UJM 4 (1) (2015)

UNNES Journal of Mathematics

http://journal.unnes.ac.id/sju/index.php/ujm

PERBANDINGAN PERAMALAN CURAH HUJAN DENGAN METODE BAYESIAN MODEL AVERAGING DAN KALMAN FILTER

Noviesag Artanto✉, Arief Agoestanto

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 Lt.1, Kampus Sekaran Gunungpati, Semarang 50299

Info Artikel

Sejarah Artikel:
Diterima Oktober 2014
Disetujui Desember 2014
Dipublikasikan Mei 2015

Keywords:
Peramalan Curah Hujan; Bayesian Model Averaging; Kalman Filter

Abstrak

Pada era sekarang tanpa disadari dampak dari pemanasan global semakin terasa, salah satunya berupa perubahan iklim. Perubahan cuaca seperti musim hujan yang tak menentu merupakan salah satu tanda berubahnya iklim. Padahal dalam suatu jenis pekerjaan tertentu hujan yang turun dapat mempengaruhi hasil produksi. Oleh karena itu dibutuhkanlah prediksi untuk mengetahui curah hujan yang turun pada suatu tempat sehingga tidak merugikan aktifitas manusia. Dalam penelitian ini akan dikaji tentang membandingan dua metode peramalan yaitu metode Bayesian Model Averaging dengan metode Kalman Filter. Dengan membandingan kedua metode peramalan tersebut diharapkan dapat diketahui metode mana yang ternyata memberikan hasil prediksi yang lebih optimal dalam menentukan tingginya curah hujan.

Abstract

In the current era, we almost do not realize that the impact of global warming is getting more real. One of the impact is in the form of climate change. Changes in weather such erratic rainy season refers to one sign of the changing of climate. Where as, in a certain type of work, rain can affect the production process. Therefore, it is needed the prediction to determine the rainfall in one particular place so that rain will not be detrimental to human activities. This study will as sesson comparing that who methods of forecasting namely Bayesian method with averaging models and Kalman Filter. By comparing the two methods of forecasting, it is expected to be known which method turned out to providea more optimal out come prediction in determining the high intensity of rainfall.

✉Alamat korespondensi:
E-mail: noviesag@gmail.com

ISSN 2252-6943

Pendahuluan

Pada era sekarang tanpa disadari dampak dari pemanasan global semakin terasa, salah satunya berupa perubahan iklim. LAPAN (2009) menyatakan bahwa dampak ekstrem dari perubahan iklim terutama terjadinya kenaikan temperatur serta pergeseran musim. Selain itu, perubahan iklim dapat memberikan dampak pada berubahnya suhu udara, curah hujan, tekanan udara, dan kecepatan angin. Dari keempat faktor ini, curah hujan merupakan salah satu faktor yang memberikan pengaruh besar dalam kehidupan manusia, antara lain sebagai faktor penentu penerbangan, pelayaran, pertanian, serta merupakan faktor penting dalam mempengaruhi prakiraan cuaca. Perubahan iklim yang terjadi saat ini membuat curah hujan menjadi tak menentu dibandingkan musim-musim sebelumnya. Untuk meminimalisir dampak yang ditimbulkan oleh curah hujan yang tak menentu yakni dengan menyediakan informasi mengenai peluang terjadinya cuaca ekstrem seperti prediksi curah hujan di suatu daerah dalam jangka waktu tertentu. Pentingnya meramalkan curah hujan inilah yang berguna untuk menghindari timbulnya kerugian bagi kehidupan manusia. Inilah yang menjadi latar belakang dari adanya penelitian ini, terutama bagi kota besar seperti Kota Semarang.

Beberapa penelitian mengenai peramalan telah dilakukan sebelumnya, antara lain oleh Raftery (2003) dengan menggunakan metode Bayesian Model Averaging, Sloughter (2010) menggunakan metode peramalan *ensemble* dan Bayesian Model Averaging, dan Tresnawati (2010) dengan menggunakan metode Kalman Filter. Sementara itu banyak pusat prediksi di negara-negara maju seperti NAEFS, telah mengembangkan model peramalan dengan *ensemble* (*Ensemble Prediction System*, EPS) yang akan memberikan luaran berupa taksiran interval. Zhu (2005) juga menyatakan, kelebihan dari penggunaan model *ensemble* dibandingkan peramalan yang menghasilkan taksiran titik adalah dapat menangkap adanya unsur / faktor ketidakpastian. Raftery dkk (2005) menyatakan salah satu karakteristik dari peramalan *ensemble* adalah sering kali bersifat *underdispersive* yaitu hasil peramalan yang dihasilkan cenderung terpusat pada suatu titik tertentu dengan nilai varians yang relatif rendah atau terkadang *overdispersive* yaitu hasil peramalan dengan nilai varians yang relatif besar. Oleh karena itu yang akan dilakukan ialah suatu proses kalibrasi yang bertujuan untuk memperbaiki peramalan *ensemble* yaitu dengan menggunakan metode Bayesian Model Averaging (BMA-EM). Bayesian Model Averaging merupakan suatu metode kalibrasi untuk menggeser atau melakukan penyesuaian terhadap nilai mean dan varian dari hasil peramalan sedemikian rupa sehingga dapat mendekati nilai observasi. Kelebihan dari metode Bayesian Model Averaging selain dapat menggeser nilai mean sehingga dapat memperkecil bias yaitu dengan adanya bobot yang diperoleh dari algoritma EM. Kelemahan dari penelitian tersebut yakni perlu menggunakan banyak model arima sehingga terbentuk model peramalan semakin akurat, penentuan pemilihan training set atau rentang waktu mempengaruhi hasil peramalan sehingga perlu dilakukan berbagai macam uji coba.

