BAB II

TINJAUAN PUSTAKA

2.1 Curah Hujan

Curah hujan merupakan ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap, dan tidak mengalir. Salah satu batasan yang melandasi pengertian curah hujan yaitu tinggi air hujan (dalam mm) yang diterima permukaan sebelum mengalami aliran permukaan, evaporasi dan peresapan atau perembesan ke dalam tanah (Handoko el al., 1988). Oleh Sosrodarsono dan Takeda (1987), intensitas curah hujan atau disebut juga derajat curah hujan adalah jumlah curah hujan dalam suatu satuan waktu (biasanya dalam mm/jam), dan dapat dibaca dari kemiringan (tangens kurva) yang dicatat oleh alat ukur curah hujan otomatis.

Berdasarkan inforrnasi yang diperoleh dari Badan Meteorologi dan Geofisika (BMG), tinggi curah hujan 1 mm sama dengan jumlah air hujan sebanyak 1 liter dalam luasan 1 meter persegi (1 mm = 1 liter/m2). Keadaan curah hujan dikatakan musim kering jika curah hujan kurang dari 50 mm/10 hari (< 50 mm/10 hari) dan musim hujan jika curah hujan mencapai lebih dari atau sarna dengan 50 mm/10 hari (≥ 50 mm/10 hari). Kriteria curah hujan dalam bulanan terbagi menjadi 3 kategori :

- rendah : 0 - 100 mm
- menengah : 100 - 300 mm
- tinggi : 300 - 500 mm

2.2 Deret Waktu (*Time Series*)

Deret waktu (*time series*) adalah rangkaian nilai pengamatan yang diamati secara berurutan selama kurun waktu tertentu, pada umumnya dalam interval-interval yang sama panjang. Secara matematis, deret waktu didefinisikan oleh nilai-nilai $Z_1, Z_2, \ldots, Z_n$ dari suatu variabel Z untuk titik-titik waktu $t_1, t_2, \ldots$ (Wei, 2006). Secara umum tahapan pemodelan data runtun waktu adalah Identifikasi model, Estimasi Parameter, Verifikasi Model, Peramalan (Aswi,2006).

Hal yang perlu diperhatikan pada peramalan data *time series* adalah galat (*error*). Hasil dari prediksi biasanya berbeda dengan data sesungguhnya, maka hal yang bisa dilakukan adalah meminimalkan galatnya. Untuk meramalkan data *time series* dibutuhkan teknik peramalan yang baik. Teknik peramalan dapat bermacam-macam tergantung pada pola data yang ada Menurut Makridakris *et al* (1992) ada empat macam tipe pola data yaitu:

a. Pola Data Horizontal

Pola data horizontal terjadi saat data observasi berfluktuasi di sekitaran suatu nilai konstan atau *mean* yang membentuk garis horizontal. Data ini disebut juga dengan data stasioner. Contoh suatu produk penjualannya tidak meningkat atau menurun selama waktu tertentu.

b. Pola Data Musiman

Pola data musiman terjadi bilamana suatu data dipengaruhi oleh faktor musiman. Pola data musiman dapat mempunyai pola musim yang berulang dari periode ke periode berikutnya. Misalnya pola yang berulang setiap bulan tertentu, tahun tertentu atau pada minggu tertentu. Contoh suplai bahan makanan tiap bulan.

c. Pola Data *Trend*

Pola data *trend* terjadi bilamana data pengamatan mengalami kenaikan atau penurunan selama periode jangka panjang. Suatu data pengamatan yang mempunyai *trend* disebut data nonstasioner.

d. Pola Data Siklis

Pola data siklis terjadi bilamana data dipengaruhi oleh fluktuasi ekonomi jangka panjang seperti yang berhubungan dengan siklus bisnis.

