ISBN:1858-1382

J M S K

Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 6, No.2, 72-85, Januari 2010

Aplikasi Kalman Filter pada Data Survival

Erna Tri Herdiani, Nuravia, Amran, & Sri Astuti Thamrin*

Abstrak

Kalman Filter merupakan metode untuk memprediksi nilai suatu peubah di masa yang akan datang dengan mempertimbangkan data-data sebelumnya yang senantiasa di up-date. Metode ini selanjutnya akan diaplikasikan pada data survival penderita penyakit *Tuberculosis* (TB) dari penduduk Amerika Serikat. Metode ini sangat menarik untuk digunakan karena Tan (2004) hanya memanfaatkan metode state space saja dalam memprediksi nilai peubahnya. Oleh karena itu, pada paper ini akan memanfaatkan metode Kalman Filter dalam memprediksi nilai suatu peubah dari data survival penderita TB di Amerika Serikat.

Kata Kunci: *Metode state space, Kalman Filter, data survival, noise, model stokastik.*

1. Pendahuluan

Tuberculosis (TB) merupakan salah satu penyakit yang membawa kematian sekitar dua sampai tiga juta orang di seluruh dunia setiap tahunnya dan sebesar satu persen sudah tertular oleh bakteri TB. Penderita TB yang tidak berobat dapat menularkan penyakitnya kepada sekitar sepuluh sampai limabelas orang dalam jangka waktu sepuluh tahun. Bakteri *Mycobacterium tuberculosis* penyebab TB menyebar di udara pada saat seorang penderita TB batuk, bersin, meludah ataupun berbicara (PPTI, 2008). Penemuan penderita baru dan pengobatan dini akan memberikan keuntungan bagi program pemberantasan TB. Penyebaran penyakit TB ini dapat dimodelkan secara stokastik. Dikarenakan jumlah penderita penyakit TB setiap tahunnya berubah-ubah, maka hal ini memungkinkan untuk menerapkan analisis deret waktu (*time series*) untuk memprediksi banyaknya penderita TB pada tahun berikutnya berdasarkan data pada tahun-tahun sebelumnya.

Pada kasus penyakit TB ini, model state berupa model stokastik yang diperoleh berdasarkan mekanisme biologi dasar dari penyakit. Jika beberapa data pengamatan baru diperoleh, maka model pengamatan dapat dibentuk. Kombinasi dari model stokastik dan model pengamatan inilah yang dikenal sebagai model State Space (Tan, 2004). Model state space yang dibentuk dari analisa survival ditentukan koefisien parameter melalui Gibs sampling. Selanjutnya model yang diperoleh langsung digunakan untuk memprediksi jumlah penderita TB di Amerika Serikat di waktu yang akan dating. meramalkan Pada paper ini, penulis tertarik untuk mengkaji mengenai aplikasi Kalman Filter pada model State Space penyakit TB yang merupakan model survinal, khususnya dalam memprediksi banyaknya penderita penyakit TB di Amerika Serikat.

* *Prodi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin, email: herdiani_86@yahoo.com, via6188@ymail.com, amran_rmuh@yahoo.com, tutithamrin@yahoo.com*

2. Kalman Filter

Kalman Filter adalah sebuah algoritma untuk memproses data rekursif optimal yang sederhana dan terdiri dari dua fase yang berbeda yaitu fase prediksi dan fase *update* (Orderud, 2005). Metode ini memasukkan semua informasi yang disediakan untuk memproses semua pengamatan yang ada. Metode ini memperhatikan ketelitian dalam menduga nilai dari variabel penting (Maybeck, 1979). Selanjutnya penurunan State Space pada Kalman Filter dapat dilihat sebagai berikut.

2.1. Penurunan State Space

Suatu proses dapat digambarkan dalam bentuk model persamaan:

$$y_k = a_k x_k + v_k, \tag{1}$$

dimana y_k adalah nilai pengamatan yang bergantung pada waktu, a_k adalah jangka waktu penambahan (*gain term)*, x_k adalah informasi, dan v_k adalah gangguan (*noise*). Tujuan yang akan dicapai dari model persamaan diatas adalah menaksir nilai x_k. Selisih dari nilai taksiran $\hat{x}_k$ dengan nilai x_k adalah kesalahan (*error*) dari pengukuran yang dituliskan dalam bentuk fungsi sebagai berikut:

$$f(e_k) = f(x_k - \hat{x}_k). \tag{2}$$

Bentuk khusus dari $f(e_k)$ bergantung pada aplikasinya, tetapi fungsi ini akan positif dan monoton naik yang ditunjukkan oleh fungsi kuadrat error.

$$f(e_k) = (x_k - \hat{x}_k)^2. \tag{3}$$

Nilai harapan dari fungsi kuadrat error disebut fungsi kehilangan (*loss function*) dan menghasilkan fungsi kesalahan kuadrat rata-rata (*Mean Square Error*) atau MSE:

$$\varepsilon(t) = E(e_k^2). \tag{4}$$

Suatu proses yang digambarkan dengan model persamaan sebagai berikut:

$$x_k = \Phi x_{k-1} + w_k, \tag{5}$$

dimana x_k adalah vektor state pada waktu ke-k (**nx1**), Φ adalah matriks state transisi dari state k-1 ke state k dan diasumsikan stasioner (**nxm**), dan w_k adalah proses white noise dengan kovariansi diketahui (**nx1**). Sedangkan model persamaan untuk pengamatan adalah sebagai berkut:

$$z_k = Hx_k + v_k, \tag{6}$$

dimana z_k adalah pengamatan terbaru dari x pada waktu k (**mx1**), H adalah matriks yang menghubungkan antara model state dengan model pengamatan (*noiseless connec*tion) (**mxn**), v_k adalah kesalahan pengamatan yang juga diasumsikan sebagai proses white noise dengan kovariansi diketahui dan korelasi silang (*cross-corelation*) dengan gangguan (*noise*) proses bernilai nol (**mx1**).