Selain dengan metode BMA penelitian lain dilakukan dengan metode Kalman Filter. Kalman Filter merupakan salah satu metode runtun waktu yang dapat digunakan dalam menentukan ramalan ke depan. Metode ini bekerja secara rekursif untuk meminimalkan ketidaktepatan dalam peramalan. Kalman Filter terdiri dari persamaan keadaan dan observasi. Metode ini digunakan untuk menyatakan suatu model runtun waktu yang ditampilkan dalam bentuk linier *state space* (Brockwell and Davis, 1991). Menurut Meinhold dan Singpurwala (1983), model, teknik, dan notasi dari Kalman Filter hampir sama dengan model regresi linier dan analisis runtun waktu. Perbedaannya terletak pada sifat rekursif yang ada pada Kalman Filter (Welch and Gary, 2001).

Kalman Filter juga memiliki kelemahan menurut Wei (2006) yaitu keberhasilan dalam mendapatkan hasil prediksi optimal bergantung pada ketepatan estimasi keadaan (*state*) awal pada data observasi terbaru. Oleh karena itu, dalam penelitian ini peneliti akan membandingkan hasil penaksiran curah hujan antara BMA-EM dengan metode Kalman Filter untuk tahun 2013. Sehingga dapat di peroleh hasil yang berbeda dan lebih akurat dari penelitian yang sudah dilakukan sebelumnya.

Runtun waktu atau *time series* adalah data yang dikumpulkan dari waktu ke waktu untuk menggambarkan perkembangan suatu kegiatan (perkembangan produksi, harga, hasil penjualan, jumlah personil, penduduk, jumlah kecelakaan, jumlah kejahatan, dan lain sebagainya) (Supranto,200:214).

Peramalan dengan metode Bayesian Model Averaging merupakan peramalan yang dapat untuk meramalkan data musiman, oleh karena itu metode ini merupakan bagian dari metode Autoregressive Integrated Moving Average (SARIMA).

Musiman adalah kecenderungan mengulangi pola tingkah gerak dalam periode musim, biasanya satu tahun untuk data bulanan. Model ARIMA Musiman merupakan model ARIMA yang digunakan untuk menyelesaikan time series musiman yang terdiri dari dua bagian, yaitu bagian tidak musiman (non-musiman) dan bagian musiman. Bagian non-musiman dari metode ini adalah model ARIMA. Secara umum bentuk model ARIMA musiman atau ARIMA $(p,d,q)(P,Q,S)^S$adalah:

$$\Phi_P(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D X_t = \Theta_q(B)\Theta_q(B^S)e_t$$

Pada analisis statistik, seperti analisis regresi, biasanya mengacu pada hasil bersyarat pada satu asumsi model statistik. Seringkali model ini telah dipilih dari antara beberapa model yang mungkin untuk data, dan analis data tidak yakin bahwa itu adalah yang terbaik. Karena model yang masuk akal lainnya juga bisa memberikan perbedaan jawaban. Ini merupakan sumber ketidakpastian dalam menarik kesimpulan, dan pendekatan yang khas, yang kemungkinan pada model tunggal dianggap terbaik dengan mengabaikan sumber ketidakpastian, sehingga meremehkan ketidakpastian tersebut.

Perataan model bayes (Leamer 1978; Kass dan Raftery 1995; Hoeting, Madigan, Raftery, dan Volinsky 1999) mengatasi masalah ini dengan pengkondisian, tidak pada satu model terbaik, tetapi pada seluruh *ensemble* model statistik pertama yang dipertimbangkan. Dalam kasus y kuantitas yang akan diperkirakan berdasarkan data training y^T menggunakan model statistik K,$\{M_1,\dots,M_K\}$, hukum probabilitas jumlah memberitahu bahwa perkiraan PDF, $p(y)$, diberikan melalui

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

Dimana $p(y|M_k)$ adalah ramalan PDF berdasarkan M_k saja, dan $p(M_k|Y^T)$ adalah probabilitas posterior M_k yang diberikan pada training data, dan mencerminkan seberapa baik Model M_k cocok dengan data pelatihan. Model posterior probabilitas menambahkan hingga satu , sehingga $\sum_{k=1}^{K} p(M_k|Y^T) = 1$, sehingga hal itu bisa dipandang sebagai bobot. BMA PDF adalah rata-rata tertimbang dari PDF yang diberikan model individual, dihitung dengan probabilitas posterior Model mereka. BMA memiliki berbagai sifat optimalitas teoritis dan telah menunjukkan kinerja yang baik dalam berbagai situasi data simulasi dan nyata (Raftery dan Zheng 2003).

