

**PERBAIKAN MODEL *SEASONAL* ARIMA  
DENGAN METODE *ENSEMBLE* KALMAN FILTER  
PADA HASIL PREDIKSI CURAH HUJAN**  
*(Improvement Seasonal ARIMA Model Using Ensemble Kalman Filter Methods  
for Rainfall Prediction Results)*

**Dwi Anugrah Wibisono, Dian Anggraeni, Alfian Futuhul Hadi**

Jurusan Matematika, Fakultas MIPA, Universitas Jember  
Jl. Kalimantan 37 Jember 68121, Indonesia  
E-mail: danugrah1995@gmail.com, {dian\_a, afhadi}.fmipa@unej.ac.id

**Abstract.** Forecasting is a time series analytic that used to find out upcoming improvement in the next event using past events as a reference. One of the forecasting models that can be used to predict a time series is Kalman Filter method. The modification of the estimation method of Kalman Filter is Ensemble Kalman Filter (EnKF). This research aims to find the result of EnKF algorithm implementation on SARIMA model. To start with, precipitation forecast data is changed in the form of SARIMA model to obtain some SARIMA model candidates. Next, this best model of SARIMA applied to Kalman Filter models. After Kalman Filter models created, forecasting could be done by applying pass rainfall data to the models. It can be used to predict rainfall intensity for next year. The quality of this forecasting can be assessed by looking at MAPE's value and RMSE's value. This research shows that enkf method relative can fix sarima method's model, proved by mape and rmse values which are smaller and indicate a more accurate prediction.

**Keywords:** Ensemble Kalman Filter, Forecast, SARIMA

**MSC 2010:** 62M10

## 1. Pendahuluan

Peramalan (*forecasting*) adalah analisis data time series yang menggunakan kejadian di masa lalu untuk mengetahui perkembangan kejadian di masa yang akan datang [1]. Peramalan merupakan suatu teknik untuk memperkirakan suatu nilai pada masa yang akan datang dengan memperhatikan data masa lalu maupun data masa saat ini. Peranan peramalan menjelajah ke dalam banyak bidang seperti ekonomi, keuangan, pemasaran, produksi, riset operasional, administrasi negara, meteorologi, geofisika, kependudukan, dan pendidikan [2]. Badan Meteorologi dan Geofisika (BMKG) adalah salah satu instansi yang menerapkan ilmu peramalan dalam meramalkan jumlah curah hujan di suatu daerah. Curah hujan adalah jumlah air hujan yang turun pada suatu daerah dalam waktu tertentu. Letak topografi suatu daerah mempengaruhi curah hujan yang akan terjadi. Selain itu, informasi curah hujan sangat bermanfaat, misalny dalam bidang perhubungan. Informasi mengenai curah hujan, baik curah hujan tinggi maupun rendah sangat membantu seorang

pilot untuk mengambil tindakan dalam proses penerbangan, baik ketika pesawat sedang mendarat maupun pesawat yang akan lepas landas. Informasi yang berkaitan dengan kondisi di masa yang akan datang tidak dapat ditentukan secara pasti, tetapi bisa diprediksi. Dalam memprediksi suatu kejadian, terdapat metode yang mampu memprediksi kejadian di masa mendatang, yaitu dengan menggunakan analisis deret waktu. Analisis deret waktu digunakan untuk melakukan analisis data yang mempertimbangkan pengaruh waktu. Data yang digunakan secara periodik berdasarkan urutan waktu, bisa dalam jam, hari, minggu, bulan dan tahun [3]. Salah satu dari model peramalan yang dapat digunakan untuk melakukan peramalan suatu deret waktu adalah dengan model *Kalman Filter*. *Kalman Filter* adalah sebuah model bagian dari *state space* (ruang keadaan) yang dapat diterapkan dalam model peramalan. Model ini menggunakan teknik rekursif dalam mengintegrasikan data pengamatan terbaru ke model untuk mengoreksi prediksi sebelumnya dan melakukan prediksi selanjutnya secara optimal berdasarkan informasi di masa lalu maupun berdasarkan informasi data saat ini [4]. Metode estimasi modifikasi dari algoritma *Kalman Filter* yang dapat digunakan untuk mengestimasi model sistem linear maupun nonlinear adalah *Ensemble Kalman Filter* (EnKF).