2.3 Koefesien Korelasi

Korelasi merupakan salah satu teknik analisis dalam statistik yang digunakan untuk mencari hubungan antara dua variabel yang bersifat kuantitatif. Hubungan dua variabel tersebut dapat terjadi karena adanya hubungan sebab akibat atau dapat pula terjadi karena kebetulan saja. Dua variabel dikatakan berkorelasi apabila perubahan pada variabel yang satu akan diikuti perubahan pada variabel yang lain secara teratur dengan arah yang sama (positif) atau berlawanan (negatif). Koefisien korelasi berkisar antara 1 sampai dengan -1. Nilai yang semakin mendekati 1 atau -1 berarti hubungan antara dua variabel semakin kuat. Sebaliknya, jika nilai mendekati 0 berarti hubungan antara dua variabel semakin lemah. Untuk mencari koefisien korelasi dapat digunakan rumus sebagai berikut :

$$r_{ij} = \frac{n(\Sigma z_i z_j) - (\Sigma z_i)(\Sigma z_j)}{\sqrt{\left[n((\Sigma z_i) - (\Sigma z_i)^2)\right]\left[n((\Sigma z_i) - (\Sigma z_j)^2)\right]}}$$

Dengan :

r_{ij} : korelasi sampel antara variabel i dan j .

n : ukuran sampel.

z_i : variabel pada lokasi ke-i.

z_j : variabel pada lokasi ke-j.

Koefisien korelasi digunakan untuk mengukur kekuatan dan arah hubungan linier di antara dua variabel saja. Untuk menguji signifikansi dari koefisien korelasi sederhana dilakukan uji-t dengan uji sebagai berikut:

Hipotesis :

H_0 : tidak ada hubungan secara signifikan antar variabel

H_1 : terdapat hubungan secara signifikan antar variabel

Statistik Uji :

$$t_{hitung} = \frac{r_{ij}\sqrt{n-2}}{\sqrt{1-r_{ij}^2}}$$

Kriteria pengujian :

Dengan α = 5% = 0,05, jika $|t_{hitung}| > t_{\frac{\alpha}{2},n}$, maka H0 ditolak artinya terdapat hubungan secara signifikan antara variabel ke-I dan variabel ke-j. Untuk $t_{\frac{\alpha}{2},n}$ diperoleh dari tabel distribusi t.

2.4 Uji Heterogenitas Lokasi

Menurut Karlina, et al., (2014) metode indeks Gini merupakan rasioanalisis yang sangat merepresentatif data dalam masyarakat yang heterogen. Metode indeks Gini biasanya digunakan untuk mengetahui tingkat pemerataan pendapatan masyarakat dengan melihat nilai indeks Gini yang dibagi menjadi beberapa kriteria diantaranya Gn = 0 artinya pemerataan sempurna dan Gn = 1 artinya pemerataan tidak sempurna. Indeks Gini merupakan suatu metode yang dapat digunakan untuk membandingkan dari satu waktu ke waktu atau dari satu lokasi ke lokasi yang lain. Berikut uji hipotesis heterogenitas.

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$ (terdapat homogenitas spasial)

$H_1 : \sigma_1^2 \neq \sigma_1^2$ (paling tidak terdapat satu heterogenitas spasial)

Kriteria Uji :

Ho ditolak jika nilai G lebih dari sama dengan 1.

Statistik uji :

$$G = 1 + \frac{1}{n} - \frac{2}{n^2 \overline{y}_i} \times \sum_{i=1}^{n_i} y_i$$

dengan

G : indeks Gini,

n : jumlah data,

n_i : jumlah data pada lokasi ke-i,

$\overline{y}_i$: merupakan rata-rata dari masing-masing variabel yang diamati.

2.5 Kestasioneran terhadap Mean dan Varian

Suatu data dikatakan stasioner pada data time series jika nilai mean dan varian konstan atau tidak mengalami perubahan yang sistematik. Makridakis et al. (1992) menyatakan bentuk visual dari suatu plot data time series seringkali cukup untuk meyakinkan bahwa data tersebut adalah stasioner atau tidak stasioner.