Selanjutnya memperluas BMA dari model statistik untuk model dinamis. Ide dasarnya adalah bahwa untuk setiap perkiraan yang diberikan merupakan model terbaik, namun tidak tahu apa itu, dan ketidakpastian tentang model terbaik adalah kuantitas oleh BMA. Sekali lagi, y dilambangkan dengan kuantitas yang akan diramalkan. Setiap peramalan deterministik, $\tilde{f}_k$, bisa bias dikoreksi, menghasilkan peramalan bias dikoreksi$\tilde{f}_k$. Peramalan $\tilde{f}_k$kemudian dikaitkan dengan PDF bersyarat, $g_k(y|\tilde{f}_k)$, yang dapat diartikan sebagai PDF bersyarat y tergantung pada $\tilde{f}_k$, mengingat bahwa $\tilde{f}_k$ adalah perkiraan terbaik dalam *ensemble*. Model prediksi BMA menjadi

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

dimana w_k adalah probabilitas posterior dari perkiraan k menjadi yang terbaik, dan didasarkan pada keterampilan perkiraan k pada periode pelatihan. w_k adalah probabilitas dan sehingga mereka menambahkan hingga 1, $\sum_{k=1}^{K} w_k = 1$.

Ketika peramalan suhu dan tekanan permukaan laut, sering tampak masuk akal untuk mendekati PDF bersyarat oleh distribusi normal berpusat di $\tilde{f}_k$, sehingga $g_k(y|\tilde{f}_k)$ adalah PDF normal dengan mean $\tilde{f}_k$dan sebuah *ensemble* anggota spesifik deviasi standar k. Sehingga diperoleh situasi seperti ini:

$$y|\tilde{f}_k \sim \mathrm{N}(\tilde{f}_k\sigma_k^2)$$

Dalam hal ini, prediksi rata-rata BMA hanya ekspektasi bersyarat dari y diberikan perkiraan, yaitu

$$E[y|f_1,\dots,f_k] = \sum_{k=1}^{K} w_k\tilde{f}_k$$

(Raftery dkk:2003) (1)

Hal ini dapat dilihat sebagai perkiraan deterministik dalam dirinya sendiri, dan dibandingkan dengan perkiraan individu dalam *ensemble* atau rata-rata *ensemble*.

Algoritma EM adalah metode untuk menemukan estimator maksimum likelihood yang bermasalah dalam hal data yang hilang sehingga jika dapat mengetahui data yang hilang masalah estimasi akan mudah. Data yang hilang tidak harus data aktual yang hilang dan sebaliknya, data tersebut sering tidak teramati, pengetahuan yang akan menyederhanakan masalah estimasi model BMA adalah model campuran terbatas (McLachlan dan Peel 2000). Disini data yang hilang z_{kst}yang mana z_{kst} = 1 jika anggota *ensemble* k adalah perkiraan terbaik untuk tempat dan waktu t, dan z_{kst} = 0 sebaliknya. Untuk setiap (s,t), hanya satu dari $\{Z_{1st},\dots,Z_{Kst}\}$ sama dengan 1 yang lain adalah nol.

Algoritma EM merupakan proses iteratif dan bergantian antara dua langkah, E langkah (expectation atau harapan), dan M langkah (maximization atau maksimalisasi). Dimulai dengan menebak awal $\theta^{(0)}$, untuk vektor parameter θ. Pada langkah E, z_{kst} yang diperkirakan diberikan menebak saat ini untuk parameter perkiraan z_{kst} yang tidak selalu bilangan bulat, meskipun nilai-nilai yang benar adalah 0 atau 1. Pada langkah M, θ diperkirakan mengingat nilai-nilai saat ini dari z_{kst}tersebut.

Untuk langkah E adalah

$$\hat{z}_{kst}^{(j)} = \frac{g(y_{st}|\tilde{f}_{kst},\sigma_k^{(j-1)})}{\sum_{i=1}^{K} g(y_{st}|\tilde{f}_{ist},\sigma_i^{(j-1)})}$$

dimana j *superscript* mengacu pada iterasi j dari algoritma EM, dan $g(y_{st}|\tilde{f}_{kst},\sigma_k^{(j-1)})$ adalah densitas normal dengan mean $\tilde{f}_{kst}$dan standar deviasi $\sigma_k^{(j-1)}$ dievaluasi pada y_{st}. Langkah M kemudian terdiri dari memperkirakan w_k dan σ_k menggunakan sebagai bobot perkiraan saat z_{kst}, yaitu $\hat{Z}_{kst}^{(j)}$. Demikian

$$w_k^{(j)} = \frac{1}{n}\sum_{s,t} \hat{z}_{kst}^{(j)},$$

$$\sigma_k^{2(j)} = \frac{\sum_{s,t} \hat{z}_{kst}^{(j)} (y_{st} - \tilde{f}_{kst})^2}{\sum_{s,t} \hat{z}_{kst}^{(j)}}$$

(Raftery dkk : 2003)

di mana n adalah jumlah observasi yaitu nilai-nilai yang berbeda (s,t)).

