m)	Kosong (%)	CH Nol (%)	Rerata Tahunan
1	96935	Juanda	JND	112.77	-7.37	3	0.2	18.2	2143
2	96973	Kalianget	KLG	113.97	-7.05	0	0.2	16.2	1415
3	96937	Perak II	PRK2	112.74	-7.21	3	0	15.6	1562
4	96975	Nganjuk	NGJ	111.79	-7.47	723	19.7	14.3	3011
5	96933	Perak I	PRK1	112.46	-7.13	3	0	14	1659
6	96925	Bawean	BWN	112.63	-5.85	3	0	11.2	2339
7	96845	Yogyakarta	YGY	110.30	-7.82	153	3.7	8.3	2436
8	96949	Karangates	KRS	112.45	-8.15	285	18.4	7.9	2174
9	96791	Jatiwangi	JTW	108.27	-6.75	85	2.9	7.5	2690
10	96943	Karangploso	KRO	112.60	-7.90	590	19.3	6.6	1856
11	96797	Tegal	TGL	109.15	-6.85	1	5.3	6.4	1788
12	96945	Tretes	TRS	112.64	-7.70	832	5.9	6.4	3783
13	96741	Tj Priuk	TJP	106.88	-6.11	3	1.3	4.4	1761
14	96747	Halim	HLM	106.90	-6.25	4	0.7	4.2	2236
15	96835	Semarang Klim	SMK	110.42	-6.98	6	0	3.7	2289
16	96839	AYani Semarang	ASM	110.38	-6.89	5	2.6	3.7	2322
17	96745	Kemayoran	KMY	106.84	-6.16	4	0	3.5	1888
18	96737	Serang	SRG	106.11	-6.11	100	0	3.3	1618
19	96749	Cengkareng	CGK	106.65	-6.12	11	11.6	3.3	1703
20	96987	Banyuwangi	BWI	114.35	-8.21	52	1.8	2.9	1395
21	96805	Cilacap	CLP	109.02	-7.73	8	0	2.6	3481
22	96733	Pd. Betung	PDB	106.75	-6.26	27	0.4	2.4	2301
23	96739	C. Tangerang	CRG	106.57	-6.29	42	2	2.4	2261
24	96783	G. Bandung	GBD	107.60	-6.88	791	0.2	1.3	2192
25	96751	Citeko	CTO	106.85	-6.70	920	10.5	0.2	3114
26	96753	Darmaga	DMG	106.75	-6.50	207	7.9	0	3822

Resolusi spasial model ini sekitar  $0,7^0 \times 0,7^0$  dan mempunyai 15 anggota *ensemble* untuk periode 1981 – 2018, terdiri dari produk *reforecast* selama 34 tahun (1981-2014) dan produk *forecast* selama 4 tahun (2015-2018). Inisial prediksi curah hujan model ECS4 dibuat pada tanggal 1 setiap bulan untuk prediksi 215 hari atau sekitar tujuh bulanan (Wang et al., 2019). Misal inisial prediksi 1 Jan 2018, terdiri dari prediksi untuk bulan Jan hingga Juli dengan Jan disebut prediksi lead time nol (LT0), Feb LT1, Mar LT2 dan seterusnya. Terdapat tujuh LT untuk prediksi curah hujan bulanan, namun hanya LT0 yang dipilih untuk dikalibrasi. Data model pada tiap stasiun diperoleh dengan metode *nearest neighbour*, yaitu dengan mengambil nilai grid model ECS4 terdekat dengan koordinat stasiun. Data model ECS4 sebelum dikoreksi ini disebut sebagai RAW model.

### Koreksi Bias dengan Quantile Mapping

RAW model mempunyai bias dan kesalahan dispersi (Han et al., 2016). Oleh karena itu, pada tahap awal dilakukan identifikasi bias dan kesalahan dispersi pada RAW model ECS4. Bias diidentifikasi dengan

menghitung selisih antara rerata anggota *ensemble* (*ensemble mean*) RAW model terhadap observasi. Bias tersebut dikoreksi dengan metode yang merujuk pada kajian Piani et al. (2010). Metode ini termasuk pada koreksi bias menggunakan distribusi parametrik, disebut juga sebagai *Gamma Quantile Mapping* (gQM) (Pastén-Zapata et al., 2020).

1. Metode gQM mengasumsikan curah hujan dapat didekati secara teoritis dengan distribusi Gamma. PDF distribusi Gamma pada Persamaan (1) digunakan untuk mendapatkan parameter bentuk dan skala di periode training.
2. Selanjutnya, hubungan fungsi transfer *inverse Gamma* CDFs antara curah hujan bulanan observasi dengan tiap *ensemble* RAW model dibuat dengan Persamaan (2). Dari hubungan tersebut diperoleh Faktor Koreksi (FK) dengan persamaan regresi linier atau *polynomial*. FK diperoleh dengan skema validasi silang *Leave One Out Cross Validation* (LOOCV) per tahun. Prosedur yang digunakan yaitu dengan mengeluarkan curah hujan bulanan untuk satu tahun sebagai periode *testing* lalu

menggunakan tahun lainnya sebagai periode *training* untuk mendapatkan FK.

3. FK kemudian diterapkan pada tiap anggota *ensemble* (Pelosi *et al.*, 2017; Piani *et al.*, 2010).
4. Selanjutnya, prediksi *ensemble* RAW model yang terkoreksi (*Bias Correction Quantile Mapping*, BCQM), dievaluasi kembali nilai bias dan kesalahan dispersinya dan membandingkannya dengan RAW model.

$$pdf(y) = \frac{e^{(-\frac{y}{\beta})} y^{(\alpha-1)}}{\Gamma(\alpha)\beta^\alpha} \quad (1)$$

$$cdf(y) = \int_0^y \frac{e^{(-\frac{y}{\beta})} y^{(\alpha-1)}}{\Gamma(\alpha)\beta^\alpha} + cdf(0) \quad (2)$$

Keterangan:  $y$  = curah hujan bulanan,  $\alpha$  = parameter bentuk, dan  $\beta$  = parameter skala, serta  $cdf(0)$  = fraksi curah hujan sama dengan nol.

### Kalibrasi Prediksi *Ensemble* dengan *Bayesian Model Averaging*

Selain melakukan koreksi bias dengan gQM, tahapan utama pada penelitian ini adalah menerapkan BMA pada RAW model ECS4. BMA adalah metode *post processing* statistik yang menghasilkan prediksi probabilistik dari prediksi *ensemble* dalam bentuk fungsi kepadatan peluang atau *Probability Density Function* (PDF) prediktif (Abraham and Puthiyadam, 2016; Baran *et al.*, 2019; Ji *et al.*, 2019; Liu *et al.*, 2019; Song *et al.*, 2018; Xu *et al.*, 2019). Misalkan  $y$  adalah observasi dan  $f = f_1, f_2, \dots, f_k$  adalah prediksi *ensemble* RAW model ECS4 dengan  $k$  anggota maka setiap anggota *ensemble*  $f_k, k = 1, 2, \dots, K$  mempunyai PDF posterior  $g_k(y|f_k)$ . Nilai PDF posterior menunjukkan besar peluang  $f_k$  sebagai syarat untuk terjadinya  $y$  dengan  $f_k$  dikatakan sebagai prediksi *ensemble* terbaik. Jenis distribusi PDF posterior bergantung pada variabel cuaca yang dikalibrasi, seperti: T2m (suhu permukaan) dan *sea level pressure* (SLP) menggunakan distribusi Normal (Raftery *et al.*, 2005), curah hujan harian dengan gamma nol (Sloughter *et al.*, 2007), dan kecepatan angin dengan gamma (Sloughter *et al.*, 2010).

Menurut Gneiting (2014) dan Sloughter *et al.* (2007), distribusi PDF prediktif BMA untuk curah hujan harian adalah gamma nol. Pada gamma nol, PDF BMA dibentuk dari distribusi gamma (untuk kejadian curah hujan besar dari nol) dan distribusi peluang kejadian curah hujan sama dengan nol. Hal ini disebabkan, pada skala harian kejadian hujan sama dengan nol selalu muncul atau peluangnya selalu ada meskipun pada periode musim hujan. Namun untuk skala bulanan (jumlah curah hujan dalam satu bulan), peluang curah hujan sama dengan nol sangat sedikit atau bahkan tidak ada, seperti pada lokasi stasiun yang berada di topografi tinggi. Oleh karena itu, penelitian ini menggunakan distribusi gamma untuk mendapatkan PDF prediktif BMA. Distribusi gamma hanya berlaku

untuk curah hujan lebih besar dari nol. Sehingga untuk mengatasi masalah jika terdapat curah hujan bulanan dengan nilai nol, maka sebelum kalibrasi BMA diterapkan, terlebih dahulu dimodifikasi curah hujan sama dengan nol mm menjadi 0,1 mm. Hal ini diasumsikan bahwa kedua nilai tersebut tidak mempunyai perbedaan yang berarti dalam konteks jumlah curah hujan satu bulan. Model PDF prediktif BMA berdistribusi gamma dinyatakan pada Persamaan (3).