2.5.1 Uji *Augmented Dickey Fuller* (ADF)

Uji ADF merupakan pengujian stasioner dengan menentukan apakah data runtun waktu *(time series)* mengandung akar unit (*unit root*). Uji ADF diperkenalkan oleh David Dickey dan Wayne Fuller dengan model sederhana yang digunakan adalah $\Delta Y_t = \beta_1 + \delta Y_{t-1} + e_t$ dengan $\delta = \rho - 1$. Hipotesis yang digunakan dalam persamaan tersebut adalah

H_0: $\delta = 1$ (tidak stasioner)

H_1: $\delta < 1$ (stasioner)

Dapat dilakukan dengan uji statistik τ yaitu $\tau = \frac{P}{SE(p)}$. Jika statistik τ lebih besar dari nilai kritis ADF maka terima H_0 dan disimpulkan Y_t mempunyai akar unit (tidak stasioner), dan apabila statistik τ kurang dari nilai kritis ADF dengan taraf nyata tertentu atau *p-value* < 5% maka tolak H_0 dan disimpulkan Y_t tidak mempunyai akar unit atau stasioner.

2.5.2 Transformasi *Box Cox*

Transformasi *Box Cox* adalah transformasi pangkat pada respon. *Box Cox* mempertimbangkan kelas transformasi berparameter tunggal yaitu λ yang dipangkatkan pada variabel respon Y, sehingga transformasinya menjadi Y^λ. λ adalah parameter yang perlu diduga. Tabel dibawah adalah beberapa nilai λ dengan transformasinya (Ispriyanti, 2004).

Tabel 2. 1 Tranformasi

λ	**Transformasi**
2	Y^2
0.5	$\sqrt{Y}$
0	log Y/ln Y
-0.5	$1/\sqrt{Y}$
-1.0	$1/Y$

Cryer (2008) mendefinisikan transformasi *box cox* sebagai berikut :

$$\mathrm{T}(Y_t) = \begin{cases} Y_t = \frac{Y_t{}^{\lambda}-1}{\lambda} & ; untuk\ \lambda \neq 0 \\ \log(Yt) & ; untuk\ \lambda = 0 \end{cases}$$

Diperlukan pemenuhan asumsi tentang kestasioneran data pada pemodelan deret waktu yang berarti bahwa tidak terdapat pertumbuhan atau penurunan data. Bila tidak terdapat perubahan pada *trend* waktu maka dapat disebut stasioner dimana rata-rata deret pengamatan di sepanjang waktu selalu konstan. Jika suatu deret data bukan merupakan data yang stasioner maka sebelum melakukan pembuatan model deret waktu perlu dilakukan pembedaan atau transformasi (Cowpertwait, 2008).

2.6 Model *Space Time*

Tidak jarang kita menemukan data yang tidak hanya bergantung pada nilai pengamatan pada waktu sebelumnya, tetapi juga dipengaruhi oleh pengamatan di lokasi lain. Data seperti itu disebut dengan data deret waktu dan spasial sehingga untuk model yang sesuai adalah model dengan unsur ruang dan waktu. Model *space time* adalah salah satu model yang menggabungkan unsur dependensi waktu dan lokasi pada data deret waktu. Model ini merupakan pemodelan dari sejumlah pengamatan $Z_i(t)$ yang terdapat pada tiap N lokasi dalam suatu ruang (i = 1,2,…, N) terhadap t periode waktu. Efek waktu dirumuskan sebagai model deret waktu dan efek lokasi dirumuskan sebagai matriks bobot spasial (Daraputri, 2015).

2.6.1 *Space Time Autoregressive* (STAR)

Model STAR merupakan suatu model yang dikategorikan berdasarkan lag yang berpengaruh secara linier baik dalam lokasi maupun waktu (Pfeifer dan Deutsch, 1980). Model STAR mengasumsikan bahwa penelitian di waktu sekarang dipengaruhi oleh waktu sebelumnya di lokasi tetentu, lokasi-lokasi yang diteliti adalah sama sehingga model ini hanya dapat diterapkan pada lokasi yang bersifat homogen (seragam). Dalam notasi matriks, Model STAR dengan derajat *autoregressive* p dan derajat spasial $\lambda_1, \lambda_2, \dots, \lambda_p$, STAR ($p$; $\lambda_1, \lambda_2, \dots, \lambda_p$) dirumuskan sebagai berikut:

$$Z(t) = \sum_{k=1}^{p} \sum_{l=0}^{\lambda k} \phi_{kl} W^{(l)} Z(t-k) + e(t)$$

Dengan :