Penelitian sebelumnya yang berkaitan dengan peramalan curah hujan dilakukan oleh Lusiani [5] pemodelan ARIMA curah hujan di kota Bandung, Ukhra [6] pemodelan dan peramalan data deret waktu dengan metode SARIMA, Retnaningrum [7] penerapan model STAR (*Space Time Autoregressive*) dan ARIMA untuk peramalan data curah hujan di kabupaten Jember dan Nikmatillah [8] *prediction interval* pada model SARIMA untuk peramalan curah hujan dan kekeringan. Sedangkan untuk penelitian sebelumnya yang berkaitan dengan *Kalman filter* dilakukan oleh Pamungkas [9] estimasi parameter model ARIMA menggunakan *Kalman filter* untuk peramalan permintaan darah dan Nugrahini [10] perbandingan metode *Kalman Filter* dan metode *Ensemble Kalman Filter* dalam mendeteksi gangguan konduksi panas pada batang logam.

Penelitian ini bertujuan untuk melihat hasil prediksi curah hujan dengan menerapkan algoritma *Ensemble Kalman Filter* pada model SARIMA. Sebelumnya data curah hujan dibentuk ke dalam model SARIMA, setelahnya akan didapatkan beberapa kandidat model SARIMA. Beberapa model kandidat SARIMA tersebut selanjutnya menerapkan metode *Ensemble Kalman Filter* terhadap kandidat model-model SARIMA. Setelah model *Kalman Filter* terbentuk, dilakukan peramalan pada data curah hujan perbulan di Kabupaten Jember untuk 1 tahun ke depan yaitu tahun 2017, untuk melihat bagaimana hasil peramalan digunakan MAPE (*Mean Absolute Percentage Error*) dan RMSE (*Root Mean Square Error*). Data penelitian ini menggunakan data curah hujan yang didapat dari 77 stasiun hujan di Kabupaten Jember.

## 2. Metodologi

Data yang digunakan dalam penelitian ini merupakan data sekunder dan kuantitatif dengan jenis *time series* yang diambil dari penelitian Retnaningrum [7], penelitian Yudistira [11], penelitian Nikmatillah [8] dan Dinas PU Bina Marga dan Sumber Daya Air. Data yang digunakan adalah data curah hujan (mm) di kabupaten Jember dari bulan Januari 2005 sampai dengan bulan Desember 2017. Data curah hujan ini dibagi menjadi dua kelompok data yaitu data training dan data *testing*. Data training digunakan untuk membentuk model sedangkan data *testing* digunakan untuk memeriksa daya ramal yang terbentuk dari data training. Data training meliputi data bulan Januari 2005 sampai dengan data bulan Desember 2016. Data testing meliputi data bulan Januari 2017 sampai dengan bulan Desember 2017. Pada kasus ini variabel-variabel yang digunakan adalah:

$X_1(t)$  : Rata-rata curah hujan di 12 wilayah Jember zona 1

$X_2(t)$ : Rata-rata curah hujan di 20 wilayah Jember zona 2

$X_3(t)$  : Rata-rata curah hujan di 14 wilayah Jember zona 3

$X_4(t)$  : Rata-rata curah hujan di 31 wilayah Jember zona 4

Langkah awal yang dilakukan adalah mencari studi literatur. Studi literatur dilakukan untuk mendapatkan informasi dari buku, jurnal dan skripsi terkait. Kemudian data terkait SARIMA (curah hujan, lokasi) akan diteliti dikumpulkan untuk kemudian diolah dengan menggunakan program R dan MATLAB. Pengolahan data berupa mengubah model SARIMA ke dalam bentuk state space, membangkitkan sejumlah  $N_\varepsilon$  ensemble (100, 200, 500) dengan mean  $\bar{x}_0$  dan kovariansi yang bersesuaian dengan waktu, dan tentukan nilai awal :

$$\hat{x}_0 = \frac{1}{N_\varepsilon} \sum_{i=1}^{N_\varepsilon} x_{0,i}$$

Tahap Prediksi :

$$\hat{x}_{k,i}^- = f(\hat{x}_{k-1}, u_{k-1}) + w_k \text{ dengan } w_{k,i} \sim N(0, Q_k)$$

$$\text{estimasi : } \hat{x}_k^- = \frac{1}{N_\varepsilon} \sum_{i=1}^{N_\varepsilon} \hat{x}_{k,i}^-$$

$$\text{kovarian error : } P_k^- = \frac{1}{N_\varepsilon - 1} \sum_{i=1}^{N_\varepsilon} (\hat{x}_{k,1}^- - \hat{x}_k^-) (\hat{x}_{k,1}^- - \hat{x}_k^-)^T$$