$$p(y|f_1, f_2, \dots, f_k) = \sum_{k=1}^K w_k g_k(y|f_k) \quad (3)$$

dengan  $w_k$  adalah bobot yang diperoleh berdasarkan kemampuan relatif dari anggota *ensemble* ke- $k$  pada periode *training*. Nilai  $w_k$  adalah *non negative* dan berjumlah satu. PDF posterior  $g_k(y|f_k)$  berdistribusi gamma dengan parameter skala ( $\alpha_k$ ) dan parameter bentuk ( $\beta_k$ ) dinyatakan dalam Persamaan (4).

$$g_k(y|f_k) = \frac{e^{(-\frac{y}{\beta_k})} y^{(\alpha_k-1)}}{\Gamma(\alpha_k)\beta_k^{\alpha_k}} \quad (4)$$

Parameter bentuk dan skala pada Persamaan (4),  $\alpha_k = \frac{\mu_k^2}{\sigma_k^2}$  dan  $\beta_k = \frac{\sigma_k^2}{\mu_k}$  bergantung pada anggota *ensemble*  $f_k$  melalui hubungan dari rerata (mean)  $\mu_k = b_{0,k} + b_{1,k}f_k$  dan variasinya (variance)  $\sigma_k^2 = c_0 + c_1f_k$ , untuk  $K = 1, 2, 3, \dots, 15$ . Pada periode *training*, pendugaan parameter  $b_{0,k}, b_{1,k}$  dilakukan dengan regresi linier  $b_{0,k} + b_{1,k}f_k$ , sedangkan parameter  $w_k$ , dan  $c_0, c_1$  diduga dengan metode *Expectation Maximization* (Sloughter *et al.*, 2010). Penjelasan lebih rinci tentang pendugaan parameter BMA dapat ditemukan di literatur (e.g. Fang and Li, 2015; He *et al.*, 2018; Zhong *et al.*, 2018). Pada penelitian ini *post processing* BMA dilakukan dengan fungsi *ensemble* BMA di aplikasi paket R (Fralely *et al.*, 2007).

Periode *training* BMA pada penelitian ini dibuat dengan dua skema jendela *training*: BMA-JTS dan BMA-JTC. BMA-JTS merupakan skema yang digunakan pada saat pertama kali BMA diperkenalkan untuk kalibrasi prediksi *ensemble* jangka pendek (Feldmann *et al.*, 2015; Raftery *et al.*, 2005; Sloughter *et al.*, 2007). Pada BMA-JTS, jendela *training* diperbaharui dengan prediksi *ensemble* dan observasi curah hujan bulanan terbaru untuk mengkalibrasi prediksi pada bulan berikutnya, sedangkan BMA-JTC menggunakan bulan yang sama dari tahun sebelumnya untuk memperbaharui jendela *training*. Hasil kalibrasi BMA-JTS dan BMA-JTC adalah berupa kurva PDF, oleh karena itu set prediksi *ensemble* keduanya diperoleh dari hasil *sampling* secara acak terhadap PDF tersebut sebanyak 15 anggota.

### Ukuran Evaluasi

Evaluasi difokuskan pada tiga set prediksi *ensemble*: RAW model, hasil kalibrasi BMA-JTS, dan

BMA-JTC. Tahapan evaluasi terbagi menjadi dua: Pertama, evaluasi karakteristik distribusi prediksi *ensemble*. Karakteristik tersebut diketahui dari bias, akurasi dan kesalahan dispersi. Akurasi diukur berdasarkan hubungan korelasi, standar deviasi dan RMSE dari siklus tahunan curah hujan antara ketiga set prediksi *ensemble* tersebut terhadap observasi. Set prediksi paling akurat ditunjukkan oleh posisinya yang paling mendekati titik referensi pada *Taylor Diagram* (Taylor, 2001). Selanjutnya, akurasi dari distribusi set prediksi *ensemble* diukur menggunakan *Continuous Rank Probability Score* (CRPS). CRPS digunakan untuk mengidentifikasi seberapa akurat distribusi seluruh anggota *ensemble* terhadap observasi. Prediksi *ensemble* terkalibrasi jika nilai CRPS BMA lebih kecil dari CRPS RAW model (Hudson and Ebert, 2017). Selain akurasi, kesalahan dispersi juga diukur dengan analisis bentuk VRH. Bentuk tersebut digunakan untuk mengidentifikasi kemampuan sebaran anggota *ensemble* dalam mewakili ketidakpastian nilai observasi. Ciri prediksi *ensemble* terkalibrasi jika bentuk VRH mendekati *uniform* atau bar histogram lebih datar, artinya nilai observasi selalu berada diantara sebaran anggota *ensemble* (Taillardat et al., 2016).

Kedua, evaluasi prediksi *ensemble* yang dinyatakan dalam bentuk prediksi probabilistik kejadian Bawah Normal (BN) dan Atas Normal (AN). Persentase probabilistik kejadian BN (AN) dihitung dari luas area PDF yang kecil dari persentile ke-33 (besar dari persentile-66). Kualitas prediksi probabilistik dinilai berdasarkan atribut reliabilitas, resolusi dan kehandalan (*skill*) yang diukur menggunakan: *Reliability Diagram* (RD), *Brier Score* (BS) dan *Brier Skill Score* (BSS). Reliabilitas diukur berdasarkan bentuk kurva RD yang menunjukkan hubungan jumlah frekwensi kejadian observasi sebagai fungsi dari tiap probabilistiknya. Pada penelitian ini lima kategori kurva RD digunakan berdasarkan Weisheimer and Palmer (2014) seperti ditunjukkan oleh Gambar 1a. Terdapat lima kategori kurva RD berdasarkan kemiringannya terhadap diagonal 1:1 (*reliability perfect*). Lima kategori tersebut menunjukkan seberapa berguna prediksi probabilistik dapat dimanfaatkan dalam pengambilan keputusan terkait prediksi musim, kategori-1: *dangerous* (berbahaya), -2: *not useful* (tidak berguna), -3: *marginally useful* (cukup berguna), -4: *still very useful* (sangat berguna) dan -5 *perfect* (sempurna). Kategori 4 dan 5, artinya prediksi probabilistik kejadian BN dan AN, berguna bahkan sangat dianjurkan digunakan dalam pengambilan keputusan terkait prediksi musim di suatu wilayah. Kategori 3, berarti prediksi yang dibuat masih berkaitan dengan kejadian yang diamati. Kategori 2, berarti prediksi yang dibuat cenderung keliru dibandingkan dengan kejadian sebenarnya sehingga

dianggap tidak berguna untuk pengambilan keputusan. Kategori 1 menunjukkan kondisi yang lebih buruk dan prediksi probabilistik kejadian BN dan AN yang dihasilkan dapat menimbulkan kesalahan informasi bahkan berbahaya jika dipakai dalam pengambilan keputusan terkait prediksi musim (Weisheimer and Palmer, 2014).

Selanjutnya BS digunakan untuk mengukur seberapa besar tingkat kesalahan prediksi probabilistik kejadian AN atau BN (Peeters, 2018), dan BSS digunakan untuk mengukur kehandalan relatif BS tersebut terhadap klimatologisnya. Nilai BS semakin baik jika mendekati nol dan sebaliknya BSS semakin baik jika mendekati 1. BSS sama dengan nol berarti kehandalan prediksi probabilistik yang dibuat sama dengan klimatologisnya sedangkan BSS negatif artinya prediksi tersebut mempunyai kehandalan yang buruk. Perhitungan semua skor pada tahapan evaluasi ini dilakukan dengan fungsi *SpecsVerification* didalam aplikasi paket R (Siebert et al., 2014).

## HASIL DAN PEMBAHASAN

### Klimatologi Curah Hujan Bulanan di Pulau Jawa

Pola siklus tahunan curah hujan di P. Jawa termasuk pada Tipe A menurut pengelompokan tiga pola dominan curah hujan di Indonesia (Supari et al., 2018). Ciri Tipe A ini terdapat satu puncak curah hujan tertinggi di sekitar bulan Des-Feb (DJF), sedangkan curah hujan terendah terjadi pada bulan Jun-Agu (JJA). Pada Gambar 1b, meskipun mempunyai pola yang sama (tipe A), terdapat perbedaan rerata curah hujan bulanan antar stasiun, yaitu >300 mm/bulan pada DJF dan <100 mm/bulan pada JJA. Tinggi boxplot pada Gambar 1b juga menggambarkan variasi curah hujan bulanan antar stasiun. Hal tersebut seperti terlihat pada periode Nov-Apr (NDJFMA), dimana terdapat variasi curah hujan 100-500 mm/bulan antar stasiun di P. Jawa. Variasi ini juga diketahui dari rerata total curah hujan tahunan masing-masing stasiun sebesar 1000-4000 mm/tahun (Tabel 1). Sementara itu, berdasarkan kategori musim hujan dan musim kemarau yang digunakan BMKG dengan batasan 150 mm/bulan, maka periode NDJFMA termasuk pada kategori periode musim hujan sedangkan Mei-Oktober (MJJASO) termasuk pada periode musim kemarau.