Untuk meminimalkan MSE agar diperoleh filter yang optimal, maka model dari kesalahan (*error*) sistem harus diubah dengan menggunakan distribusi Gauss. Kovariansi dari dua model noise diasumsikan stasioner dan dituliskan dalam bentuk:

$$Q = E[w_k {w_k}^T], \tag{7}$$

$$R = E[v_k {v_k}^T], \tag{8}$$

dimana **Q** adalah kovariansi pada vektor state dan **R** adalah kovariansi pada pengamatan. MSE ekuivalen dengan bentuk persamaan berikut:

$$E[e_k {e_k}^T] = P_k \ , \tag{9}$$

dimana P_k adalah matriks kovariansi kesalahan pada waktu k, (**nxn**). Persamaan di atas dapat dikembangkan sehingga diperoleh persamaan sebagai berikut:

$$P_k = E[e_k {e_k}^T] = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]. \tag{10}$$

Diasumsikan taksiran prior dari $\hat{x}_k$ dinamakan ${\hat{x}_k}'$, didapatkan dari informasi dalam sistem. Selanjutnya dapat dibuat suatu persamaan update untuk taksiran terbaru dengan menggabungkan taksiran lama dengan data pengukuran sebagai berikut:

$$\hat{x}_k = {\hat{x}_k}' + K_k(z_k - H{\hat{x}_k}') \ , \tag{11}$$

dimana K_k adalah nilai Kalman. $z_k - H{\hat{x}_k}'$ pada persamaan adalah inovasi atau residu pengukuran: $i_k = z_k - H{\hat{x}_k}'$. (12)

Dengan mensubstitusikan persamaan (6) ke (11), diperoleh persamaan:

$$\hat{x}_k = {\hat{x}_k}' + K_k(Hx_k + v_k - H{\hat{x}_k}'). \tag{13}$$

Selanjutnya, dengan mensubstitusikan persamaan (13) ke (10) diperoleh

$$P_k = E[[(I - K_k H)(x_k - {\hat{x}_k}') - K_k v_k][(I - K_k H)(x_k - {\hat{x}_k}') - K_k v_k]^T], \tag{14}$$

dimana $x_k - {\hat{x}_k}'$ adalah kesalahan dari taksiran prior yang tidak berkorelasi dengan noise pengukuran, sehingga dapat dituliskan seperti pada persamaan berikut:

$$P_k = (I - K_k H)E\left[(x_k - {\hat{x}_k}')(x_k - {\hat{x}_k}')^T\right](I - K_k H)^T + K_k E[v_k v_k^T]K_k^T. \tag{15}$$

Dengan melakukan substitusi persamaan (8) dan (10) ke (14), diperoleh

$$P_k = (I - K_k H){P_k}'(I - K_k H)^T + K_k R {K_k}^T, \tag{16}$$

dimana ${P_k}'$ adalah taksiran prior dari P_k. Persamaan (16) adalah persamaan kovariansi kesalahan ter-update. Diagonal dari matriks kovariansi adalah MSE, yang ditunjukkan sebagai berikut:

$$P_{kk} = \begin{bmatrix} E[e_{k-1}{e_{k-1}}^T] & E[e_k {e_{k-1}}^T] & E[e_{k+1}{e_{k-1}}^T] \\ E[e_{k-1}{e_k}^T] & E[e_k {e_k}^T] & E[e_{k+1}{e_k}^T \\ E[e_{k-1}{e_{k+1}}^T] & E[e_k {e_{k+1}}^T] & E[e_{k+1}{e_{k+1}}^T] \end{bmatrix}. \tag{17}$$

Pada trace matriks kovariansi eror merupakan MSE yang dapat diminimumkan dengan memperkecil trace dari P_k yang juga akan memperkecil trace P_{kk}. Kondisi minimum diperoleh dari turunan pertama trace P_k terhadap K_k dan hasilnya disamakan dengan nol (Lacey, 2007). Perluasan persamaan (16) adalah

$$P_k = P_k' - K_k H P_k' - P_k' H^T {K_k}^T + K_k(H P_k' H^T + R){K_k}^T. \tag{18}$$

Karena trace suatu matriks sama dengan trace transposenya (Lacey, 2007), maka trace dari persamaan (18) dapat dituliskan seperti persamaan berikut:

$$T[P_k] = T[P_k'] - 2T[K_k H P_k'] + T[K_k(H P_k' H^T + R){K_k}^T], \tag{19}$$

dimana $T[P_k]$ adalah trace dari matriks P_k dan selanjutnya didiferensialkan terhadap K_k, sehingga diperoleh

$$-2\left(HP_k'\right)^T + 2K_k\left(HP_k'H^T + R\right) = 0. \tag{20}$$

Selanjutnya, $\left(HP_k'\right)^T = K_k\left(HP_k'H^T + R\right)$ (21)

Sehingga $K_k = P_k'H^T\left(HP_k'H^T + R\right)^{-1}.$ (22)

Persamaan (22) disebut persamaan penambahan Kalman (*Kalman gain*)**.** Persamaan tersebut digunakan untuk memperbaharui pengamatan. i_k pada persamaan (12) berhubungan dengan *kovariansi prediksi pengukuran (inovasi kovariansi)*, yang didefinisikan oleh persamaan berikut:

$$S_k = HP_k'H^T + R. \tag{23}$$

Dengan mensubtitusikan persamaan (22) ke (18) diperoleh

$$P_k = \left(I - K_k H\right)P_k'. \tag{24}$$

Persamaan (24) merupakan matriks kovariansi kesalahan yang diupdate melalui penambahan optimal (*optimal gain*). Untuk menduga variabel x_k digunakan persamaan (13), (22), dan (24). Proyeksi state diperoleh dengan menggunakan

$$\hat{x}_{k+1}' = \Phi\hat{x}_k. \tag{25}$$

Dengan proyeksi matriks kovariansi eror untuk interval waktu berikutnya, $k+1$

$$e_{k+1}' = x_{k+1} - \hat{x}_{k+1}' = \Phi e_k + w_k. \tag{26}$$

Dengan mengembangkan persamaan (10) untuk $k+1$ diperoleh

$$P_{k+1}' = E\left[e_{k+1}' e_{k+1}^{T'}\right] = E[(\Phi e_k + w_k)(\Phi e_k + w_k)^T], \tag{27}$$

e_k dan w_k mempunyai nilai korelasi silang sama dengan nol karena w_k terakumulasi diantara *k* dan *k+1* sedangkan eror e_k meningkat sampai waktu *k*. Oleh karena itu persamaan (27) dapat dituliskan:

$$P_k' = \Phi P_k \Phi^T + Q \ . \tag{28}$$

Persamaan ini merupakan taksiran prior pada matriks kovariansi kesalahan yang dapat digunakan dalam memprediksi state selanjutnya (Lacey, 2007).