Tahap koreksi :

$$z_{k,i} = z_k + v_{k,i} \text{ dengan } v_{k,i} \sim N(0, R_k)$$

$$\text{kalman gain : } K_k = P_k^- H^T (H P_k^- H^T + R_k)^{-1}$$

$$\text{estimasi : } \hat{x}_{k,i} = \hat{x}_{k,i}^- + K_k (z_{k,i} - H \hat{x}_{k,i}^-)$$

$$\hat{x}_k = \frac{1}{N_\varepsilon} \sum_{i=1}^{N_\varepsilon} \hat{x}_{k,i}$$

kovarian *error* :  $P_k = (I - K_k H)P_k^-$

dengan

$x_k$ : variabel keadaan sistem pada waktu  $k$  yang nilai estimasi awalnya adalah  $\hat{x}_0$  dan kovarian awal  $P_{x_0}$

$u_k$ : variabel input deterministik pada waktu  $k$

$w_k$ : noise pada model sistem

$z_k$ : variabel pengukuran

$H$ : matriks pengukuran

$v_k$ : *noise* pada model pengukuran

$A_k, B_k, G_k$ : matriks-matriks koefisien di dalam ukuran yang bersesuaian dengan

$A = n \times n, B = m \times m$ , dan  $H = p \times 1$

Nilai MAPE dan RMSE digunakan untuk hasil peramalan yang nantinya digunakan untuk melihat seberapa akurat hasil peramalan yang kita lakukan.

RMSE digunakan untuk mencari keakuratan hasil peramalan dengan data history dengan menggunakan rumus sebagai berikut :

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (Z_t - \hat{Z}_t)^2}{N}} \quad (1)$$

Dengan  $T$  merupakan banyaknya pengamatan,  $Z_t$  adalah nilai pengamatan pada waktu ke- $t$ , dan  $\hat{Z}_t$  adalah nilai ramalan pada waktu ke- $t$ . Semakin kecil nilai RMSE yang dihasilkan semakin bagus pula hasil peramalan yang dilakukan [2].

Sama seperti halnya RMSE, MAPE menghitung kualitas dari data observasi dan prediksi yang dimunculkan dalam nilai persen. Nilai MAPE dapat dihitung dengan persamaan berikut :

$$MAPE = \left(\frac{100\%}{n}\right) \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2)$$

dengan

$A_t$  : Nilai data asli

$F_t$  : Nilai data peramalan

### 3. Hasil dan Pembahasan

Berdasarkan 4 (empat) pengelompokan wilayah Kabupaten Jember dan data yang diperoleh, Nikmatillah [8] mengungkapkan bahwa beberapa model kandidat *Seasonal* ARIMA yang didapatkan terdapat pada Tabel 1.

Tabel 1. Kandidat model SARIMA

	Model 1	Model 2	Model 3	Model 4
Z1	SARIMA (2,0,2)(1,0,0) <sup>12</sup>	SARIMA (2,0,1)(1,0,0) <sup>12</sup>	SARIMA (2,0,2)(1,0,1) <sup>12</sup>	SARIMA (1,0,0)(1,0,0) <sup>12</sup>
Z2	SARIMA (1,0,0)(2,0,0) <sup>12</sup>	SARIMA (1,0,0)(1,0,0) <sup>12</sup>	SARIMA (2,0,0)(2,0,0) <sup>12</sup>	SARIMA (1,0,1)(2,0,0) <sup>12</sup>
Z3	SARIMA (1,0,0)(2,0,0) <sup>12</sup>	SARIMA (1,0,0)(1,0,0) <sup>12</sup>	SARIMA (2,0,0)(2,0,0) <sup>12</sup>	SARIMA (1,0,1)(2,0,0) <sup>12</sup>
Z4	SARIMA (1,0,0)(2,0,0) <sup>12</sup>	SARIMA (1,0,0)(1,0,0) <sup>12</sup>	SARIMA (1,0,1)(2,0,0) <sup>12</sup>	SARIMA (2,0,1)(2,0,0) <sup>12</sup>

Tabel 1 menunjukkan beberapa kandidat model yang cocok untuk tiap zona. Menggunakan fungsi `auto.arima` pada R tidak hanya mendapatkan model SARIMA terbaik menurut AIC, namun bisa juga digunakan untuk mencari kandidat model lain di tiap zona. Model-model di atas dapat ditulis secara sistematis dan diterapkan dengan metode EnKF yang disimulasikan menggunakan 200 *ensemble* sebagai berikut.