### Pengaruh Koreksi Bias terhadap Prediksi *Ensemble* RAW Model ECS4

#### *Kondisi Bias pada RAW Model ECS4*

Hamburan titik merah yang tersebar di kiri dan kanan diagonal *scatter plot* pada Gambar 2a, menunjukkan bahwa RAW model mempunyai bias yang

bervariasi, yaitu: bias negatif atau lebih rendah (*under estimate*) dan bias positif atau lebih tinggi dari observasi (*over estimate*). Namun demikian, untuk curah hujan tinggi > 400 mm/bulan, RAW model cenderung *under estimate* dibandingkan observasi yang mencapai 700 mm/bulan. Oleh karena itu, pada Gambar 2a, banyak titik merah berada pada sisi kanan diagonal yang menunjukkan RAW model lebih rendah dari observasi. Variasi bias ini juga terlihat secara temporal dari rerata curah hujan bulanan (Jan-Des) untuk semua stasiun BMKG di P. Jawa (Gambar 1b). Pada periode hujan (kemarau) RAW model (garis merah) cenderung *under estimate* dibandingkan observasi (garis hitam).

Selain bias, RAW model juga memuat kesalahan dispersi berdasarkan bentuk VRH (*bar* merah pada Gambar 2b). Bentuk VRH mempunyai *spike* pada kiri dan kanan (*U-shaped*) yang berarti bahwa pada sebagian stasiun nilai observasi selalu berada di luar (lebih besar atau lebih kecil) dari semua anggota *ensemble* RAW model. Dengan kata lain, bentuk VRH seperti itu menunjukkan adanya masalah terkait sebaran anggota *ensemble* yang terlalu sempit atau disebut *underdispersive*. Kondisi bias pada RAW model ini dapat diatasi dengan koreksi bias menggunakan metode *Quantile Mapping*.

**Koreksi Bias dengan Gamma Quantile Mapping**

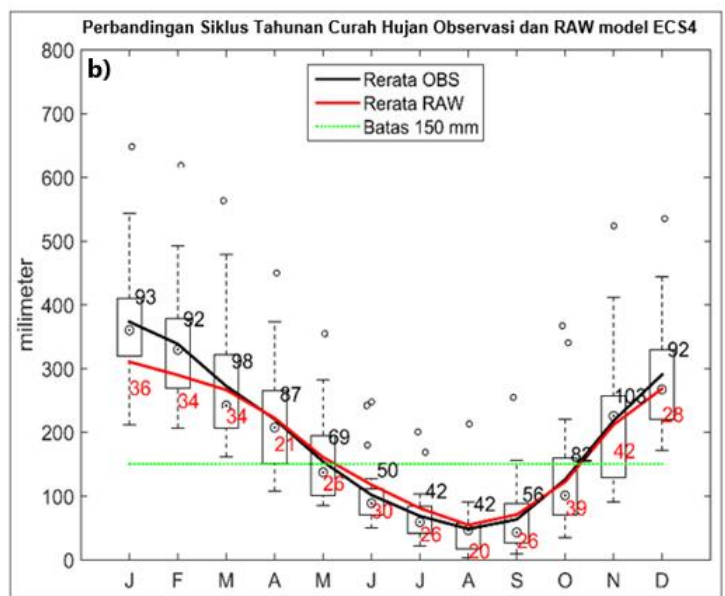
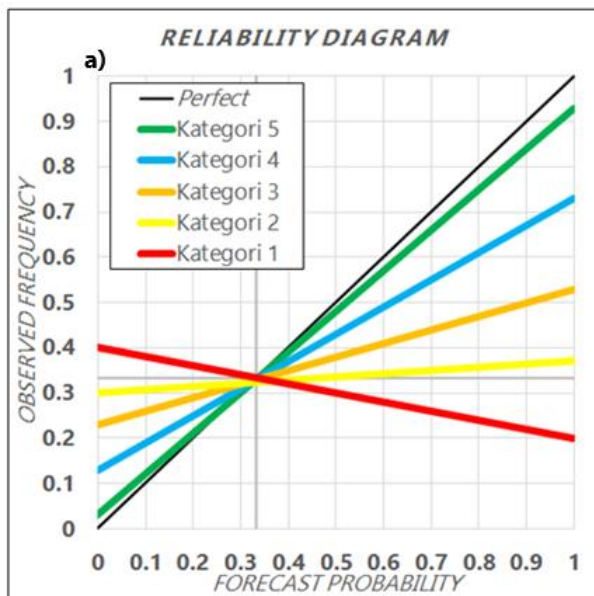
Bias pada BCQM lebih baik dibanding RAW model ECS4 karena ditandai oleh hamburan titik biru semakin mendekati diagonal *scatter plot* dan peningkatan nilai  $R^2$  dari 0,7069 menjadi 0,8913

(Gambar 2a). Dengan kata lain, metode gQM dapat mengurangi selisih antara RAW model terhadap observasi. Namun, meskipun dapat mengurangi bias, BCQM tidak memperbaiki sebaran anggota *ensemble* secara keseluruhan karena bentuk VRH pada Gambar 2b (*bar* biru) masih *U-shaped*. Hasil ini sepakat dengan pendapat Gneiting (2014), bahwa koreksi bias hanya memperbaiki kesalahan pada variasi yang ada di setiap individu anggota *ensemble*, tetapi tidak mampu mengalibrasi semua anggota *ensemble* secara keseluruhan. Oleh karena itu, BCQM masih mempunyai kesalahan dispersi, sehingga diperlukan metode *post processing* BMA untuk mengatasi masalah ini.

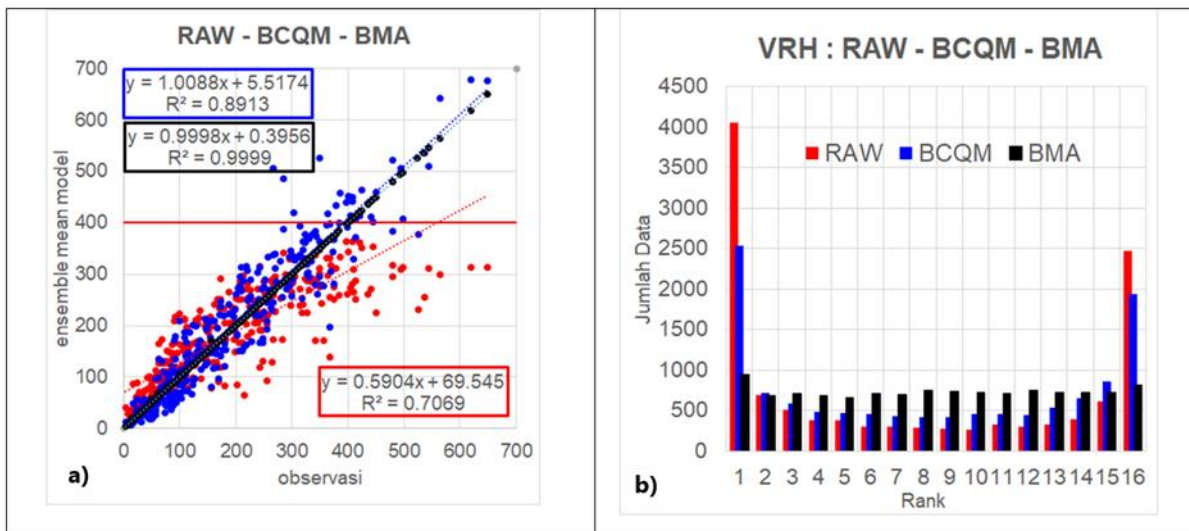
**Kemampuan BMA dalam Memperbaiki Kualitas Prediksi Ensemble Model ECS4**

**BMA dengan Jendela Training Sequential**

Prediksi *ensemble* RAW model ECS4 berasal dari model tunggal yang diberikan gangguan acak untuk menghasilkan beberapa prediksi individu atau disebut sebagai anggota *ensemble* (Molteni et al., 2011). Secara statistik setiap anggota *ensemble* tersebut tidak dapat dibedakan kemampuannya. Oleh karena itu pada penelitian ini, BMA-JTS dan BMA-JTC dilakukan dengan konsep *exchangeable*. Artinya, setiap anggota *ensemble* dapat saling menggantikan satu sama lain, atau tidak ada anggota yang dianggap lebih baik atau lebih buruk kemampuannya dari anggota lainnya. Oleh karena itu, bobot  $w_k$  pada Persamaan (3) bernilai sama 1/15 untuk tiap anggota (Schmeits and Kok, 2010).



**Gambar 1.** (a) Lima kategori reliabilitas berdasarkan bentuk kurva *Reliability Diagram* (dimodifikasi dari Weisheimer and Palmer, 2014 dan (b) Rerata siklus tahunan curah hujan dari semua stasiun BMKG di P. Jawa untuk observasi (hitam), RAW model ECS4 (merah) dengan angka pada garis adalah standar deviasi tiap bulan. sedangkan rerata curah hujan bulanan untuk observasi tiap stasiun ditampilkan dalam bentuk boxplot. Batas pembeda musim kemarau dan musim hujan ditunjukkan oleh garis horizontal hijau putus-putus pada nilai 150 mm/bulan.