Dari tahap E dan M kemudian diiterasi untuk konvergensi, yang didefinisikan sebagai perubahan tidak lebih besar dari beberapa toleransi kecil di salah satu log-likelihood, nilai-nilai parameter, atau $\hat{z}_{kst}^{(j)}$ dalam satu iterasi. Log-likelihood dijamin akan meningkat pada setiap iterasi EM (Wu 1983), yang menunjukkan bahwa secara umum menyatu dengan maksimum lokal dari kemungkinan. Konvergensi maksimum global yang tidak dapat dijamin, sehingga solusi yang dicapai oleh algoritma dapat peka terhadap nilai-nilai awal. Mulai nilai berdasarkan pengalaman masa lalu biasanya memberikan solusi yang baik. Dalam implementasi, *training set* terdiri dari perkiraan dan pengamatan untuk $m$waktu ke waktu sebelumnya.

Hasil peramalan dari kalibrasi ensembel berubah sebuah taksiran interval dalam bentuk pdf. Oleh karena itu, evaluasi kebaikan model tidak bisa dilakukan dengan prosedur standar MSE atau MAPE. Sehingga dalam penelitian ini akan digunakan evaluasi kebaikan model dengan prosedur CRPS. CRPS berkaitan dengan probabilitas rank / pangkat, yang membandingkan distribusi hasil ramalan dengan pengamatan, dimana keduanya direpresentasikan sebagai fungsi distribusi kumulatif (cdf).

$$CRPS = \frac{1}{ncase}\sum_{i=1}^{ncase}\int_{X=-x}^{X=x} (F_i^f(x) - F_i^0(x))^2\ dx$$

(Raftery dkk : 2003)

$F_i^f(x)$: cdf dari hasil peramalan (ke-i)
$F_i^0(x)$: cdf dari pengamatan sebenarnya (ke-i)
$ncase$: jumlah peramalan

Hasil dari CRPS ini berupa suatu nilai, dimana peramalan akan dikatakan bagus bila CRPS yang dihasilkan sangat kecil atau mendekati nol.

Kalman Filter adalah sebuah metode bagian dari ruang keadaan (*state space*) yang dapat diterapkan dalam model prakiraan statistik. Sesuai dengan Wei (2006), metode ini menggunakan teknik rekursif dalam mengintegrasikan data pengamatan terbaru ke model untuk mengoreksi prediksi sebelumnya dan melakukan prediksi selanjutnya secara optimal berdasarkan informasi data di masa lalu maupun berdasarkan informasi data saat ini.

Persamaan umum dari model Kalman Filter sesuai dengan Hamilton (1994) jika diberikan y_t yang dinotasikan sebagai vector berukuran $(n \, x \, 1)$ variabel terobservasi pada waktu t, model dinamis untuk y_t yang mungkin tidak teramati dapat dijelaskan pada vektor yang berukuran $(r \, x \, 1)$. Representasi ruang keadaan untuk y diberikan pada persamaan berikut :
Persamaan keadaan : $\xi_{t+1} = F\xi_t + v_{t+1}$
Persamaan observasi : $y_t = H'^{\xi_t} + w_t$
Dimana
ξ_t : vector keadaan berukuran $(r \times 1)$.
F : matriks parameter berukuran $(r \times r)$ pada persamaan keadaan
H' : matriks parameter berukuran $(n \times r)$ pada persamaan observasi
v : vektor noise berukuran $(r \times 1)$ pada persamaan keadaan
w : vektor noise berukuran $(n \times 1)$ pada persamaan observasi
y : vector dari variable terobservasi berukuran $(n \times 1)$
Q : matriks varian kovarian berukuran $(r \times r)$ pada noise persamaan keadaan
R : matriks varian kovarian berukuran $(n \times n)$ pada noise persamaan observasi
r : orde ARIMA dari data pengamatan
n : jumlah variable bebas pada data pengamatan

Vektor v_t dan w_t harus mengikuti asumsi *white noise* dengan :

$$E(v_t v'_t) = \begin{cases} Q, \text{ untuk } t = \tau \\ 0, \text{ lainnya} \end{cases}$$

Dimana Q merupakan matriks berukuran $(r \times r)$.

$$E(w_t w'_t) = \begin{cases} R, \text{ untuk } t = \tau \\ 0, \text{ lainnya} \end{cases}$$

Dimana R adalah matriks berukuran $(n \times n)$.

Vektor *noise* v_t dan w_t diasumsikan tidak berkorelasi dengan semua lag sehingga untuk semua t dan τ nilai $E(v_t w'_t) = 0$.

Metode

Pada penelitian ini data curah hujan yang digunakan merupakan data curah hujan dari tahun 2003-2013 untuk wilayah Kecamatan Gunungpati Kota Semarang, yang diambil langsung dari Badan Meteorologi Klimatologi dan Geofisika Provinsi Jawa Tengah. Langkah – langkah analisis yang akan dilakukan pada penelitian ini sebagai berikut: Pertama membuat model ARIMA dari data observasi yang telah didapat. Kedua membagi model ARIMA yang akan digunaka untuk metode Bayesian Model Averaging (BMA) dan Kalman Filter. Dalam penelitian ini untuk metode BMA model arima yang digunakan sebanyak empat model yang dipilih berdasarkan syarat memenuhi white noise dan residual normal. Sedangan untuk metode Kalman Filter model arima yang digunakan merupakan model paling terbaik.