(t) : vektor acak ukuran $(n \times l)$ pada waktu t

ϕ_{kl} : parameter STAR pada lag waktu k dan lag spasial l

$W^{(l)}$: matriks bobot ukuran $(n \times n)$ pada lag spasial l (dengan $l = 0,1, \dots n$)

λ_k : spasial lag dari bentuk *autoregressive* orde p

e(t) : vektor *noise* ukuran (n x l) berdistribusi normal multivariat dengan mean 0 dan matriks varian-kovarian $\sigma^2 I_N$

2.6.2 *Generalized Space Time Autoregressive* (GSTAR)

Generalized Space Time Autoregressive (GSTAR) pertama kali diperkenalkan oleh Borovkova, *et al* pada tahun 2002. Model GSTAR merupakan model pengembangan dari model *Space Time Autoregressive* (STAR) dimana GSTAR lebih fleksibel dibandingkan dengan model STAR. Menurut Suhartono dan Subanar (2006) secara matematis notasi dari model GSTAR (1: p) dengan model STAR (1: p). Pada model GSTAR nilai-nilai parameter lag spasial yang sama diperbolehkan berlainan, sedangkan pada model STAR pada parameter *autoregressive* diasumsikan sama pada semua lokasi.

Jika suatu deret $\{(t): t = 0, \pm1, \pm2, \ldots, \pm n\}$ merupakan *multivariat time series* dari N lokasi, maka model GSTAR (p; $\lambda_1, \lambda_2, \ldots, \lambda_p$) :

$$Z(t) = \sum_{k=1}^{p} [\phi_{k0} + \phi_{k1} W] Z_{t-k} + e(t)$$

Dengan :

ϕ_{k0} : matriks diagonal parameter *space time* lag spasial 0 dan parameter *autoregressive* lag waktu ke- k

ϕ_{k1} : matriks diagonal parameter *space time* lag spasial 1 dan parameter *autoregressive* lag waktu ke- k

W : matriks pembobot (n x n)

$e(t)$: vektor *noise* ukuran (n x1)

λ_k : orde spasial dengan k = 1,2,…,p

2.7 Identifikasi Orde Model GSTAR

Menurut Wutsqa (2010) pemilihan orde spasial model GSTAR pada umumnya dibatasi pada orde 1, karena orde yang lebih tinggi akan sulit untuk diinterpretasikan. Sedangkan pada orde waktu (*autoregressive*) dapat ditentukan dengan AIC (*Akaike Information Criterian*). Pemilihan orde model terbaik pada GSTAR dapat ditentukan dengan melihat nilai AIC terkecil.

Perhitungan nilai AIC sebagaimana menurut Akaike (1973, 1974) dalam Lutkepohl (2005) yaitu:

$$AIC(p) = ln\left|\Sigma_{u(p)}\right| + \frac{2p}{T}K^2$$

Dimana $\Sigma_{u(p)} = T^{-1}\sum_{t=1}^{T} \hat{u}_t \, \hat{u}_t'$ adalah matriks taksiran kovarian residual dari model *vector autoregressive* (p), T merupakan jumlah residual dan K merupakan jumlah variabel.

2.8 Pemilihan Bobot Lokasi

Pemilihan atau penentuan bobot lokasi merupakan pemasalahan utama pada pemodelan GSTAR. Penentuan bobot lokasi yang sering digunakan dalam apikasi model GSTAR adalah bobot sergam, invers jarak, biner, korelasi silang. Pada penelitian ini akan digunakan bobot invers jarak pada model GSTAR. Bobot invers jarak adalah pembobotan yang mengacu pada jarak antar lokasi. Semakin lokasi berdekatan maka semakin besar nilai bobot lokasinya. Pembobotan dengan invers jarak mengacu pada jarak antar lokasi, misalkan jarak diantara 4 lokasi didefinisikan :