**Gambar 2.** Bias pada RAW model ECS4 (merah), BCQM (biru), BMA-JTC (hitam) terhadap observasi yang ditampilkan dalam bentuk (a) *scatter plot* dan (b)VRH. Keduanya dihasilkan dari siklus tahunan curah hujan (1981-2018) untuk 26 stasiun di P. Jawa.

Penerapan BMA-JTS dilakukan dengan menguji sebanyak 12 panjang jendela *training* (JT) mulai dari JT 10 bulan hingga JT 120 bulan. Misal pada JT 10 bulan, proses BMA-JTS dimulai dari 10 bulan pertama pada data, yaitu dengan periode *training* Jan 1981-Okt 1981 dan *testing* Nov 1981. Sedangkan pada JT 20 bulan, proses BMA-JTS dimulai dari 20 bulan pertama, yaitu dengan periode *training* Jan 1981-Agt 1982 dan *testing* Sep 1982, dan hal yang sama seterusnya untuk JT 30, 40 hingga 120 bulan. Sehingga semakin panjang JT yang dipakai maka panjang periode *testing* yang diperoleh semakin berkurang, begitu sebaliknya.

Namun tidak semua JT digunakan dalam proses kalibrasi BMA-JTS. Penentuan panjang JT optimal dipilih berdasarkan nilai rerata CRPS. CRPS yang semakin kecil menunjukkan bahwa PDF prediktif BMA-JTS (selanjutnya: PDF BMA-JTS) mempunyai interval yang lebih sempit dibandingkan PDF prediktif RAW model (selanjutnya: PDF RAW model) sehingga diartikan bahwa nilai observasi lebih banyak berada di dalam interval PDF BMA-JTS dibandingkan RAW model.

Rerata dari semua stasiun menunjukkan bahwa pada JT 10 bulan nilai CRPS berkisar 67 mm, kemudian menurun pada JT 20 bulan menjadi 60 mm. Selanjutnya pada JT 40 bulan juga terdapat penurunan nilai CRPS menjadi sekitar 58 mm. Namun setelah JT 30 bulan nilai keduanya relatif datar dan tidak mengalami penurunan. Hasil ini sama dengan temuan (Sloughter *et al.*, 2007) dengan JT 30 hari pada penelitiannya yang mengkalibrasi prediksi curah hujan harian atau temuan (Sloughter *et al.*, 2010) yang menyebutkan JT 25-40 hari sebagai JT optimal pada saat mengkalibrasi prediksi kecepatan angin harian. Dengan demikian, kalibrasi BMA-JTS menunjukkan bahwa periode *training* yang diperlukan untuk mencapai hasil yang optimal cukup sekitar 30

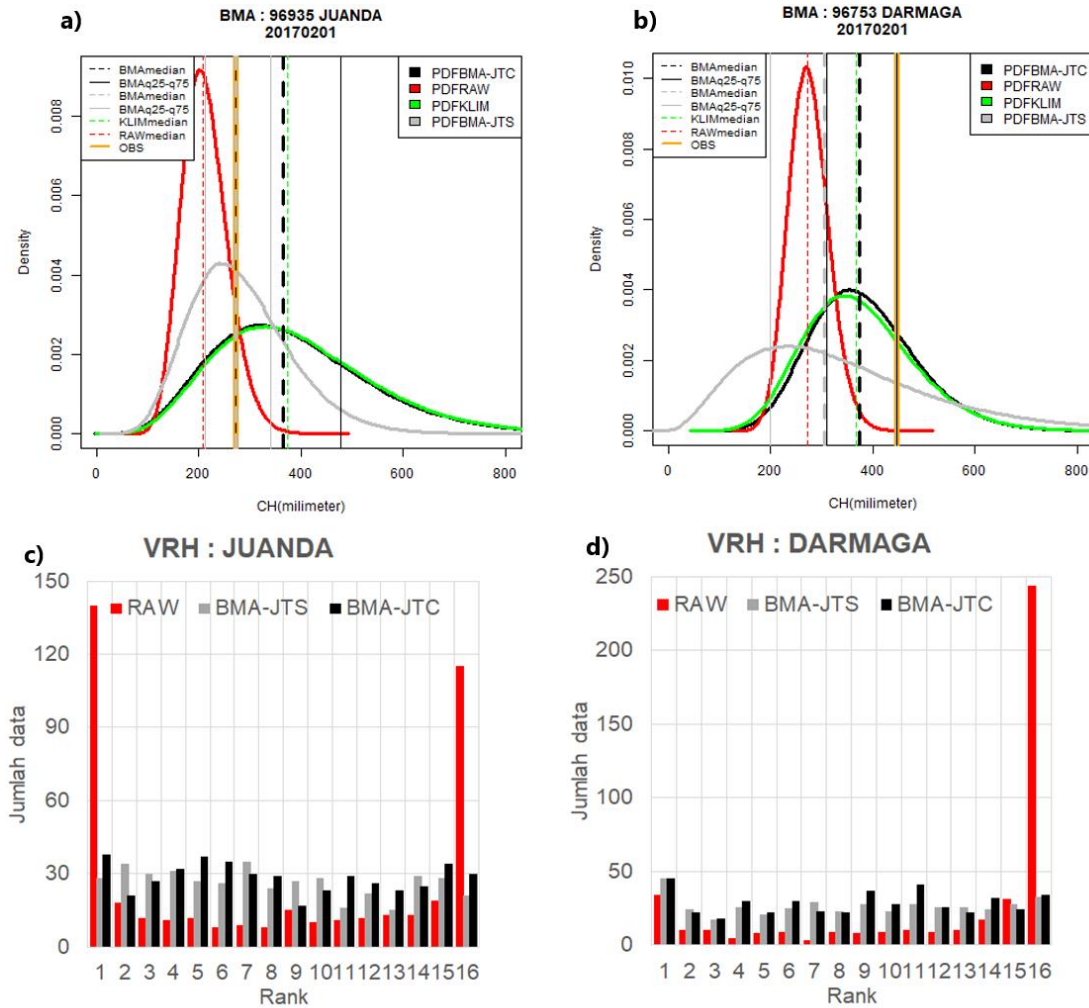
data (bisa hari atau bulan tergantung satuan data yang digunakan).

Berdasarkan deret curah hujan bulanan dari Jan 1981-Des 2018 maka dengan JT 30 bulan, diperoleh panjang periode *testing* dari Juli 1983-Des 2018 (426 bulan) dengan setiap proses *training* dilakukan pendugaan parameter  $b_{0,k}, b_{1,k}$  serta  $c_0, c_1$  sehingga diperoleh parameter skala ( $\alpha_k$ ) dan parameter bentuk ( $\beta_k$ ) untuk PDF BMA-JTS yang berbeda. Gambar 3a dan 3b adalah contoh hasil kalibrasi BMA-JTS dengan JT 30 bulan (periode *training* Agt 2014-Jan 2017), pada stasiun Juanda (JND) (kiri) dan Dramaga (DMG) (kanan) untuk bulan Februari 2017, dan ditampilkan bersamaan dengan PDF Klimatologi curah hujan bulan Februari 1981-2010 dan PDF RAW model.

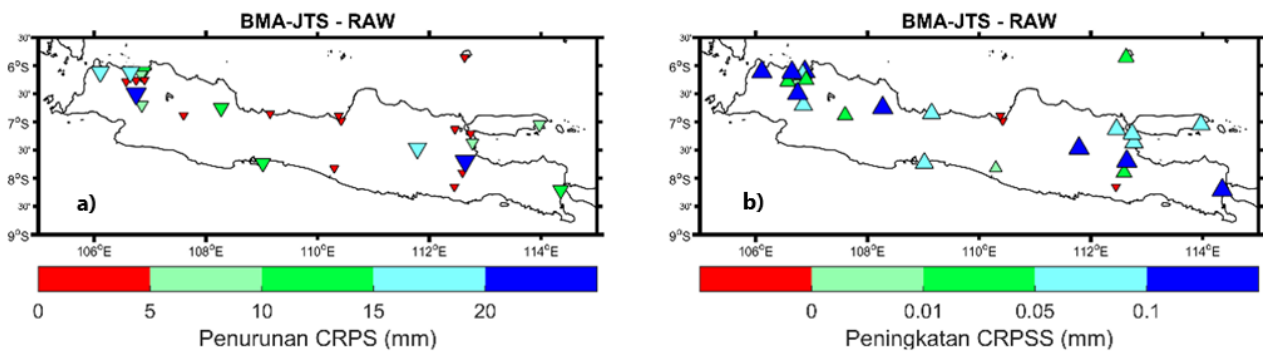
Pada stasiun JND, nilai observasi (garis vertikal orange) berada dalam 50% interval PDF BMA-JTS (quantile-25 hingga quantile-75). Pada Gambar 3, jika dibandingkan dengan median PDF RAW model (garis vertikal merah putus-putus), maka median dari PDF BMA-JTS (garis vertikal hitam putus-putus) berhimpit dengan nilai observasi untuk Februari 2017 di stasiun JND. Hal tersebut indikasi BMA-JTS lebih akurat dari RAW model, dengan selisih antara median PDF BMA-JTS terhadap observasi mendekati nol mm. Begitu juga untuk stasiun DMG, BMA-JTS lebih unggul dari RAW model karena median PDF BMA-JTS lebih dekat dengan observasi, dengan selisih sekitar 50 mm dibandingkan terhadap median PDF RAW model dengan selisih sekitar 100 mm. Meskipun PDF BMA-JTS untuk setiap bulan *testing* dan setiap lokasi mempunyai hasil beragam, bergantung pada ruang (lokasi stasiun) dan waktu (bulan yang dikalibrasi), namun secara umum kalibrasi BMA-JTS mampu memperbaiki karakteristik distribusi prediksi *ensemble* RAW model.