Langkah ketiga yakni melakukan proses peramalan dengan BMA , dengan membangkitkan data ensamble tiruan dari keempat model *time series.* Proses ini dilakukan dengan memperbarui model peramalan dari hari ke hari, sehingga bukan peramalan sekaligus dalam banyak periode seperti pada umumnya, melainkan peramalan satu tahap kedepan. Setelah itu menggunakan pendekatan EM , dengan melalui langkah kalibrasi ini akan didapat parameter – parameter kalibrasi yang akan digunakan dalam mendapatkan taksiran interval. Lalu mengevaluasi ketepatan peramalan dengan menggunakan CPRS.

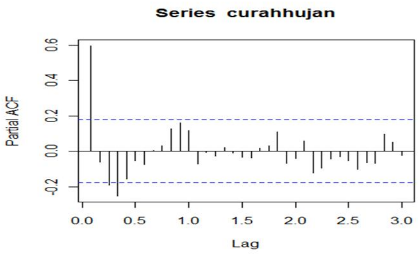
Langkah keempat melakukan proses permalan dengan Kalman Filter yaitu dengan membentuk sebuah fungsi pada ruang keadaan yang mewakili proses ARIMA terbaik dari data. Selanjutnya mengestimasi nilai awal parameter yang belum diketahui dari fungsi yang telah dibentuk. Lalu melakukan pemfilteran dari data curah hujan dengan output parameter dari hasil fungsi yang telah diestimasi. Menghitung nilai residual, membuat plot hasil filtering dan plot normalitas residual. Melakukan prediksi curah hujan dari data yang telah difilter kemudian mencari batas atas dan bawah hasil prediksi.

Langkah kelima membandingan hasil peramalan antara metode BMA dengan Kalman Filter melalui Mean Squared Error (MSE). Maka dapat diperoleh metode peramalan yang lebih akurat.

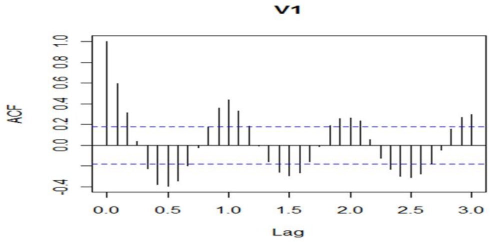
Pembahasan

Pada penelitian ini akan dilakukan peramalan dengan membandingkan data asli dengan data hasil peramalan. Metode peramalan yang akan dibandingkan ialah metode Bayesian Model Averaging dan metode Kalman Filter. Untuk mengolah data tersebut maka progam yang digunakan untuk mempermudah pengerjaan ada dua yakni dengan progam R dan minitab. Data

curah hujan yang digunakan adalah data curah hujan untuk Kecamatan Gunung Pati Kota Semarang periode 2003 – 2013 yang diambil dari Badan Meteorologi Klimatologi dan Geofisika (BMKG) Kota Semarang.



(a)



(b)