r_1 = Jarak antara lokasi 1 dengan lokasi 2

r_2 = Jarak antara lokasi 1 dengan lokasi 3

r_3 = Jarak antara lokasi 1 dengan lokasi 4

r_4 = Jarak antara lokasi 2 dengan lokasi 3

r_5 = Jarak antara lokasi 2 dengan lokasi 4

r_6 = Jarak antara lokasi 3 dengan lokasi 4

Dituliskan dalam bentuk matriks sebagai berikut :

$$\begin{bmatrix} 0 & \frac{r_2+r_3}{r_1+r_2+r_3} & \frac{r_1+r_3}{r_1+r_2+r_3} & \frac{r_1+r_2}{r_1+r_2+r_3} \\ \frac{r_4+r_5}{r_1+r_4+r_5} & 0 & \frac{r_1+r_5}{r_1+r_4+r_5} & \frac{r_1+r_4}{r_1+r_4+r_5} \\ \frac{r_4+r_6}{r_2+r_4+r_6} & \frac{r_2+r_6}{r_2+r_4+r_6} & 0 & \frac{r_2+r_4}{r_2+r_4+r_6} \\ \frac{r_5+r_6}{r_3+r_5+r_6} & \frac{r_3+r_6}{r_3+r_5+r_6} & \frac{r_3+r_5}{r_3+r_5+r_6} & 0 \end{bmatrix} = \begin{bmatrix} 0 & W_{12} & W_{13} & W_{14} \\ W_{21} & 0 & W_{23} & W_{24} \\ W_{31} & W_{32} & 0 & W_{34} \\ W_{41} & W_{42} & W_{43} & 0 \end{bmatrix}$$

Matriks diatas distandarkan dalam bentuk W_{ij^*} untuk mendapatkan $\sum_{i \neq j}{}^{(l)} = 1$ (Anggraeni *et al.*, 2013).

2.9 Penduga Parameter Model GSTAR

Pendugaan parameter model GSTAR dilakukan pada semua bobot lokasi dengan menggunakan Metode Kuadrat Terkecil. Berdasarkan model GSTAR dengan orde p=1 dan orde spasial 1 dimana kita dapat tuliskan $\phi_{ki} = \phi_1^{(i)}$ untuk k = 0,1 dapat diturunkan sebagai:

$$Z_i(t) = \phi_{10}^{(i)} Z_i(t-1) + \phi_{11}^{(i)} \sum_{j=1}^{N} W_{ij} Z_j(t-1) + e_i(t)$$

dengan $Z_i(t)$ menyatakan pengamatan pada $t = 0,1, \ldots T$, untuk lokasi $i = 1,2, \ldots, N$ maka:

$$V_i(t) = \sum_{j=1}^{N} W_{ij} Z_j(t)$$

Hal ini berlaku bentuk linier yaitu $Y_i = X_i\beta_i + e_i$

$$Y_i = \begin{bmatrix} Z_i(1) \\ Z_i(2) \\ \vdots \\ Z_i(t) \end{bmatrix}, X = \begin{bmatrix} Z_i(0) & V_i(1) \\ Z_i(1) & V_i(2) \\ \vdots & \vdots \\ Z_i(t-1) & V_i(t-1) \end{bmatrix}, e_i = \begin{bmatrix} e_i(1) \\ e_i(2) \\ \vdots \\ e_i(t) \end{bmatrix}$$

$\beta = (\phi_{10}{}^{(1)}, \phi_{10}{}^{(2)}, \dots, \phi_{10}{}^{(N)}; \phi_{11}{}^{(1)}, \phi_{11}{}^{(2)}, \dots, \phi_{11}{}^{(N)})$.

Penyamaan model untuk semua model linier yaitu Y=Xβ+e , dengan

$Y = (Y_1', \dots, Y_N')'$, $X = \mathrm{diag}(X_1, \dots, X_N)$, $\beta = (\beta_1', \dots, \beta_N')'$, $e = (e_1', \dots, e_N')'$.