Berdasarkan bentuk VRH (Gambar 3c dan 3d), proporsi *bar* VRH BMA-JTS mendekati *uniform* atau lebih datar. BMA-JTS juga mempunyai nilai CRPS yang lebih kecil dari RAW model karena terdapat penurunan nilai CRPS BMA-JTS terhadap RAW model (Gambar 4a). Hal ini berarti PDF BMA-JTS lebih akurat dibanding PDF RAW model. Selanjutnya, BMA-JTS juga meningkatkan

*skill* CRPS (CRPSS) yang menunjukkan bahwa selain lebih baik dari RAW model, hasil kalibrasi BMA-JTS juga lebih baik dari nilai klimatologi. Penurunan (peningkatan) CRPS (CRPSS) BMA terhadap RAW model ini terjadi pada semua stasiun BMKG di P. Jawa (Gambar 4a dan 4b).



**Gambar 3.** Panel a dan b: PDF BMA-JTS (abu-abu), BMA-JTC (hitam), PDF RAW model (merah), PDF klimatologi (hijau), observasi (garis vertikal orange) untuk Februari 2017; Panel c dan d: VRH BMA-JTS (*bar* abu-abu), VRH BMA-JTC (hitam) dan VRH RAW model (merah). Kolom kiri (kanan) untuk stasiun Juanda (Darmaga).



**Gambar 4.** Penurunan nilai *Continuous Rank Probability Source* (CRPS) (a) dan peningkatan *skill* CRPS (b) hasil BMA-JTS dengan JT 30 bulan terhadap RAW model ECS



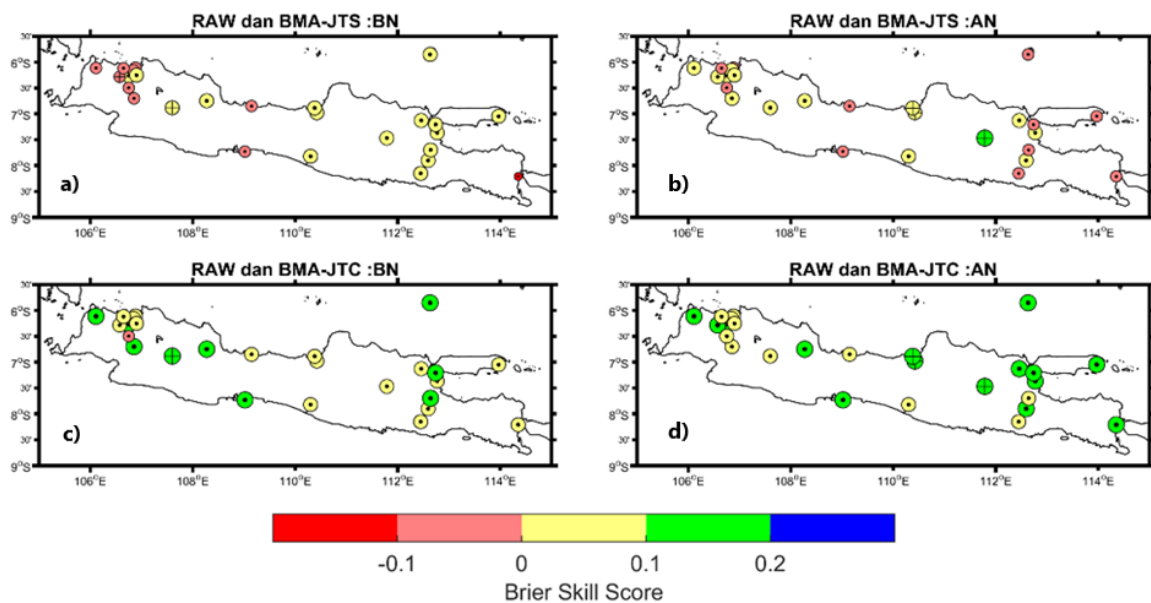
Karakteristik distribusi prediksi *ensemble* yang semakin baik juga diikuti dengan perbaikan evaluasi prediksi probabilistik kejadian AN dan BN yang dihasilkan dari *set* prediksi *ensemble* kalibrasi BMA-JTS dibandingkan RAW model. Pertama, berdasarkan nilai BSS (Gambar 5a dan 5b), diketahui nilai BSS RAW model umumnya bernilai negatif sebanyak 24(22) stasiun pada BN(AN), setelah dilakukan kalibrasi BMA-JTS terdapat peningkatan nilai BSS menjadi positif dan tersisa 11(10) stasiun pada BN(AN) yang bernilai negatif. Peningkatan BSS secara umum lebih banyak terjadi pada stasiun-stasiun di wilayah Jawa bagian timur dan tengah dibandingkan Jawa bagian barat.

Kedua, berdasarkan bentuk *Reliability Diagram* (RD) yang dibagi menjadi 5 kategori, menunjukkan RD kejadian BN dan AN pada RAW model sebagian besar berkategori 3, 2 dan 1 atau *marginally useful*, *not useful* dan *dangerous* (Gambar 7a dan 7b). Kemudian, hasil kalibrasi BMA-JTS dapat meningkatkan kategori RD menjadi 4 dan 5 (Gambar 7c dan 7d). Peningkatan reliabilitas ini menunjukkan BMA-JTS mampu memperbaiki reliabilitas RAW model sehingga informasi prediksi probabilistik kejadian BN dan AN menjadi berguna dan sangat dianjurkan untuk dipakai dalam pengambilan keputusan yang berkaitan dengan prediksi musim di P. Jawa, seperti menentukan apakah prediksi curah hujan di bulan berikutnya akan lebih basah atau lebih kering dari kondisi normal.

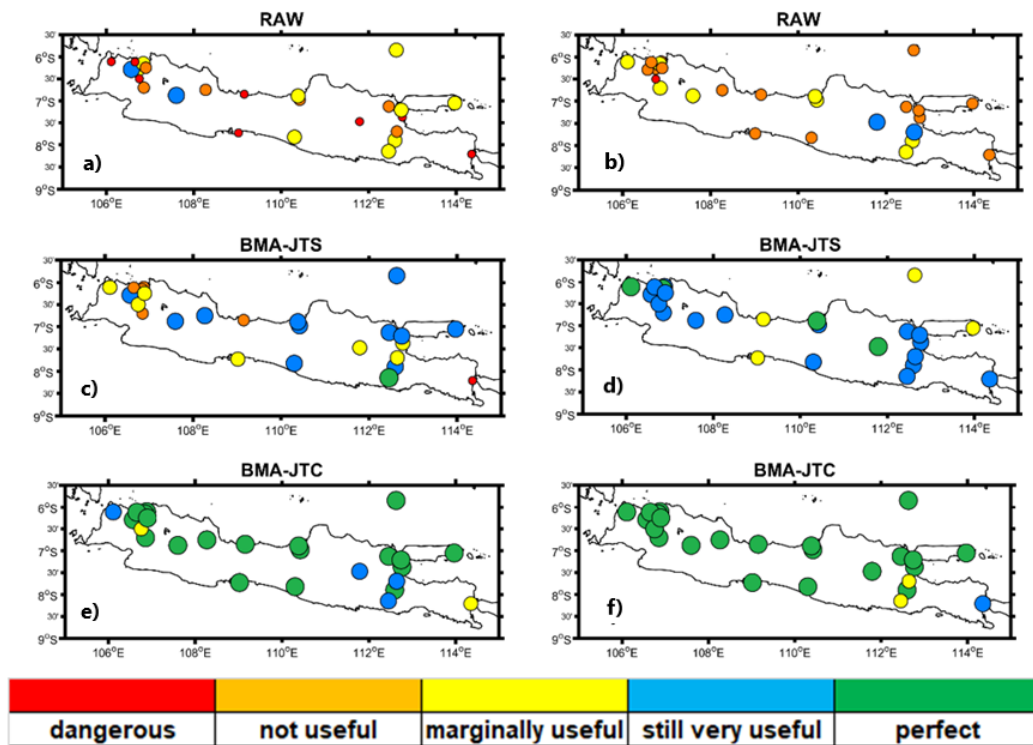
**BMA dengan Jendela Training Conditional**