Model sistem dan pengukuran

$$x_{k+1} = f(x_k, u_k) + w_k$$

dengan  $f(x_k, u_k) = A_k x_k + E_k d_k$

$$\begin{bmatrix} \emptyset_1 \\ \cdot \\ \cdot \\ \cdot \\ \emptyset_p \\ \mathbf{Y}_t \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & \vdots \\ \vdots & 0 & 1 & 0 & \dots & \vdots \\ \vdots & \dots & 0 & 1 & 0 & \vdots \\ 0 & 0 & 0 & 0 & 1 & \vdots \\ \mathbf{U}_1 & \dots & \dots & \dots & \mathbf{U}_p & 0 \end{bmatrix} \begin{bmatrix} \emptyset_1 \\ \cdot \\ \cdot \\ \cdot \\ \emptyset_p \\ \mathbf{Y}_t \end{bmatrix}_k + \begin{bmatrix} 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \dots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \varepsilon_t \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix}_k + w_k$$

$$z_k = H_k x_k + v_k = [0 \quad \dots \quad \dots \quad 0 \quad 1] \begin{bmatrix} \emptyset_1 \\ \cdot \\ \cdot \\ \cdot \\ \emptyset_p \\ \mathbf{Y}_t \end{bmatrix}_k + v_k$$

Dengan  $w_k$  adalah noise sistem yang timbul akibat proses pembentukan model, sedangkan  $v_k$  adalah noise hasil pengukuran. Setelah diperoleh model sistem dan pengukuran, selanjutnya dilakukan inisiasi. Untuk nilai awal  $Y_t$  diambil dari mean data. Untuk penambahan nilai *noise* model sistem  $w_t$  dibangkitkan dengan sejumlah bilangan acak komputer menggunakan program Matlab. Nilai awal variansi dari noise  $Q = 10^{-6}$  dan  $R = 10^{-2}$ .

Sedangkan nilai awal  $\bar{x}_0$  diberikan sebagai berikut  $\bar{x}_0 = \begin{bmatrix} \emptyset_1 \\ \cdot \\ \cdot \\ \cdot \\ \emptyset_p \\ Y_t \end{bmatrix}$

Simulasi untuk mendapatkan model pada penelitian kali ini dilakukan dengan menerapkan metode EnKF terlebih dahulu dan mencoba 200  $N_e$  ensemble. Penerapan metode EnKF pada model SARIMA diharapkan mampu untuk memperbaiki hasil prediksi curah hujan yang terjadi di kabupaten Jember. Setelah simulasi dilakukan, penulis mendapatkan model baru hasil penerapan metode EnKF terhadap model SARIMA dan melakukan peramalan untuk tahun 2017 dimana data hasil peramalan nantinya akan dievaluasi apakah hasil model EnKF SARIMA lebih baik dari pada model SARIMA tanpa penerapan metode EnKF.

Dalam hal ini suatu peramalan tidak lepas dari suatu kesalahan perhitungan ramalan. Ada beberapa metode untuk mengevaluasi hasil peramalan yang nantinya digunakan untuk melihat seberapa akurat hasil peramalan yang kita lakukan, salah satunya adalah *Mean Absolute Percentage Error* (MAPE) dan *Root Mean Square Error* (RMSE). Semakin kecil nilai RMSE dan MAPE yang dihasilkan semakin bagus pula hasil peramalan yang dilakukan.

Tabel 2. Nilai MAPE dan RMSE dari model terbaik

			<b>SARIMA</b>	<b>EnKF-SARIMA</b>
Zona 1	SARIMA (1,0,0)(1,0,0) <sup>12</sup>	MAPE	444 %	80 %
		RMSE	89,4	86,26
Zona 2	SARIMA (1,0,0)(1,0,0) <sup>12</sup>	MAPE	659 %	127 %
		RMSE	132,21	125,65
Zona 3	SARIMA (1,0,0)(1,0,0) <sup>12</sup>	MAPE	1286 %	125 %
		RMSE	100,31	78,64
Zona 4	SARIMA (1,0,0)(2,0,0) <sup>12</sup>	MAPE	503 %	132 %
		RMSE	81,23	79,34

Pada fungsi SARIMA, selain model terbaik yang dilihat dari nilai AIC terdapat kandidat model-model lain yang dapat digunakan untuk meramalkan nilai-nilai curah hujan bulan selanjutnya. Tabel 2 menunjukkan nilai MAPE dan RMSE dari model terbaik dari keempat zona yang berbeda. Nilai MAPE dan RMSE dari zona 1 yang terbaik terdapat pada model 4. Baik tidaknya model EnKF di tiap zona dapat dilihat dari nilai MAPE dan RMSE yang lebih kecil dibandingkan nilai MAPE dan RMSE pada pasangan model SARIMANYA. Semakin kecil nilai MAPE dan RMSE, menunjukkan makin akuratnya hasil peramalan yang dilakukan.