2.2. Update Kovariansi Model

Model kovariansi dari parameter dalam bentuk inversnya dikenal sebagai *matriks informasi*. Kalman Filter disebut *Filter Informasi* ketika dibangun berdasarkan matriks informasi tersebut. Selanjutnya dapat diformulasikan persamaan *update* sebagai alternatif untuk matriks kovariansi dengan menggunakan penjabaran standar eror sebagai berikut:

$$P_k^{-1} = P_k'^{-1} + HR^{-1}H^T. \tag{29}$$

Dapat dilihat bahwa persamaan (29) dan (24) ekuivalen. Ini dapat dibuktikan dengan menggunakan identitas $P_k \times P_k^{-1} = I$. Dengan bentuk persamaan kovariansi yang diperbaharui adalah $P_k = (I - K_k H)P_k'$ dan $P_k^{-1} = P_k'^{-1} + HR^{-1}H^T$, oleh karena itu

$$(I - K_k H)P_k' \times \mathrm{P}_k'^{-1} + HR^{-1}H^T = I. \tag{30}$$

Dalam tutorial Kalman Filter (Tony Lacey), dengan mensubtitusikan K_k diperoleh

$$[P_k' - P_k' H^T (H P_k' H^T + R)^{-1} H P_k'] \left[P_k'^{-1} + H^T R^{-1} H \right] = I. \qquad (31)$$

Jadi terbukti bahwa persamaan (29) dan (24) ekuivalen.

2.3. Model State Space

Model State Space menggabungkan model stokastik dengan model statistik untuk menguraikan suatu proses. Model State Space dari penyakit Tuberculosis menyimpulkan informasi biologi dari sistem melalui model sistem stokastik dan menggabungkannya dengan data terbaru melalui persamaan pengamatan. Keuntungan spesifik dari model ini adalah adanya prosedur optimal untuk memperbaharui model data baru yang merupakan langkah penghalusan pada model State Space (Tan, 2004)**.**

3. Sekilas Mengenai Penyakit Tuberculosis

Tuberkulosis (TB) adalah salah satu jenis penyakit menular yang disebabkan oleh bakteri "*Mycobacterium tuberculosis* ". Bakteri ini dapat menyerang semua bagian tubuh manusia, dan yang paling sering terkena adalah organ paru (90%). Bakteri ini berbentuk batang dan bersifat tahan asam sehingga dikenal juga sebagai Batang Tahan Asam (BTA) (PPTI, 2008). Penyakit TB dapat menyerang siapa saja tak terkecuali pria, wanita, tua, muda, kaya dan miskin serta dimana saja.

Seperti halnya flu, bakteri TB menyebar di udara pada saat seorang penderita TB batuk, bersin, meludah ataupun berbicara. Penderita TB yang tidak berobat dapat menularkan penyakitnya kepada sekitar 10-15 orang dalam jangka waktu 10 tahun (PPTI, 2008), dan yang menularkan adalah mereka yang di dalam dahaknya terdapat bakteri TB (penderita TB BTA positif) (Poeloengan, 2005). Setelah lima tahun, 50% dari penderita TB tanpa pengobatan akan meninggal, 25% akan sembuh sendiri dengan daya tahan tubuh tinggi, dan 25% sebagai kasus kronik yang tetap menular.

Timbulnya *HIV* sangat berpengaruh terhadap peningkatan jumlah penderita TB. Infeksi *HIV* mengakibatkan kerusakan luas sistem daya tahan tubuh seluler (*Cellular Immunity*), sehingga jika terjadi infeksi oportunistik, seperti Tuberkulosis, maka yang bersangkutan akan menjadi sakit parah bahkan mengakibatkan kematian. Bila jumlah orang terinfeksi *HIV* meningkat, maka jumlah penderita TB akan meningkat, dengan demikian penularan TB di masyarakat akan meningkat pula (Poeloengan, 2005).

Perawatan bagi penderita TB diharuskan minum obat dengan teratur dan benar sesuai dengan anjuran dokter selama enam bulan berturut-turut tanpa terputus dengan melibatkan petugas kesehatan atau anggota keluarga untuk mengawasi dan memastikan penderita TB minum obat dengan teratur dan benar (PPTI, 2008). Menurut Poelongan, 2005, pencegahan terhadap TB dapat dilakukan dengan pemberian imunisasi BCG (*Bacillus Calmette Guerin*) pada bayi serta meningkatkan daya tahan tubuh dengan makanan bergizi dan menjaga pola hidup bersih.

Proses Stokastik dari Penyakit Tuberculosis

Sebuah proses stokastik $\{X(t), t \in T\}$ adalah kumpulan dari variabel acak, yaitu untuk setiap $t \in T, X(t)$ adalah sebuah variabel acak, dimana t adalah waktu dan $X(t)$ sebagai state dari proses pada waktu *t*. Himpunan T disebut indeks himpunan dari proses, jika T adalah himpunan yang dapat dihitung (*countable set*), proses stokastik dikatakan proses waktu diskrit (*discrete-time*). Sedangkan jika T adalah sebuah interval dari garis bilangan real, maka proses stokastik dikatakan proses waktu kontinu (*continuous-time*). Sebagai contoh, $\{X_n, n = 0,1,..,n\}$ adalah sebuah proses stokastik dengan waktu diskrit yang dilambangkan dengan bilangan bulat non-negatif; sedangkan $\{X(t), t \geq 0\}$ adalah sebuah proses stokastik dengan waktu kontinu yang dilambangkan dengan bilangan real non negatif (Ross, 1997).