Penerapan BMA-JTC dilakukan untuk setiap bulan menggunakan skema validasi silang LOOCV per

bulan per tahun. Misal untuk mendapatkan hasil kalibrasi BMA-JTC pada bulan Januari, maka Jan 1981 digunakan untuk bulan testing dan Jan 1982 hingga Jan 2018 sebagai periode training, lalu jika bulan Jan 1982 untuk bulan testing maka Jan 1981, Jan 1983 hingga Jan 2018 sebagai periode training, begitu seterusnya dan hal yang sama dilakukan untuk bulan lainnya. Pada tiap bulan testing yang dikalibrasi, diperoleh PDF BMA yang berbeda. PDF BMA-JTC ditunjukkan oleh kurva hitam pada Gambar 3a dan 3b. Kurva ini terlihat berhimpit dengan PDF klimatologi karena skema kalibrasi BMA-JTC dilakukan untuk karakter bulan yang sama dengan klimatologi. Hal ini menyebabkan, secara umum, hasil kalibrasi dengan pendekatan BMA-JTC lebih baik dari BMA-JTS. Keunggulan BMA-JTC diketahui dari bentuk pola siklus curah hujan tahunan diantara keduanya. Siklus tahunan dari median BMA-JTC lebih dekat dengan observasi dibandingkan BMA-JTS, baik stasiun JND maupun DMG (Gambar 7a, 7b, 7c dan 7d). Begitu juga pola dari sebaran prediksi dari 50% PDF BMA-JTC lebih menyerupai pola observasi dibandingkan BMA-JTS, khususnya pada periode musim hujan di stasiun JND (Gambar 7c). Selain itu pada musim kemarau sebaran BMA-JTC lebih tajam dari BMA-JTS. Kesesuaian pola siklus tahunan pada Gambar 7a, 7b, 7c dan 7d juga dapat diketahui dari *Taylor Diagram* pada Gambar 7e dan 7f. Pada kedua stasiun JND dan DMG, terlihat jelas bahwa median PDF BMA-JTC lebih akurat karena posisinya lebih dekat ke titik referensi dibandingkan median PDF BMA-JTS RAW model.



**Gambar 5.** Perbandingan nilai BSS RAW model ECS4 dan hasil kalibrasi BMA-JTS dengan JT 30 bulan dan BMA-JTC untuk kejadian BN (a, c) dan AN (b, d). Tanda + dan • pada simbol stasiun artinya BSS RAW model bernilai positif dan negatif. Warna pada simbol adalah nilai BSS setelah dikalibrasi BMA.



**Gambar 6.** Perbandingan kategori *Reliability Diagram*. RAW model ECS4 (a, b), hasil kalibrasi BMA-JTS (c, d) dan BMA-JTC (e, f) untuk prediksi probabilistik kejadian BN (kolom kiri) dan AN (kolom kanan)

Perbandingan kedua terlihat dari nilai BSS dan kategori RD. BMA-JTC mampu meningkatkan BSS pada kejadian BN dan AN. Peningkatan BSS positif sekitar 0,1 sampai 0,2 hampir ditemukan pada semua stasiun yaitu sebanyak 25(26) stasiun untuk kejadian BN(AN) (Gambar 5c dan 5d). Dibandingkan dengan BMA-JTS (Gambar 5a dan 5b), BSS positif lebih sedikit hanya sekitar 15(16) pada BN(AN) dan sisanya BSS bernilai negatif. Hal ini menunjukkan bahwa BMA-JTC lebih baik dalam mengurangi tingkat kesalahan dalam menghasilkan prediksi probabilistik kejadian BN dan AN dibandingkan BMA-JTS. Begitu juga pada kategori RD, meskipun BMA-JTS meningkatkan reliabilitas RAW model (Gambar 6c dan 6d), namun peningkatan yang dihasilkan berdasarkan BMA-JTC lebih baik karena kategori RD didominasi oleh kelas 4 dan 5 atau *still very useful* dan *perfect*, untuk kejadian BN dan AN (Gambar 6e dan 6f).

Secara umum uraian di atas menunjukkan hasil secara spasial bahwa BMA-JTC mempunyai kemampuan yang lebih baik dari BMA-JTS hampir di semua stasiun di P. Jawa. Selanjutnya, Gambar 2a menunjukkan terdapat korelasi sangat tinggi ( $R^2 = 0,99$ ) antara *mean* dari PDF BMA-JTC terhadap observasi dibandingkan korelasi antara *ensemble mean* RAW ( $R^2 = 0,70$ ), BCQM ( $R^2 = 0,89$ ) terhadap observasi. Selain itu bentuk VRH BMA-JTC yang lebih datar (Gambar 2b) juga menandakan bahwa BMA-JTC mampu memperbaiki kesalahan dispersi yang terdapat dalam anggota *ensemble* RAW model.

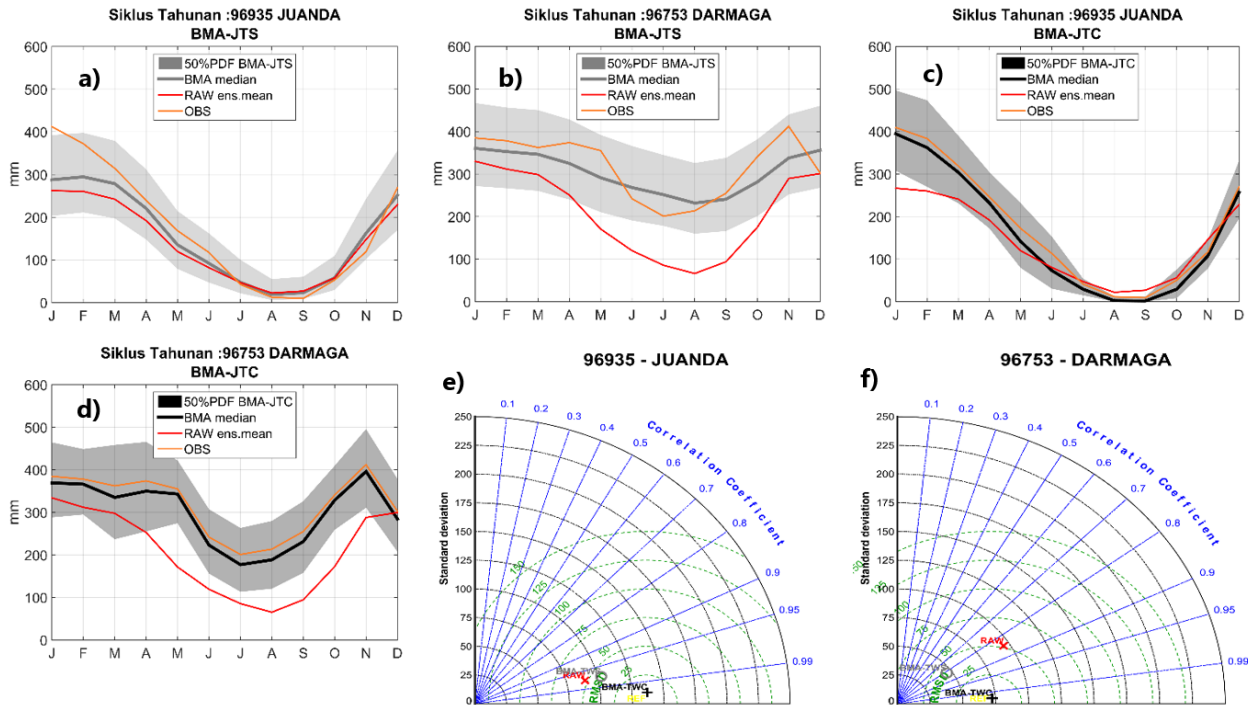
## KESIMPULAN

Penelitian ini mengkaji kemampuan metode BMA dalam memperbaiki kualitas prediksi musim di Indonesia khususnya P. Jawa. Prediksi musim yang dikalibrasi berupa prediksi *ensemble* curah hujan bulanan dari keluaran model ECS4. BMA dilakukan dengan dua jenis jendela *training*, yaitu: BMA-JTS dan BMA-JTC. Kedua model memperbaharui informasi prediksi dan observasi pada jendela *training* tetapi dengan cara yang berbeda. BMA-JTC memperbaharui jendela *training* dengan bulan yang sama dengan periode klimatologinya, sementara BMA-JTS tanpa memerhatikan jenis bulan dalam jendela *training*.

Kemampuan metode BMA-JTS dan BMA-JTC telah dievaluasi, dan hasilnya menunjukkan bahwa keduanya mampu memperbaiki karakteristik distribusi RAW model, seperti dispersi anggota *ensemble* yang lebih baik berdasarkan VRH yang lebih datar dan tidak *U-shaped*. Kemudian BMA-JTS dan BMA-JTC juga mampu memperbaiki kualitas prediksi probabilitas kejadian BN dan AN untuk semua stasiun BMKG di P. Jawa. Namun demikian, kemampuan BMA-JTC lebih unggul dari BMA-JTS. Keunggulan BMA-JTC diketahui dari jumlah stasiun dengan reliabilitas *still very useful* dan *perfect*, serta BSS positif, lebih banyak dibandingkan BMA-JTS dan juga RAW model. Selain itu, akurasi pola siklus tahunan dari BMA-JTC lebih mendekati pola observasi dibandingkan BMA-JTS dan RAW model.