Sehingga bentuk estimasi kuadrat terkecil $\widehat{\beta}_T$ adalah

$$\widehat{\beta}_T = (X'X)^{-1}X'y$$

(Borovkova, 2008).

2.10 *Diagnostic Checking*

Cek diagnosa bertujuan untuk memeriksa apakah model yang diestimasi sudah cocok dengan data yang dimodelkan. Pengujian residual meliputi dua tahapan yaitu pengujian residual *white noise* dan uji multivariat normal. Berikut penjelasan dari masing-masing pengujian residual.

2.10.1 Pengujian Residual *White Noise*

Residual dikatakan *white noise* jika memenuhi dua sifat yaitu bersifat identik yang berarti mempunyai varians yang konstan dan bersifat *independent* yang berarti antar residual tidak saling berkorelasi atau residual bersifat homogen serta berdistribusi normal. Berikut hipotesis pengujian *white noise* menggunakan Ljung-Box. (Wei, 2006: 153)

Ho : $\rho_1 = \rho_2 = \dots = \rho_k$ = Residual *White Noise*

H1 : minimal ada satu $\rho_k \neq 0$ dengan k = 1, 2, ..., K (residual tidak *white noise*)

Statistik uji yang digunakan adalah:

$$Q = n(n+2)\sum_{k=1}^{K} \frac{\hat{p}k^2}{n-k}$$

Dengan:

Q : statistik uji Ljung-Box

n : jumlah data pengamatan

$\hat{p}k^2$: ACF residual lag ke-k.

Tolak H0 , jika Q > $X^2_{\alpha;K\text{-}p\text{-}q}$ atau p-value kurang dari α

2.10.2 Residual Multivariat Normal

Pengujian terhadap asumsi ini bermaksud untuk mengetahui *error* dari peramalan dengan model GSTAR mengikuti distribusi normal multivariat atau tidak. Jika pada plot didapat sebaran residual mendekati garis lurus maka residual mengikuti distribusi normal multivariat. Selain menggunakan plot, untuk melihat residual-residual berdistribusi normal atau tidak dapat dilakukan uji formal menggunakan uji kolmogorov smirnov, skewness, kurtosis.

2.11 Metode Kalman Filter

Filter Kalman diperkenalkan oleh R.E Kalman pada tahun 1960. Kalman Filter (KF) adalah suatu metode estimasi variabel keadaan dari sistem dinamik stokastik linear diskrit yang meminimumkan kovariansi *error* estimasi (Welch & Bishop, 2006). Kalman Filter metode pendekatan yang digunakan untuk memperkirakan fungsi parameter dalam peramalan *time series*. Salah satu keunggulan dari metode ini adalah mampu menyelesaikan kasus dengan data yag minim atau sedikit. Data yang dimaksudkan merupakan data alat ukur. Data pengukuran atau data aktual menjadi yang terpenting dari algoritma Kalman Filter karena selisih data berguna untuk koreksi hasil prediksi, jadi menghasilkan estimasi yang mendekati data actual (Masduqi, 2008).

Kalman Filter sebuah metode yang disebut sebagai atau bagian dari ruang keadaan berguna untuk peramalan statistik. Metode ini menggunakan peramalan dengan cara menggunakan data pengamatan sebelumnya dan dilakukan ramalan berikutnya secara optimal berdasar pada informasi tentang data masa lalu dan data saat ini. Model sistem dan model pengukuran :

$$x_{k+1} = A_k\, x_k + B_k u_k + G_k w_k$$

Dimana pengukuran $Z_k \in R^p$ memenuhi

$$Z_k = H_k x_k + v_k$$

$$x_0 \sim (x_0,\, P_{xo}),\; w_k \sim (0,\, Q_k),\; v_k \sim (0,\, R_k)$$

dengan:

x_0 = inisialisasi sistem

x_{k+1} = variabel keadaan pada waktu k +1 dan berdimensi

x_k = variabel keadaan pada waktu k yang nilai estimasi awal x_0 dan kovarian awal P_{xo} , $x_k \in R^p$