Pada populasi yang beresiko terkena penyakit Tuberculosis, kelompok manusia dapat dibagi menjadi 2 kelompok yaitu kelompok manusia yang sehat (N_1) yaitu kelompok yang tidak menderita penyakit TB dan kelompok manusia yang sakit (N_2) dalam hal ini penderita TB. Jadi $X(t)$ terdiri dari $N_1(t)$ dan $N_2(t)$. Pada kedua kelompok itu di dalamnya terdapat proses kelahiran, kematian, imigrasi, sakit dan sembuh dari penyakit TB.

Selama interval waktu $[t, t + \Delta t)$, proses kelahiran-kematian-sakit-sembuh mengikuti distribusi Multinomial dan proses imigrasi adalah proses Poisson. Untuk memperoleh persamaan diferensial stokastik untuk variabel state $N_1(t)$ dan $N_2(t)$, amati bahwa $X(t + \Delta t)$ diperoleh dari $X(t)$ secara stokastik melalui proses kelahiran-kematian-sakit-sembuh dan proses imigrasi. Sehingga menurut Tan, 2004, persamaan diferensial stokastik untuk $N_i(t), i = 1,2$ adalah:

$$\Delta N_1(t) = N_1(t + \Delta t) - N_1(t), \quad (32)$$
$$\Delta N_2(t) = N_2(t + \Delta t) - N_2(t). \quad (33)$$

4. Aplikasi Kalman Filter Pada Model Survival

Pada sebuah populasi yang mempunyai individu beresiko terkena penyakit Tuberculosis (TB) terbagi menjadi 2 kelompok yaitu kelompok manusia yang sehat (N_1) yaitu kelompok yang tidak menderita penyakit TB dan kelompok manusia yang sakit dalam hal ini terinfeksi Tuberculosis dan sakit (N_2), dimana hal ini merupakan proses stokastik dua dimensi yaitu N_1 dan N_2. Ketika suatu populasi beresiko sakit, maka kemungkinan yang dapat terjadi adalah:

1. Anggota kelompok N_1 menderita sakit karena kontak dengan anggota N_2 atau melalui perantaranya.
2. Anggota kelompok N_2 dapat meninggal akibat penyakit TB atau sebab lainnya sedangkan anggota kelompok N_1 hanya dapat meninggal akibat penyakit lainnya.
3. Anggota kelompok N_2 dapat disembuhkan dengan obat sehingga masuk ke kelompok N_1.
4. Proses kelahiran diabaikan dengan asumsi bahwa sangat kecil kemungkinan bayi yang baru lahir menderita penyakit TB.
5. Selain itu terdapat proses imigrasi dalam populasi.

Jadi dalam sistem stokastik perpindahan penyakit Tuberculosis memasukkan proses kelahiran, kematian, imigrasi, sakit dan sembuh dari penyakit. Misalkan $X(t), t \geq 0 = N_1(t)$ dan $N_2(t)$, dimana $N_i(t)$ menyatakan banyaknya orang $N_i(i = 1,2)$ dalam populasi pada waktu t.

Variabel-variabel transisi yang akan digunakan dalam persamaan sistem stokastik sebagai berikut: $F_1(t)$ = Banyaknya orang sehat yang menjadi sakit selama $[t, t + \Delta t)$; $F_2(t)$ = Banyaknya orang sakit yang disembuhkan dengan obat selama $[t, t + \Delta t)$; $B_1(t)$ = Banyaknya kelahiran dari orang sehat selama $[t, t + \Delta t)$; $B_2(t)$ = Banyaknya kelahiran dari orang sakit selama $[t, t + \Delta t)$; $D_1(t)$ = Banyaknya kematian dari orang sehat selama $[t, t + \Delta t)$; $D_2(t)$ = Banyaknya kematian dari orang sakit selama $[t, t + \Delta t)$; $R_1(t)$ = Banyaknya imigrasi orang sehat selama $[t, t + \Delta t)$; $R_2(t)$ = Banyaknya imigrasi orang sakit selama $[t, t + \Delta t)$; $\alpha_1 =$ laju penyakit (laju transisi dari $N_1 \rightarrow N_2$); $\alpha_2 =$ laju kesembuhan (laju transisi $N_2 \rightarrow N_1$); $b_1 =$ laju kelahiran dari orang sehat; $b_2 =$ laju kelahiran dari orang sakit; $d_1 =$ laju kematian dari orang sehat; $d_2 =$ laju kematian dari orang sakit; λ_1 = laju imigrasi dari orang sehat dan λ_2 = laju imigrasi dari orang sakit.

Perpindahan keadaan dari sakit ke sehat atau dari sehat ke sakit akan digunakan sebagai model state. Model state pada penyakit TB dapat diperoleh berdasarkan mekanisme biologi dasar dari penyakit TB yaitu melalui proses penularannya. Selama interval waktu $[t, t+\Delta t)$, proses kelahiran, kematian, sakit dan sembuh mengikuti distribusi Multinomial dengan parameter yang digunakan adalah $\{N_i(t); b_i(\mathrm{t})\Delta t, d_i(\mathrm{t})\Delta t, \alpha_i(\mathrm{t})\Delta t\}$ sedangkan proses imigrasi merupakan proses Poisson dengan rata-rata (mean) $\lambda_i(t)\Delta t$. Karena proses imigrasi merupakan proses Poisson, maka model state dapat diperoleh dari persamaan diferensial stokastik variabel state $N_1(t)$ dan $N_2(t)$, yang terlebih dahulu diamati bahwa $X(t+\Delta t)$ diperoleh dari $X(t)$ secara stokastik melalui proses kelahiran-kematian-sakit-sembuh dan proses imigrasi.