Pada zona 2 dan zona 3 model terbaik dilihat dari nilai MAPE dan RMSE adalah model 2, dimana model 2 dari zona 2 adalah SARIMA (1,0,0)(1,0,0)<sup>12</sup>. Model terbaik dari nilai MAPE dan RMSE pada zona 3 adalah SARIMA (1,0,0)(1,0,0)<sup>12</sup>. Zona 4 model 1 yang mana adalah SARIMA (1,0,0)(2,0,0)<sup>12</sup> memiliki nilai MAPE dan RMSE yang konsisten lebih kecil jika dibandingkan dengan nilai MAPE dan RMSE model SARIMAny.

Tabel 2 merupakan hasil perbandingan model EnKF SARIMA dengan model SARIMA. Dimana model EnKF SARIMA signifikan memperbaiki hasil peramalan curah hujan model SARIMA dilihat dari nilai MAPE dan RMSE yang relatif lebih kecil atau lebih akurat. Hasil penelitian ini dapat disimpulkan bahwa penerapan metode EnKF pada model SARIMA untuk memprediksi curah hujan tahun 2017 kabupaten Jember, mampu untuk memperbaiki hasil peramalan yang dilakukan.

#### 4. Kesimpulan

Berdasarkan hasil analisis data dan pembahasan diperoleh kesimpulan bahwa hasil prediksi dengan menggunakan metode EnKF SARIMA lebih baik atau akurat dari pada metode SARIMA, ditandai dengan nilai MAPE dan RMSE pada EnKF SARIMA yang lebih kecil. Untuk pengembangan penelitian selanjutnya, *Ensemble Kalman Filter* dapat diterapkan untuk mengestimasi parameter model peramalan time series lainnya, seperti model GSTAR dan sebagainya.

#### Daftar Pustaka

- [1] Assauri, S. (1984), *Teknik dan Metode Peramalan*. Jakarta: Penerbit Fakultas Ekonomi Universitas Indonesia.
- [2] Aswi dan Sukarna. (2006). *Analisis Deret Waktu*. Makassar: Andira Publisher.
- [3] Lusiani, A. 2011. Pemodelan *autoregressive integrated moving average* (arima) curah hujan di kota Bandung. *Sigma-Mu*. 3(2).
- [4] Makridakis, S., Wheelwright, S.C., dan McGee, V.E. (1999). *Forecasting: Methods and applications*. Second Edition. New York: Wiley. Terjemahan oleh U.S. Ardiyanto dan A. Basith. *Metode dan Aplikasi Peramalan Jilid 1*. Edisi Kedua. Jakarta: Penerbit Erlangga.
- [5] Nikmatullah, V. (2018). Prediction Interval pada Model *Seasonal Autoregressive Integrated Moving Average* (SARIMA) untuk Peramalan Curah Hujan dan Kekeringan. *Skripsi*. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember

- [6] Nugrahini, T. (2012). Perbandingan Metode *Kalman Filter* dan Metode *Ensemble Kalman Filter* dalam Mendeteksi Gangguan Konduksi Panas pada Batang Logam. *Skripsi*. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- [7] Pamungkas, H. (2016). Estimasi Parameter Model ARIMA Menggunakan *Kalman Filter* untuk Peramalan Permintaan Darah. *Skripsi*. Surabaya: Institut Sepuluh November.
- [8] Retnaningrum. (2015). Penerapan model STAR dan ARIMA untuk peramalan Curah Hujan di kabupaten Jember. *Skripsi*. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- [9] Ukhra, A. (2014). Pemodelan dan peramalan data deret waktu dengan metode seasonal ARIMA. *Jurnal Matematika UNAND*. 3(3):59 – 67.
- [10] Wei, W.W.S. (2006). *Time Series Analysis : Univariate and Multivariate Methods*. Edisi Kedua. New York :Pearson Education, Inc. Jember : Jember University Press.
- [11] Yudistira, I. (2017). Pengelompokan Stasiun Hujan melalui Variabel Geografis pada Pemodelan GSTAR Musiman untuk Peramalan Curah Hujan di Kabupaten Jember. *Skripsi*. Jember: Universitas Jember.