Gambar 1. (a) Plot ACF dan (b) Plot PCAF data curah hujan bulanan

Dari Gambar 1 plot ACF dan PCAF dapat dilihat kemungkinan data observasi telah stasioner sehingga tidak perlu untuk ditransormasi ataupun dideferinsikan.

```
Augmented Dickey-Fuller Test
data:  curahhujan
Dickey-Fuller = -6.4912,
Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

berdasarkan uji augmentasi dickey fuller diatas *p-value*< 5%, ini menunjukkan data observasi telah stasioner.

Berdasarkan Tabel 1 dari enam model yang dibuat hanya empat model saja yang ternyata dapat digunakan untuk peramalan dengan metode Bayesian Model Averaging, yakni ARIMA (0,0,1) $(1,0,0)^{12}$, ARIMA (1,0,0) $(0,0,1)^{12}$, ARIMA (1,0,0) $(1,0,0)^{12}$ dan ARIMA (0,0,2) $(1,0,0)^{12}$ karena memenuhi *white noise* dan residual normal.

Sedangkan model ARIMA $(0,0,1)(1,0,0)^{12}$ akan digunakan untuk meramalkan dengan metode Kalman Filter karena pada model tersebut memiliki AIC yang paling besar.

Peramalan dengan metode Bayesian Model Averaging adalah salah satu metode kalibrasi dimana metode ini dilakukan untuk memperbaiki peramalan *ensemble* yang mempunyai sifat *underdispersive* (nilai varian terlalu rendah) ataupun *overdispersive* (nilai varian terlalu tinggi) sehingga membuat hasil peramalan tidak dapat menangkap data observasi.

Peramalan *ensemble* tiruan yang dilakukan dalam penelitian ini dilakukan dengan cara memperbarui data dengan model peramalan dari hari ke hari namun dalam penelitian ini dilakukan dari satu bulan ke bulan lain, jadi tidak langsung untuk beberapa bulan ke depan. Dengan memperbarui dari satu bulan ke bulan lain maka dimulai dari peramalan untuk bulan Januari 2013 diperoleh dengan memodelkan data *in-sample* (Januari 2003 – Desember 2012) dengan model ARIMA $(1,0,0)(1,0,0)^{12}$. Dengan memperbarui data *in-sample* (mengurangi satu data terlama dan menambahkan satu data terbaru) menjadi Februari 2003 – Januari 2013 dan dengan model ARIMA $(1,0,0)(1,0,0)^{12}$, akan diperoleh hasil peramalan untuk Februari 2013. Pembaruan data *in-sample* ini dilakukan terus menerus hingga didapatkan peramalan untuk bulan Desember 2013. Selain itu, langkah ini juga dilakukan terhadap 3 model lainnya (ARIMA $(1,0,0)(0,0,1)^{12}$, ARIMA $(0,0,1)(0,0,1)^{12}$, dan ARIMA$(0,0,2)(1,0,0)^{12}$).

Setelah didapatkan 4 buah peramalan *ensemble* untuk masing-masing model, dapat diperoleh nilai μ dan σ^2 (perbulan) dari keempat model peramalan *ensemble* tiruan tersebut. Berdasarkan penelitian sebelumnya oleh sloughter dkk (2010), data dari masing – masing *ensemble* mengikuti distribusi gamma. Sehingga parameter gamma (α dan β) akan diestimasi dari μ dan σ^2 yang telah ditemukan sebelumnya.

Tabel 1. Evaluasi Model Arima Yang Digunakan

Model	M1	M2	M3	M4	M5	M6
Arima	(0,0,1) $(1,0,0)^{12}$	(1,0,0) $(0,0,1)^{12}$	(1,0,0) $(1,0,0)^{12}$	(1,0,1) $(1,0,0)^{12}$	(0,0,2) $(1,0,0)^{12}$	(2,0,0) $(1,0,0)^{12}$
ϕ_1	-	0,7197	0,6723	0,7471	-	0,5904
P-value	-	0,000	0,000	0,000	-	0,000
ϕ_2	-	-	-	-	-	0,1162
P-value	-	-	-	-	-	0,2111
$\boldsymbol{\theta}_1$	0,4458	-	-	0,1419	0,4983	-
P-value	0,000	-	-	0,2440	0,000	-
$\boldsymbol{\theta}_2$	-	-	-	-	0,2933	-
P-value	-	-	-	-	0,0006	-
ϕ_{12}	0,5889	-	0,3848	0,3940	0,5230	0,3958
P-value	0,000	-	0,0002	0,0001	0,000	0,0001
$\boldsymbol{\theta}_{12}$	-	0,3367	-	-	-	-
P-value	-	0,004	-	-	-	-
AIC	1586,79	1568,25	1566,34	1567,03	1577,57	1566,77
White noise	Memenuhi	Memenuhi	Memenuhi	Tidak memenuhi	Memenuhi	Tidak Memenuhi
Residual Normal	Memenuhi	Memenuhi	Memenuhi	Memenuhi	Memenuhi	Memenuhi

Tabel 2. Hasil Peramalan Dengan *Ensemble* Tiruan

Bulan	Observasi	Lower	Upper
Januari	414	337	370
Februari	222	155	202
Maret	363	189	213
April	221	257	288
Mei	301	143	198
Juni	206	210	263
Juli*	131	114	158
Agustus	22	108	148
September	22	64	72
Oktober	111	116	163
November	346	221	251
Desember*	254	248	297

Berdasarkan Tabel 2 dapat diketahui bahwa hanya dua dari dua belas hasil peramalan *ensemble* tiruan yang dapat menangkap data observasi, hal ini karena nilai varian *ensemble* tiruan yang besar. Sehingga hal ini mengindikasikan bahwa hasil peramalan *ensemble* tiruan umumnya bersifat *overdispersive*. Karena peramalan untuk bulan Januari – Desember 2013 dengan menggunakan *ensemble* tiruan bersifat *overdispersive* dan masih memberikan hasil yang kurang baik, maka hasil peramalan ensembel tiruan perlu dikalibrasi. Metode kalibrasi yang digunakan dalam penelitian ini adalah metode Bayesian Model Averaging dimana parameter – parameternya (varians dan bobot) akan diestimasi dengan pendekatan Expectation Maximization. Selain itu akan digunakan beragam nilai m (*window training*) antara lain 3, 6, dan 9.