Distribusi peluang bersyarat dari $R_i(t)$ dan $\{B_i(t), D_i(t), F_i(t)\}$ dengan syarat $N_i(t)$ diketahui, dimana $i = 1,2$ adalah $R_1(\mathrm{t})|N_1(t) \sim$ Poisson dengan mean $N_1(t)\lambda_1(t)\Delta t$, $R_2(\mathrm{t})|N_2(t) \sim$ Poisson dengan mean $N_2(t)\lambda_2(t)\Delta t$, $\{B_1(t), D_1(t), F_1(t)\}|N_1(t) \sim$Multinomial $[N_1(t); b_1(\mathrm{t})\Delta t, d_1(\mathrm{t})\Delta t, \alpha_1(\mathrm{t})\Delta t]$, dan $\{B_2(t), D_2(t), F_2(t)\}|N_2(t) \sim$ Multinomial $[N_2(t); b_2(\mathrm{t})\Delta t, d_2(\mathrm{t})\Delta t, \alpha_2(\mathrm{t})\Delta t]$.

Dalam state $X(t)$, $\{R_i(t), i = 1,2,$ dan $\left[B_j(t), D_j(t), F_j(t)\right],\ j = 1,2\}$ berdistribusi saling bebas satu sama lainnya. Sehingga untuk $i = 1,2$, didefinisikan $\varepsilon_i(t)$, sebagai berikut

$$\begin{aligned}\varepsilon_i(t)\Delta t = [R_i(t) - N_i(t)\lambda_i(t)\Delta t] + [B_i(t) - b_i(t)N_i(t)\Delta t]\\ -[F_i(t) - \alpha_i(t)N_i(t)\Delta t] - [D_i(t) - d_i(t)N_i(t)\Delta t]\end{aligned} \tag{34}$$

Persamaan diferensial stokastiknya untuk $N_i(t), i = 1,2$ adalah:

$$\begin{aligned}\Delta N_1(t) = N_1(t+\Delta t) - N_1(t)\\ = \{[\lambda_1(t) + b_1(t) - d_1(t) - \alpha_1(t)]N_1(t) + \alpha_2(t)N_2(t)\}\Delta t + \varepsilon_1(t)\Delta t\end{aligned} \tag{35}$$

$$\begin{aligned}\Delta N_2(t) = N_2(t+\Delta t) - N_2(t)\\ = \{[\lambda_2(t) + b_2(t) - d_2(t) - \alpha_2(t)]N_2(t) + \alpha_1(t)N_1(t)\}\Delta t + \varepsilon_2(t)\Delta t\end{aligned} \tag{36}$$

Pada persamaan diatas, noise $\varepsilon_i(t)$, dimana $i = 1,2$ bersifat acak dengan nilai ekspektasi sama dengan nol dan tidak berkorelasi dengan variabel state $\{N_i(t), i = 1,2\}$. Pada tahap berikutnya untuk memperbaharui model state, diperlukan model pengamatan agar prediksi yang dilakukan lebih akurat.

Model pengamatan diperoleh berdasarkan data pengamatan terbaru dari penyakit TB. Misalkan $Z_i(k)$ merupakan banyaknya N_i yang diamati pada waktu $t_k, k = 1,2,..,n$ dan $\varepsilon_i(k) = [Z_i(k) - N_i(t_k)]/\sqrt{N_i(t_k)}$ berdistribusi normal dengan rata-rata 0 dan variansi σ_i^2, saling bebas untuk $i = 1,2,\ k = 1,2,\dots,n$. Maka model pengamatan ditunjukkan sebagai berikut:

$$Z_1(k) = N_1(t_k) + [N_1(t_k)]^{\frac{1}{2}}\varepsilon_1(k), \tag{37}$$

$$Z_2(k) = N_2(t_k) + [N_2(t_k)]^{\frac{1}{2}}\varepsilon_2(k),\ \text{untuk}\ k = 1,2,\dots,n, \tag{38}$$

dimana ε_1 dan ε_2 saling bebas dan berdistribusi normal dengan mean sama dengan 0 dan variansi ${\sigma_1}^2$ dan ${\sigma_2}^2$. Berdasarkan persamaan (6) untuk model pengamatan $z_k = Hx_k + v_k$, dimana H = 1, karena hanya bergantung pada x_k dan v_k adalah *noise* pengamatan yang ditunjukkan oleh $\varepsilon_i(k)$ dalam persamaan. Berdasarkan model state dan model pengamatan, selanjutnya akan diaplikasikan pada data penyakit Tuberculosis di Amerika Serikat.

4.1. Sumber Data

Data yang digunakan merupakan data sekunder yaitu untuk data state, data diperoleh dari jurnal yang berjudul “State Space Models for Survival Analysis” oleh Weiming Ke bersama Dr. Wai-yuan Tan (2004) sedangkan untuk data pengamatan digunakan data penduduk Amerika

Serikat pada tahun 1993-2007 yang bersumber dari *http://www.disastercenter.com/crime/uscrime.htm* dan data penderita penyakit Tuberculosis di Amerika Serikat pada tahun 1993-2007 yang bersumber dari *http://www.cdc.gov/tb/*. Adapun data tersebut dapat dilihat pada Tabel 1.

4.2. Analisa Data

Pertambahan penduduk Amerika Serikat pada tahun 1980 sampai 1992 menunjukkan bahwa jumlah penduduk Amerika Serikat mengalami pertambahan secara linear setiap tahunnya yaitu dari tahun 1980 sampai tahun 1992. Sedangkan pada tahun 1980 sampai tahun 1985 penderita Tuberculosis di Amerika Serikat mengalami penurunan jumlah penderita. Akan tetapi pada tahun berikutnya jumlah penderita terus bertambah sampai tahun 1992. Perubahan ini disebabkan adanya efek dari *HIV* di Amerika Serikat (Tan 2004). Oleh karena itu sebelum menyusun model state, terlebih dahulu harus diketahui parameter yang mempengaruhi perubahan N_1 dan N_2.

Tabel 1. Data Kasus Penyakit TB di Amerika Serikat Tahun 1980-1992.