Secara umum, hasil penelitian ini menunjukkan BMA dapat digunakan sebagai salah satu metode *post processing* untuk prediksi *ensemble* dari keluaran prediksi musim (jangka panjang) model iklim global,

seperti halnya SPE model ECS4. Dibandingkan metode koreksi bias gQM, BMA tidak hanya mengurangi bias namun juga memperbaiki kesalahan dispersi anggota *ensemble*.



**Gambar7.** Perbedaan pola siklus tahunan curah hujan di stasiun Juanda (JND) dan Darmaga (DMG) yang dihasilkan dari BMA-JTS (a, b) dan BMA-JTC (c, d). Taylor Diagram berupa hubungan koefisien korelasi (garis lengkung biru), standar deviasi (garis lengkung biru) dan RMSE (garis lengkung hitam), menunjukkan akurasi pola siklus tahunan antara median PDF BMA-JTS (\* abu-abu), median PDF BMA-JTC (+hitam) dan *ensemble mean* RAW model ECS4 (x merah) terhadap nilai observasi untuk stasiun Juanda (e) dan Darmaga (f).

**DAFTAR PUSTAKA**

Abraham, M.S., Puthiyidam, J.J., 2016. A Survey on Wind Data Pre-Processing in Electricity Generation. *Int. J. Cybern. Inform.* 5.

Alessandrini, S., Delle Monache, L., Sperati, S., Cervone, G., 2015. An analog *ensemble* for short-term probabilistic solar power forecast. *Appl. Energy* 157, 95–110. <https://doi.org/10.1016/j.apenergy.2015.08.011>

Baran, S., Hemri, S., El Ayari, M., 2019. Statistical Postprocessing of Water Level Forecasts Using Bayesian Model Averaging With Doubly Truncated Normal Components. *Water Resour. Res.* 55, 3997–4013. <https://doi.org/10.1029/2018WR024028>

Baran, S., Möller, A., 2017. Bivariate *ensemble* model output statistics approach for joint forecasting of wind speed and temperature. *Meteorol. Atmospheric Phys.* 129, 99–112. <https://doi.org/10.1007/s00703-016-0467-8>

Baran, S., Möller, A., 2015. Joint probabilistic forecasting of wind speed and temperature using Bayesian model averaging. *Environmetrics* 26, 120–132. <https://doi.org/10.1002/env.2316>

Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. *Nature* 525, 47–55. <https://doi.org/10.1038/nature14956>

Bazile, R., Boucher, M.-A., Perreault, L., Leconte, R., 2017. Verification of ECMWF System 4 for seasonal hydrological forecasting in a northern climate. *Hydrol. Earth Syst. Sci.* 21, 5747. <https://doi.org/10.5194/hess-21-5747-2017>

Berner, J., Fossell, K.R., Ha, S.-Y., Hacker, J.P., Snyder, C., 2015. Increasing the Skill of Probabilistic Forecasts: Understanding Performance Improvements from Model-Error Representations. *Mon. Weather Rev.* 143, 1295–1320. <https://doi.org/10.1175/MWR-D-14-00091.1>

Eide, S.S., Bremnes, J.B., Steinsland, I., 2017. Bayesian Model Averaging for Wind Speed *Ensemble*

- Forecasts Using Wind Speed and Direction. *Weather Forecast.* 32, 2217–2227. <https://doi.org/10.1175/WAF-D-17-0091.1>
- Erickson, M.J., Colle, B.A., Charney, J.J., 2012. Impact of bias-correction type and conditional training on Bayesian model averaging over the northeast United States. *Weather Forecast.* 27, 1449–1469. <https://doi.org/10.1175/WAF-D-11-00149.1>
- Fang, M., Li, X., 2015. Application of Bayesian Model Averaging in the Reconstruction of Past Climate Change Using PMIP3/CMIP5 Multimodel *Ensemble* Simulations. *J. Clim.* 29, 175–189. <https://doi.org/10.1175/JCLI-D-14-00752.1>
- Feldmann, K., Scheuerer, M., Thorarinsdottir, T.L., 2015. Spatial Postprocessing of *Ensemble* Forecasts for Temperature Using Nonhomogeneous Gaussian Regression. *Mon. Weather Rev.* 143, 955–971. <https://doi.org/10.1175/MWR-D-14-00210.1>
- Fraley, C., Raftery, A.E., Gneiting, T., Slaughter, J.M., 2007. *EnsembleBMA*: An R package for probabilistic forecasting using *ensembles* and Bayesian model averaging. WASHINGTON UNIV SEATTLE DEPT OF STATISTICS.
- Gneiting, T., 2014. Calibration of medium-range weather forecasts. European Centre for Medium-Range Weather Forecasts.
- Guan, H., Cui, B., Zhu, Y., 2015. Improvement of Statistical Postprocessing Using GEFS Reforecast Information. *Weather Forecast.* 30, 841–854. <https://doi.org/10.1175/WAF-D-14-00126.1>
- Han, K., Choi, J., Kim, C., 2018. Comparison of statistical post-processing methods for probabilistic wind speed forecasting. *Asia-Pac. J. Atmospheric Sci.* 54, 91–101. <https://doi.org/10.1007/s13143-017-0062-z>
- Han, K., Choi, J., Kim, C., 2016. Comparison of prediction performance using statistical postprocessing methods. *Asia-Pac. J. Atmospheric Sci.* 52, 495–507. <https://doi.org/10.1007/s13143-016-0034-8>
- He, S., Guo, S., Liu, Z., Yin, J., Chen, K., Wu, X., 2018. Uncertainty analysis of hydrological multi-model *ensembles* based on CBP-BMA method. *Hydrol. Res.* 49, 1636–1651. <https://doi.org/10.2166/nh.2018.160>
- Hudson, D., Ebert, B., 2017. *Ensemble* Verification Metrics. ECMWF Annual Seminar 2017 39.
- Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J.B., Grell, G., Smirnova, T.G., Benjamin, S.G., Brown, J.M., 2017. A Performance Comparison between Multiphysics and Stochastic Approaches within a North American RAP *Ensemble*. *Mon. Weather Rev.* 145, 1161–1179. <https://doi.org/10.1175/MWR-D-16-0160.1>
- Ji, L., Zhi, X., Zhu, S., Fraedrich, K., 2019. Probabilistic Precipitation Forecasting over East Asia Using Bayesian Model Averaging. *Weather Forecast.* 34, 377–392. <https://doi.org/10.1175/WAF-D-18-0093.1>
- Khajehei, S., Ahmadalipour, A., Moradkhani, H., 2018. An effective post-processing of the North American multi-model *ensemble* (NMME) precipitation forecasts over the continental US. *Clim. Dyn.* 51, 457–472. <https://doi.org/10.1007/s00382-017-3934-0>
- Kolachian, R., Saghafian, B., 2019. Deterministic and probabilistic evaluation of raw and post processed sub-seasonal to seasonal precipitation forecasts in different precipitation regimes. *Theor. Appl. Climatol.* 137, 1479–1493. <https://doi.org/10.1007/s00704-018-2680-5>
- Kontis, V., Bennett, J.E., Mathers, C.D., Li, G., Foreman, K., Ezzati, M., 2017. Future life expectancy in 35 industrialised countries: projections with a Bayesian model *ensemble*. *The Lancet* 389, 1323–1335. [https://doi.org/10.1016/S0140-6736\(16\)32381-9](https://doi.org/10.1016/S0140-6736(16)32381-9)
- Kuswanto, H., Rahadiyuza, D., Gunawan, D., 2019. Probabilistic Precipitation Forecast in (Indonesia) Using NMME Models: Case Study on Dry Climate Region, in: Chaminé, H.I., Barbieri, M., Kisi, O., Chen, M., Merkel, B.J. (Eds.), *Advances in Sustainable and Environmental Hydrology, Hydrogeology, Hydrochemistry and Water Resources*. Springer International Publishing, Cham, pp. 13–16.
- L’Heureux, M.L., Takahashi, K., Watkins, A.B., Barnston, A.G., Becker, E.J., Di Liberto, T.E., Gamble, F., Gottschalck, J., Halpert, M.S., Huang, B., Mosquera-Vásquez, K., Wittenberg, A.T., 2016. Observing and Predicting the 2015/16 El Niño. *Bull. Am. Meteorol. Soc.* 98, 1363–1382. <https://doi.org/10.1175/BAMS-D-16-0009.1>
- Liu, Y., Zhao, G., Peng, X., 2019. Deep Learning Prognostics for Lithium-Ion Battery Based on *Ensembled* Long Short-Term Memory Networks. *IEEE Access* 7, 155130–155142. <https://doi.org/10.1109/ACCESS.2019.2937798>
- Manzanas, R., Lucero, A., Weisheimer, A., Gutiérrez, J.M., 2018. Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts? *Clim. Dyn.* 50, 1161–1176. <https://doi.org/10.1007/s00382-017-3668-z>