u_k = vektor masukan, pada waktu k, $u_k \in R^p$

w_k = *noise* pada sistem dengan mean w_k = 0 dengan kovarian Q_k

Z_k = variabel pengukuran , $z_k \in R^p$

v_k = *noise* pada pengukuran dengan mean v_k= 0 dengan kovarian R_k

A_k, B_k, G_k, H_k = matriks dengan nilai elemen-elemennya merupakan koefisien variabel masing-masing. Variabel $w_k \sim (0,\, Q_k)$, $v_k \sim (0,\, R_k)$ ini diasumsikan *white* (berdistribusi normal dengan mean 0, tidak berkorelasi satu sama lain maupun dengan nilai estimasi awal x_0. Proses estimasi KF terdiri dari dua tahapan yaitu dengan memprediksi variabel keadaan berdasarkan pada siste dinamik atau yang lebih dikenal dengan istilah prediksi (*time update)* dan kedua yaitu tahap koreksi pada data-data pengukuran untuk perbaikan hasil estimasi yang telah diperoleh.

Pertama tahap prediksi digunakan prediksi variabel keadaan dengan persamaan estimasi variabel keadaan dengan akurasi dihitung menggunakan kovarian *error*. Kedua tahap koreksi dari data hasil prediksi dikoreksi menggunakan model pengukuran. Kalman Gain salah satu bagian pada tahap ini, dengan cara menentukan matriks Kalman Gain yang digunakan untuk mengecilkan kovarian *error*. Kedua tahapan tersebut dilakukan secara rekursif dengan cara mengecilkan kovarian *error* .

Setelah berbagai tahapan di atas dilakukan, maka langkah selanjutnya adalah dengan melakukan tahap inisialisasi dengan menggunakan persamaan.

$$P_0 = P_{xo} , x_0 = x_0$$

Langkah selanjutnya dari Kalman Filter adalah *time update* (prediksi) dengan melakukan estimasi atau proyeksi kondisi yang akan datang

$$x_{k+1} = A_k x_k + B_k u_k$$

dan proyeksi kovarian *error* yang akan datang

$$P_{k+1}=A_k P_k A^T+G_k Q_k G_K^T$$

Setelah tahapan tersebut dilakukan, maka langkah selanjutnya adalah melakukan koreksi hitung kalman gain, yaitu

$$K_{k+1}=P_{k+1}H_{K+1}^T(H_{k+1}P_{k+1}H_{k+1}^T+R_{k+1})^{-1}$$

Estimasi awal untuk $xk-1$ dan $Pk-1$

$$X_{k+1} = X_{k+1}^- + K_{k+1}[Z_{k+1} - H_{k+1}X_{k+1}^-]$$

dan *update* kovarian *error*

$$P_{k+1} = (1 - K_{k+1}H_{k+1})\, P_{k+1}^{-}$$

Algoritma Kalman Filter terdiri dari empat bagian yaitu pertama identifikasi model sistem dan model pengukuran, kedua merupakan inisialisasi, selanjutnya untuk ketiga prediksi dan keempat koreksi.

2.12 Kriteria Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan berdasarkan kriteria *out- sample*. Kriteria *out-sample* yang akan digunakan adalah *Root Mean Square Error* (RMSE). RMSE adalah ukuran perbedaan antara nilai prediksi dari model dengan nilai sebenarnya dari observasi. RMSE digunakan untuk memperoleh gambaran keseluruhan standar deviasi yang muncul ketika terjadi perbedaan antar model. Untuk mengetahui besarnya nilai RMSE, dapat digunakan rumus sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Z(t) - \hat{Z}(t))^2}$$

Dengan :

n : Banyak ramalan yang dilakukan

$Z(t)$: Data faktual pada waktu ke-t

$\hat{Z}(t)$: Data peramalan pada waktu ke-t

Model terbaik dapat dipilih dengan melihat nilai RMSE. Model dengan nilai RMSE lebih kecil akan lebih baik dibandingkan dengan model dengan nilai RMSE yang lebih besar.