Tahun	Total Penduduk (jiwa)	Sehat (jiwa)	Penderita TB (jiwa)
1980	226.545.805	226.517.805	28.000
1981	228.762.212	228.734.712	27.500
1982	230.978.619	230.953.219	25.400
1983	233.195.025	233.171.025	24.000
1984	235.411.432	235.389.231	22.201
1985	237.627.839	237.605.638	22.201
1986	239.844.246	239.821.246	23.000
1987	242.060.653	242.037.653	23.000
1988	244.277.059	244.254.059	23.000
1989	246.493.466	246.469.666	23.800
1990	248.709.873	248.684.373	25.500
1991	250.926.280	250.899.997	26.283
1992	253.142.687	253.116.014	26.673

Sumber: “*State Space Models for Survival Analysis*”, Tan (2004).

Adapun parameter pada model state adalah laju kelahiran, laju kematian, laju penyakit, laju kesembuhan dan laju imigrasi (Tan, 2004). Hasil pendugaan parameter pada model state adalah $d_1 =$ laju kematian dari orang sehat = 0.04, $d_2 =$ laju kematian dari orang sakit = 0.1, λ_1 = laju imigrasi dari orang sehat = 0.05, λ_2 = laju imigrasi dari orang sakit = 0.03, $\alpha_1 =$ laju penyakit = 0.2, $\alpha_2 =$ laju kesembuhan = 0.4. Jika nilai parameter disubtitusikan pada persamaan (35) dan (36), diperoleh:

$$\begin{aligned} \Delta N_1(t) &= \{[\lambda_1(t) - d_1(t) - \alpha_1(t)]N_1(t) + \alpha_2(t)N_2(t)\}\Delta t + \varepsilon_1(t)\Delta t \\ &= -0{,}19N_1(t)\Delta t + 0{,}4N_2(t)\,\Delta t + \varepsilon_1(t)\Delta t \end{aligned} \tag{39}$$

$$\begin{aligned} \Delta N_2(t) &= \{[\lambda_2(t) - d_2(t) - \alpha_2(t)]N_2(t) + \alpha_1(t)N_1(t)\}\Delta t + \varepsilon_2(t)\Delta t \\ &= -0{,}11N_2(t)\Delta t + 0{,}2N_1(t)\,\Delta t + \varepsilon_2(t)\Delta t \end{aligned} \tag{40}$$

Sehingga model state untuk $N_1(t)$ dan $N_2(t)$ adalah

$$N_1(t) = N_1(t-1) - 0{,}19N_1(t-1)\Delta t + 0{,}4N_2(t-1)\,\Delta t + \varepsilon_1(t)\Delta t \quad (41)$$
$$N_2(t) = N_2(t-1) - 0{,}11N_2(t-1)\Delta t + 0{,}2N_1(t-1)\,\Delta t + \varepsilon_2(t)\Delta t \quad (42)$$

Adapun data yang digunakan sebagai data pengamatan untuk memperbaharui model state, dapat dilihat pada Tabel 2 berikut.

Tabel 2. Data Kasus Penyakit TB di Amerika Serikat Tahun 1993-2007.

Tahun	Total Penduduk (jiwa)	Sehat (jiwa)	Penderita TB (jiwa)
1993	257.908.000	257.882.893	25.107
1994	260.341.000	260.316.795	24.205
1995	262.755.000	262.732.272	22.728
1996	265.228.572	265.207.362	21.210
1997	267.637.000	267.617.249	19.751
1998	270.296.000	270.277.713	18.287
1999	272.690.813	272.673.312	17.501
2000	281.421.906	281.405.597	16.309
2001	285.317.559	285.301.613	15.946
2002	287.973.924	287.958.868	15.056
2003	290.690.788	290.675.951	14.837
2004	293.656.842	293.642.341	14.501
2005	296.507.061	296.492.996	14.065
2006	299.398.484	299.384.730	13.754
2007	301.621.157	301.607.858	13.299

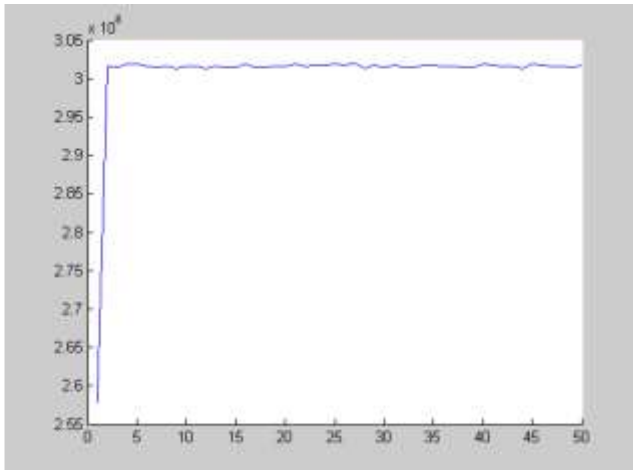
Sumber: *http://www.disastercenter.com/crime/uscrime.htm*,*http://www.cdc.gov/tb/*.

Dengan keadaan penduduk di Amerika Serikat pada tahun 1993 sampai 2007 seperti pada data di atas. Maka dapat disimpulkan bahwa jumlah penduduk di Amerika Serikat terus mengalami peningkatan dari tahun ke tahun. Namun penderita TB di Amerika Serikat mengalami penurunan dari tahun 1993 hingga tahun 2007. Berdasarkan data penderita Tuberculosis di Amerika Serikat tahun 1980 sampai 1992, Weiming Ke dan Dr. Wai Yuan Tan (2004) melakukan pendugaan terhadap peluang survival dari kelompok N_1 dan N_2 untuk beberapa tahun kedepan dengan metode *Multi Level Gibbs Sampling*.

Pada paper ini, penulis menerapkan data yang sama untuk model state dan data pengamatan sebagai data tambahan untuk memprediksi N_1 dan N_2 pada tahun berikutnya dengan metode Kalman Filter. Proses rekursif Kalman Filter untuk peramalan jumlah penduduk sehat dan penderita penyakit TB dilakukan dengan menggunakan hasil prediksi jumlah penduduk yang sehat pada iterasi ke-39 sampai 50 dapat dilihat pada Gambar 1 dan Tabel 3.