- Maraun, D., Shepherd, T.G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J.M., Hagemann, S., Richter, I., Soares, P.M.M., Hall, A., Mearns, L.O., 2017. Towards process-informed bias correction of climate change simulations. *Nat. Clim. Change* 7, 764–773. <https://doi.org/10.1038/nclimate3418>
- Mendoza, P.A., Rajagopalan, B., Clark, M.P., Ikeda, K., Rasmussen, R.M., 2015. Statistical postprocessing of high-resolution regional climate model output. *Mon. Weather Rev.* 143, 1533–1553.
- Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., Vitart, F., 2011. The new ECMWF seasonal forecast system (System 4). European Centre for Medium-Range Weather Forecasts Reading.
- Monier, E., Gao, X., Scott, J.R., Sokolov, A.P., Schlosser, C.A., 2015. A framework for modeling uncertainty in regional climate change. *Clim. Change* 131, 51–66. <https://doi.org/10.1007/s10584-014-1112-5>
- Olson, R., Fan, Y., Evans, J.P., 2016. A simple method for Bayesian model averaging of regional climate model projections: Application to southeast Australian temperatures. *Geophys. Res. Lett.* 43, 7661–7669. <https://doi.org/10.1002/2016GL069704>
- Pastén-Zapata, E., Jones, J.M., Moggridge, H., Widmann, M., 2020. Evaluation of the performance of Euro-CORDEX Regional Climate Models for assessing hydrological climate change impacts in Great Britain: A comparison of different spatial resolutions and quantile mapping bias correction methods. *J. Hydrol.* 584, 124653. <https://doi.org/10.1016/j.jhydrol.2020.124653>
- Peeters, T., 2018. Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *Int. J. Forecast.* 34, 17–29. <https://doi.org/10.1016/j.ijforecast.2017.08.002>
- Pelosi, A., Medina, H., Van den Bergh, J., Vannitsem, S., Chirico, G.B., 2017. Adaptive Kalman Filtering for Postprocessing *Ensemble* Numerical Weather Predictions. *Mon. Weather Rev.* 145, 4837–4854. <https://doi.org/10.1175/MWR-D-17-0084.1>
- Piani, C., Haerter, J., Coppola, E., 2010. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.* 99, 187–192. <https://doi.org/10.1007/s00704-009-0134-9>
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian Model Averaging to Calibrate Forecast *Ensembles*. *Mon. Weather Rev.* 133, 20. <https://doi.org/10.1175/MWR2906.1>
- Schefzik, R., 2016. A Similarity-Based Implementation of the Schaake Shuffle. *Mon. Weather Rev.* 144, 1909–1921. <https://doi.org/10.1175/MWR-D-15-0227.1>
- Schepen, A., Wang, Q.J., Everingham, Y., 2016. Calibration, Bridging, and Merging to Improve GCM Seasonal Temperature Forecasts in Australia. *Mon. Weather Rev.* 144, 2421–2441. <https://doi.org/10.1175/MWR-D-15-0384.1>
- Schmeits, M.J., Kok, K.J., 2010. A comparison between raw *ensemble* output, (modified) Bayesian model averaging, and extended logistic regression using ECMWF *ensemble* precipitation reforecasts. *Mon. Weather Rev.* 138, 4199–4211. <https://doi.org/10.1175/2010MWR3285.1>
- Seong, M.-G., Suh, M.-S., Kim, C., 2017. Intercomparison of prediction skills of *ensemble* methods using monthly mean temperature simulated by CMIP5 models. *Asia-Pac. J. Atmospheric Sci.* 53, 339–351. <https://doi.org/10.1007/s13143-017-0039-y>
- Siegert, S., Bhend, J., Kroener, I., De Felice, M., 2014. SpecsVerification: Forecast Verification Routines for *Ensemble* Forecasts of Weather and Climate.
- Sillmann, J., Thorarindottir, T., Keenlyside, N., Schaller, N., Alexander, L.V., Hegerl, G., Seneviratne, S.I., Vautard, R., Zhang, X., Zwiers, F.W., 2017. Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather Clim. Extrem.* 18, 65–74. <https://doi.org/10.1016/j.wace.2017.10.003>
- Siuta, D., West, G., Stull, R., Nipen, T., 2017. Calibrated Probabilistic Hub-Height Wind Forecasts in Complex Terrain. *Weather Forecast.* 32, 555–577. <https://doi.org/10.1175/WAF-D-16-0137.1>
- Sloughter, J.M., Gneiting, T., Raftery, A.E., 2010. Probabilistic wind speed forecasting using *ensembles* and Bayesian model averaging. *J. Am. Stat. Assoc.* 105, 25–35. <https://doi.org/10.1198/jasa.2009.ap08615>
- Sloughter, J.M.L., Raftery, A.E., Gneiting, T., Fraley, C., 2007. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Mon. Weather Rev.* 135, 3209–3220. <https://doi.org/10.1175/MWR3441.1>

- Song, X., Zhu, Y., Peng, J., Guan, H., 2018. Improving Multi-Model *Ensemble* Forecasts of Tropical Cyclone Intensity Using Bayesian Model Averaging. *J. Meteorol. Res.* 32, 794–803. <https://doi.org/10.1007/s13351-018-7117-7>
- Späth, S., von Bremen, L., Junk, C., Heinemann, D., 2015. Time-consistent calibration of short-term regional wind power *ensemble* forecasts. *Meteorol Z* 24, 381–392.
- Supari, Tangang, F., Salimun, E., Aldrian, E., Sopaheluwakan, A., Juneng, L., 2018. ENSO modulation of seasonal rainfall and extremes in Indonesia. *Clim. Dyn.* 51, 2559–2580. <https://doi.org/10.1007/s00382-017-4028-8>
- Taillardat, M., Mestre, O., Zamo, M., Naveau, P., 2016. Calibrated *Ensemble* Forecasts Using Quantile Regression Forests and *Ensemble* Model Output Statistics. *Mon. Weather Rev.* 144, 2375–2393. <https://doi.org/10.1175/MWR-D-15-0260.1>
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmospheres* 106, 7183–7192. <https://doi.org/10.1029/2000JD900719>
- Vogel, P., Knippertz, P., Fink, A.H., Schlueter, A., Gneiting, T., 2018. Skill of Global Raw and Postprocessed *Ensemble* Predictions of Rainfall over Northern Tropical Africa. *Weather Forecast.* 33, 369–388. <https://doi.org/10.1175/WAF-D-17-0127.1>
- Wang, H., Li, G., Wang, G., Peng, J., Jiang, H., Liu, Y., 2017. Deep learning based *ensemble* approach for probabilistic wind power forecasting. *Appl. Energy* 188, 56–70. <https://doi.org/10.1016/j.apenergy.2016.11.111>
- Wang, J., Chen, J., Du, J., Zhang, Y., Xia, Y., Deng, G., 2018. Sensitivity of *Ensemble* Forecast Verification to Model Bias. *Mon. Weather Rev.* 146, 781–796. <https://doi.org/10.1175/MWR-D-17-0223.1>
- Wang, Q., Shao, Y., Song, Y., Schepen, A., Robertson, D.E., Ryu, D., Pappenberger, F., 2019. An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm. *Environ. Model. Softw.* 122, 104550.
- Weisheimer, A., Palmer, T.N., 2014. On the reliability of seasonal climate forecasts. *J. R. Soc. Interface* 11, 20131162–20131162. <https://doi.org/10.1098/rsif.2013.1162>
- Wilson, L.J., Beaugregard, S., Raftery, A.E., Verret, R., 2007. Calibrated surface temperature forecasts from the Canadian *ensemble* prediction system using Bayesian model averaging. *Mon. Weather Rev.* 135, 1364–1385. <https://doi.org/10.1175/MWR3347.1>
- Xu, J., Ancil, F., Boucher, M.-A., 2019. Hydrological post-processing of streamflow forecasts issued from multimodel *ensemble* prediction systems. *J. Hydrol.* 578, 124002. <https://doi.org/10.1016/j.jhydrol.2019.124002>
- Yang, X., Sharma, S., Siddique, R., Greybush, S.J., Mejia, A., 2017. Postprocessing of GEFS Precipitation *Ensemble* Reforecasts over the U.S. Mid-Atlantic Region. *Mon. Weather Rev.* 145, 1641–1658. <https://doi.org/10.1175/MWR-D-16-0251.1>
- Žagar, N., Buizza, R., Tribbia, J., 2015. A Three-Dimensional Multivariate Modal Analysis of Atmospheric Predictability with Application to the ECMWF *Ensemble*. *J. Atmospheric Sci.* 72, 4423–4444. <https://doi.org/10.1175/JAS-D-15-0061.1>
- Zhao, T., Bennett, J.C., Wang, Q., Schepen, A., Wood, A.W., Robertson, D.E., Ramos, M.-H., 2017. How suitable is quantile mapping for postprocessing GCM precipitation forecasts? *J. Clim.* 30, 3185–3196. <https://doi.org/10.1175/JCLI-D-16-0652.1>
- Zhong, Y., Guo, S., Ba, H., Xiong, F., Chang, F.-J., Lin, K., 2018. Evaluation of the BMA probabilistic inflow forecasts using TIGGE numeric precipitation predictions based on artificial neural network. *Hydrol. Res.* 49, 1417–1433. <https://doi.org/10.2166/nh.2018.177>
- Zhu, Y., Luo, Y., 2015. Precipitation Calibration Based on the Frequency-Matching Method. *Weather Forecast.* 30, 1109–1124. <https://doi.org/10.1175/WAF-D-13-00049.1>