Penggunaan nilai m yang berbeda akan menyebabkan perbedaan waktu yang dapat diramalkan. Untuk peramalan menggunakan lead ke-1, waktu pertama yang dapat diramalkan dimulai dari satu waktu setelah *m*. Selain itu, penggunaan nilai *m* yang berbeda juga akan menyebabkan hasil peramalan yang berbeda pula.

Setelah menentukan ukuran m yang digunakan, langkah berikutnya adalah mendapatkan nilai b_0 dan b_1. Pada dasarnya b_0 dan b_1 digunakan untuk menghilangkan bias dari peramalan *ensemble*, dimana proses ini bertujuan untuk menggeser nilai μ peramalan *ensemble* tiruan agar dapat mendekati nilai observasi melalui prosedur regresi. Nilai b_0 dan b_1 dapat diperoleh dengan cara meregresikan peramalan *ensemble* terhadap observasi dengan jumlah data yang digunakan sebanyak *m* data sebelum tanggal peramalan. Sebagai ilustrasi , misalkan akan dikalibrasi peramalan *ensemble* pada bulan Desember 2013 dengan $m = 3$, maka data yang digunakan adalah data mulai September 2013 – November 2013. Melalui proses regresi ini akan diperoleh koefisien regresi $b_0 = -129$ dan $b_1 = 1{,}96$. Koreksi bias didapat melalui rumusan $\mu_k = b_0 + b_1 f_k$, dimana fk merupakan peramalan *ensemble* untuk tiap model pada bulan Desember 2013, maka akan diperoleh $\mu_1 = 212{,}362$; $\mu_2 = 196{,}0322$; $\mu_3 = 194{,}479$; dan $\mu_4 = 212{,}5073$.

Nilai varian dan *weight* diestimasi dengan pendekatan EM. Dari pendekatan EM ini akan diperoleh koefisien c_0 = 1,09 dan c_1= 1,E-05 , dengan standar deviasi $\sigma_k = c_0 + c_1 f_k$. Selanjutnya didapatkan σ_1= 1,094055 ; σ_2=1,093595 ; σ_3=1,093552 ; dan σ_4=1,094059. Serta nilai bobot untuk masing-masing model adalah w_1= 5,16E-04 ; w_2 = 9,93E-04 ; w_3 = 1,63E-03 ; dan w_4 = 6,27E-04. Semakin besar bobot dari model itu semakin tinggi tingkat kontribusi peramalan *ensemble* terhadap hasil peramalan terkalibrasi.

Karena peramalan *ensemble* untuk kecepatan angin mengikuti distribusi gamma, pmaka parameter gamma diestimasi dari μ_k dan σ^2_k yang telah ditentukan sebelumnya. Dan diperoleh α_1= 37676,97019 , β_1= 0,00563; α_2= 32132,31016 , β_2= 0,00610; α_3=31627,8147 , β_3=0,006149 ; dan α_4 = 37728,1867 , β_4= 0,005632.

Tabel 3. Parameter Distribusi untuk Bulan Desember 2013

Bobot (w)	m = 3	m=6	m=9
M1	5,16,E-04	7,72,E-04	6,06,E-04
M2	9,93,E-04	1,24,E-03	1,13,E-03
M3	1,63,E-03	1,02,E-03	8,28,E-04
M5	6,27,E-04	9,00,E-04	1,37,E-03
μ-terkalibrasi	203,8453	220,9399	205,74
σ^2-terkalibrasi	1,196432	2,049806	2,609

Berdasarkan Tabel 3 dapat diketahui bahwa hasil peramalan terkalibrasi terbaik untuk Bulan Desember 2013 adalah menggunakan m=6, dikarenakan hasil peramalan terkalibrasi memiliki bias paling kecil karena nilai meannya mendekati observasi yaitu 220,9399 serta memiliki bobot w yang mendekati nilai bobot terbaik 1 yaitu bernilai 1,02,E-03.

Evaluasi hasil peramalan antara kedua metode yang digunakan ini dilakukan dengan menggunakan Continous Ranked Probability Score (CRPS), dimana prosedur dasarnya adalah membandingkan distribusi hasil peramalan dengan pengamatan, yang keduanya direpresentasikan. Sebagai fungsi distribusi kumulatif (cdf).

Berdasarkan Tabel 4 dapat diketahui hasil peramalan dengan menggunakan BMA-EM akan lebih mendekati nilai observasi daripada hasil peramalan dengan metode *ensemble* tiruan, selain itu hasil interval peramalannya juga padat. Hal ini ditunjukkan oleh nilai CRPS metode kalibrasi BMA-EM yang lebih kecil dari pada *ensemble* tiruan. Sedangkan untuk menggunakan m dapat diketahui bahwa peramalan terbaik menggunakan m=3 dikarenakan akan menghasilkan nilai CRPS yang optimum / terkecil dibandingan dengan nilai m lainnya. Sehingga hasil prediksi untuk curah hujan Kota Semarang pada bulan Desember 2013 sebesar 203 mm.

Tabel 4. Evaluasi Hasil Peramalan

CRPS	m=3	m=6	m=9
Ensemble Tiruan	0,500301	0,500301	0,500301
BMA	0,500506	0,500354	0,500164

Persamaan Kalman Filter dibuat berdasarkan model terbaik ARIMA (0,0,0)(0,1,1)12 yang direpresentasikan dalam bentuk ruang keadaan (*state space*) sehingga didapat persamaan Kalman Filter sebagai berikut

$$\begin{vmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} \begin{vmatrix} \xi_{t+1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{vmatrix}$$

Gambar 2. Persamaan *State space* Kalman Filter

Dan untuk persamaan observasinya seperti berikut:

$$y_t = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1]\xi_t + w_t$$

Hasil peramalan curah hujan Kota Semarang untuk wilayah Kecamatan Gunung Pati tahun 2013 ditunjukkan pada Tabel 5 sebagai berikut

Tabel 5. Hasil Peramalan Kalman Filter Tahun 2013

Bulan	Peramalan Curah Hujan (mm)	Batas Atas	Batas Bawah
Januari	429	274	584
Februari	241	76	407