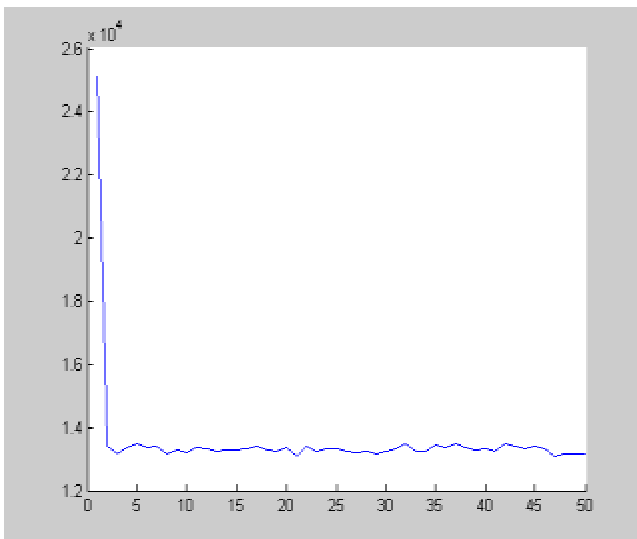
Tabel 3. Prediksi Penduduk Sehat.

Iterasi	Xa	Iterasi	Xa	Iterasi	Xa	Iterasi	Xa
39	301.440.000	42	301.520.000	45	301.850.000	48	301.550.000
40	301.890.000	43	301.610.000	46	301.690.000	49	301.420.000
41	301.670.000	44	301.330.000	47	301.630.000	50	301.730.000



Gambar 1. Plot Prediksi Penduduk Sehat.

Sedangkan pada Gambar 2 dan Tabel 4 dapat dilihat hasil prediksi jumlah penderita TB di Amerika Serikat sebagai berikut.



Gambar 2. Plot Prediksi Penderita TB.

Tabel 4. Prediksi Penderita TB.

Iterasi	Xa	Iterasi	Xa	Iterasi	Xa	Iterasi	Xa
39	13.305	42	13.485	45	13.426	48	13.179
40	13.334	43	13.392	46	13.317	49	13.182
41	13.244	44	13.348	47	13.080	50	13.163

Dari Tabel 3, diperoleh hasil prediksi banyaknya penduduk sehat (tidak menderita penyakit TB) pada tahun 2008 adalah sekitar 301.730.000 jiwa dengan nilai variansi sebesar $2,87 \times 10^{29}$. Sedangkan prediksi penderita TB pada tahun 2008 (dapat dilihat pada Tabel 4) sebanyak 13.163 jiwa dengan nilai variansi sebesar $1,15 \times 10^{15}$. Dengan nilai Kalman yang diperoleh yaitu K = 1. Nilai rata-rata yang diperoleh dari 50 *forecasting* (50 kali program dijalankan) yaitu prediksi penduduk sehat sebanyak 301.629.000 jiwa sedangkan prediksi penderita TB sebanyak 13.311 jiwa.

Selanjutnya untuk melihat besarnya simpangan baku tiap kelompok, digunakan bilangan baku dengan rumus:

$$z_i = \frac{x_i - \bar{x}}{s} \; i = 1,2,..n, \tag{43}$$

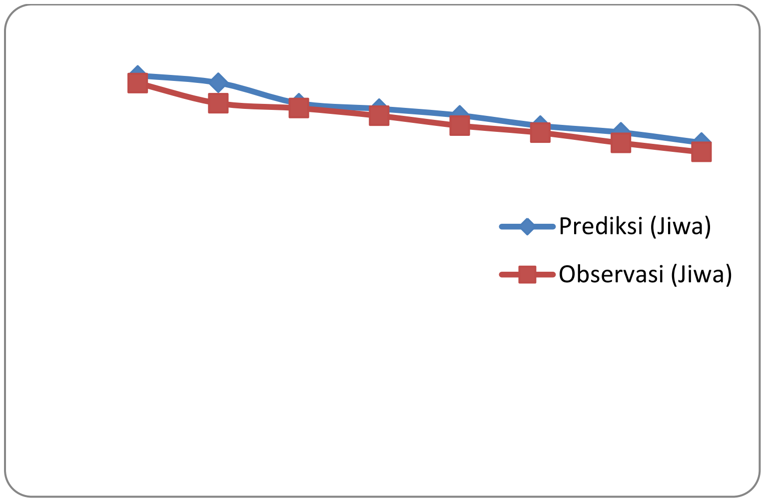
dimana z_i = bilangan baku ke-I; x_i = data ke-I; $\bar{x}$ = rata-rata data; s = simpangan baku data, (Tiro, 2004). Bilangan baku untuk kelompok penderita TB dan penduduk sehat dapat digunakan sebagai data pengamatan dalam Kalman Filter. Pada Tabel 5 dapat dilihat hasil *forcasting* penduduk sehat dan penderita TB.

Tabel 5. Prediksi Penduduk Sehat dan Penderita TB Menggunakan Data Bilangan Baku.

Penduduk Sehat	Penderita TB
1,4595	-1,1348
1,4498	-1,125
1,4517	-1,1152
1,4635	-1,1094
1,4614	-1,1328
1,4751	-1,1445
1,4606	-1,1123
1,4623	-1,1406
1,4614	-1,1289
1,4617	-1,1221

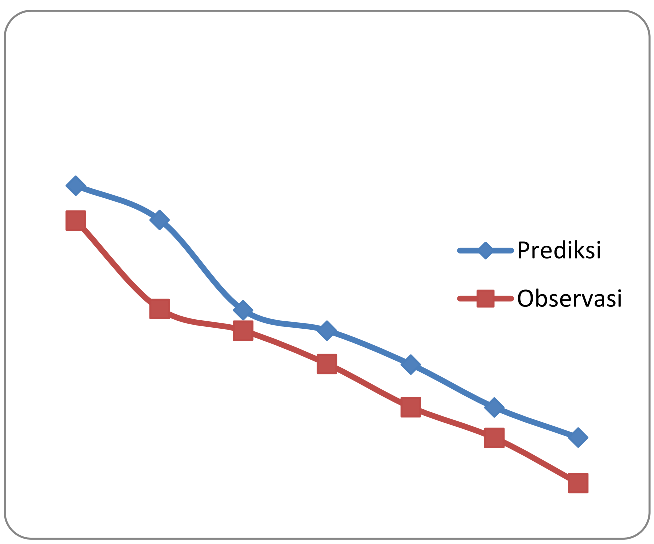
Nilai rata-rata penduduk sehat adalah 1,4607 dengan simbangan baku sebesar 0,006839, sedangkan untuk penderita TB diperoleh rata-rata = -1,12656 dengan simpangan baku sebesar 0,011928. Jika dikonversi ke nilai awal diperoleh prediksi penduduk sehat sebanyak 301.629.659 jiwa dan penderita TB sebanyak 13.326 jiwa, dimana hasil yang diperoleh tidak berbeda jauh baik menggunakan bilangan baku maupun tidak.

Berdasarkan laporan CDC Amerika Serikat diketahui bahwa jumlah penderita TB di Amerika Serikat pada tahun 2008 sebanyak 12.898 jiwa dari total penduduk 304.482.526 jiwa (penduduk sehat = 304.469.628 jiwa). Sehingga dapat diketahui persentase kesalahan (*eror*) dari prediksi penduduk sehat sebesar 0,93% dan penderita TB sebesar 3,20%. Ini menunjukkan bahwa hasil prediksi yang diperoleh dengan metode Kalman Filter cukup akurat. Pada Gambar 3 dapat dilihat perbandingan antara hasil prediksi penderita TB dengan Kalman Filter dan jumlah penderita TB sebenarnya (observasi) dari tahun 2001-2008.



Gambar 3. Plot Prediksi Penderita TB dan Jumlah Penderita TB yang Sebenarnya Tahun 2001-2008.

Dari Tabel 5 dapat dilihat bahwa rata-rata dari 50 hasil prediksi penderita TB pada tahun 2001 sebanyak 16.921 jiwa sedangkan hasil observasi (jumlah penderita TB sebenarnya pada tahun 2001) adalah 15.946 dengan tingkat kesalahan sebesar 2,16%. Plot perbandingan antara hasil prediksi dengan data sebenarnya (observasi) dapat dilihat pada Gambar 3, akan tetapi untuk menunjukkan bahwa tingkat kesalahan hasil prediksi dengan data sebenarnya cukup kecil, dapat digunakan bilangan baku seperti yang terlihat pada Gambar 4 yang diperoleh dari nilai rata-rata 10 *forecasting* pada tahun 2001-2007.



Gambar 4. Plot Prediksi Penderita TB dan Jumlah Penderita TB yang Sebenarnya Tahun 2001-2007 dengan Menggunakan Bilangan Baku.

5. Kesimpulan

Kesimpulan yang dapat dituliskan dalam paper ini adalah

1. Model state untuk penderita penyakit Tuberculosi di Amerika Serikat adalah
   $N_1(t) = N_1(t-1) - 0{,}19N_1(t-1)\Delta t + 0{,}4N_2(t-1)\,\Delta t + \varepsilon_1(t)\Delta t$
   sedangkan model state untuk penduduk yang sehat adalah
   $N_2(t) = N_2(t-1) - 0{,}11N_2(t-1)\Delta t + 0{,}2N_1(t-1)\,\Delta t + \varepsilon_2(t)\Delta t$

2. Model pengamatan untuk penderita penyakit Tuberculosis di Amerika Serikat adalah $Z_1(k) = N_1(t_k) + [N_1(t_k)]^{\frac{1}{2}}\varepsilon_1(k)$, sedangkan model pengamatan untuk penduduk yang sehat adalah $Z_2(k) = N_2(t_k) + [N_2(t_k)]^{\frac{1}{2}}\varepsilon_1(k)$.

3. Prediksi jumlah penderita penyakit Tuberculosis di Amerika Serikat untuk tahun 2008 adalah 13.311 jiwa sedangkan jumlah penduduk yang sehat sebanyak 301.629.000 jiwa.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Universitas Hasanuddin yang telah membiayai pelaksanaan penelitin ini melalui DIPA Universitas Hasanuddin Tahun 2009 sesuai dengan surat perjanjian pelaksanaan pekerjaan hibah penelitian berdasarkan nomor kontrak 2806/H4.13.2/PL.09/2009 tanggal 13 Agustus 2009.

Daftar Pustaka

[1] Itamar, K., Thomas, L., dan Yuval, M., 2008, Robotic mapping, *Jurnal Student Lecture*.

[2] Lacey, T., 2007, *Tutorial: The Kalman Filter.*

[3] Maybeck, P.S., 1979, *Stochastics Models, Estimation, and Control. Volume I,* Academic Press, New York.

[4] Orderud, F., *Comparison of Kalman Filter Estimation Approaches for State Space Models with Nonlinear Measurement*.

[5] Poeloengan, M. dkk., *Lokakarya Nasional Penyakit Zoonosis : Bahaya dan Penanganan Tuberculosis.*

[6] PPTI, 2008, http://www.dinkes-sleman.go.id/files/news47b3b12895520.pdf. [Diakses: Mei 2009]

[7] Pujiati, S.A., 2008, *Perbandingan Metode Peramalan untuk Deret Waktu Musiman*, Pasca Sarjana Jurusan Statistika: FMIPA ITS.

[8] Ross, S.M., 1997, *Introduction To Probability Models. Sixth Edition*, Academic Press.USA.

[9] Tan, W.Y., Ke, W., 2004, *State Space Models for Survival Analysis*, Elsevier Science B.V.

[10] Tiro, M. A., 2004, *Dasar-Dasar Statistika*, State University of Makassar Press, Makassar.

[11] Tuberculosis. http://www.cdc.gov/tb/. [Diakses: April 2009]

[12] USA, http:// www.disastercenter.com/crime/uscrime.htm. [Diakses: April 2009]

[13] Vazquez, A. dan Syversveen, A.R., 2006, The *Ensemble Kalman Filter*, Norsk Regnesentral: Norwegia.