Maret	254	78	429
April	168	-17	352
Mei	98	-95	291
Juni	16	-186	217
Juli	-3	-213	206
Agustus	-4	-221	212
September	52	-171	276
Oktober	139	-92	369
November	300	63	537
Desember	360	117	703

Perbandingan antara kedua metode peramalan ini bertujuan untuk mengetahui metode mana yang terbaik dalam menyelesaikan kasus penelitian ini. Setelah dilakukan peramalan untuk bulan Januari – Desember 2013 maka kedua metode tersebut akan menggunakan MSE untuk membandingkan metode mana yang merupakan metode terbaik.

Tabel 6. Hasil Perbandiangan Metode BMA-EM Dan Kalman Filter

Bulan	Obs	BMA			Kalman Filter
		m=3	m=6	m=9	
Januari	414	340	337	328	429
Februari	222	374	192	182	241
Maret	363	247	233	207	254
April	221	198	206	221	168
Mei	301	291	297	197	98
Juni	206	247	225	202	16
Juli	131	321	183	151	-3
Agustus	22	167	188	170	-4
September	22	-8	81	104	52
Oktober	111	87	119	143	139
November	346	197	259	263	300
Desember	254	204	221	206	360

Dengan menggunakan MSE didapat hasil peramalan metode BMA untuk m=3 sebesar 10627, m=6 sebesar 5571 , m=9 sebesar 6974 sedangkan untuk metode Kalman Filter MSEnya sebesar 10521.

Simpulan

Dari pembahasan dapat disimpulkan bahwa hasil peramalan dengan menggunakan metode Bayesian Model Averaging sangat bervariasi karena tergantung dari banyaknya model arima yang digunakan serta penentuan *training set* (m) yang akan digunakan. Sedangkan Peramalan dengan metode Kalman Filter menunjukan bahwa nilai observasi dapat masuk dalam rentang selang hasil peramalannya karena memiliki interval yang cukup besar.

Setelah diuji dengan MSE ternyata peramalan dengan menggunakan metode Bayesian Model Averaging lebih akurat daripada menggunakan metode Kalman Filter karena metode Bayesian Model Averaging memiliki error lebih sedikit dibandingkan dengan metode Kalman Filter, namun peramalan dengan Kalman Filter mempunyai keunggulan dapat menangkap lebih banyak data observasi karena memiliki interval yang lebih panjang daripada metode Bayesian Model Averaging.

DAFTAR PUSTAKA

Box,George E.P., Jenkins,Gwilym M., Reinsel, Gregory C. 2008. *Time Series Analysis: Forecasting and Control, 4th Edition.* NJ.USA.

Brockwell,P.J. and Davis, R.A. 1991. *Time Series :Theory and Methods*. Second Edition. Springer-Verlag, Inc. New York.

Erna, A., Sanjoyo, A., dan Dieky, A. 2011. „"The Ground Water Pollution Estimation by The *Ensemble* Kalman Filter. *Canadian Journal on Science and Engineering Mathematics*, Vol 2 No. 2.

LAPAN.(2009).*DampakPerubahan Iklim*.(http://iklim.dirgantara-lapan.or.id.id/index.php?option=com_content&view=article&id=60&Itemid=37, diaksespada 9 September 2014).

Hamilton, J.D. 1994. *Time Series Analysis*. Princeton University Press. New Jersey.

Makridakis,Spyrosdkk. 1999. *Metode dan Aplikasi Peramalan* edisi kedua. Jakarta: Erlangga

McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models*. New York: Wiley.

Meinhold, R.J. and Singpurwala, N. D. 1983. Understanding The Kalman Filter. *The American Statistician*. Volume 37 No. 2 :

123 – 127. American Statistical Association.

Petris, G. 2010. An R Package for Dynamic Linear Models. *Journal of Statistical Software*. Volume 36 Issue 12. American Statistical Association. USA.

Petris,G. and Petrone, S. 2011. State Space Models in R. *Journal of Statistical Software*. Volume 41 Issue 4. American Statistical Association. USA.

Raftery,Adrian E.2003. Using Bayesian Model Averaging to Calibrate Forecast *Ensembles*. *Jurnal of Statistic Department*. University of Washington, Seattle, Washington

Raftery, A.F., Gneiting, T., Balabdaoui, F., dan Polakowski, M. (2005). Using Bayesian Model to Calibrate Forecast *Ensembles*. *Monthly Weather Review*, Vol 133. Hal 1155-1173.

Sloughter, M.J., Gneiting, T., dan Raftery, A.E. (2010). Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model Averaging. *Journal of the American Statistical Association*. Vol 105, No 489. Hal 25-34.

Supranto, J.2000. *Statistika Teori dan Aplikasi*. Jakarta : Erlangga.

Tresnawati, R. dan Komalasari, K.E. 2011. Skenario Tenggang Waktu SST Nino 3.4 Terhadap Curah Hujan untuk Meningkatkan Akurasi Prediksi Kalman Filter. *Jurnal Meteorologi dan Geofisika*. Volume 12 No. 3 : 243 – 251. Puslitbang BMKG. Jakarta.

Wei, W.W.S. 2006. *Time Series Analysis Univariate and Multivariate Methods*. Second Edition. Pearson Education, Inc. US.

Welch, G. and Bishop, G. 2001. *An Introduction to the Kalman Filter*. ACM Inc.

Wu, C. F. J. (1983). *On The Convergence Properties of The EM Algorithm*. Annals of Statistic 11, 95-103.

Zhu, Yuejian.2005. *Ensemble* Forecast: A New Approach To Uncertainty And Predictability. *Advance In Atmosphere Science,Vol 22,No 6.* Hal 781-788