Analysis of Crimean-Congo Hemorrhagic Fever Incidents with Dynamically Weighted Particle Filter

Duchwan Ryu

with D. Bilgili, Ö. Ergönül,
F. Liang, and N. Ebrahimi

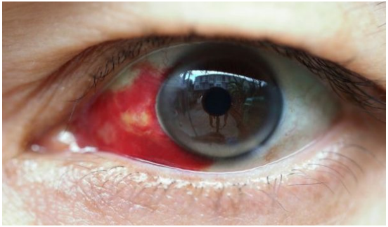
Northern Illinois University

May 18, 2018

Outline

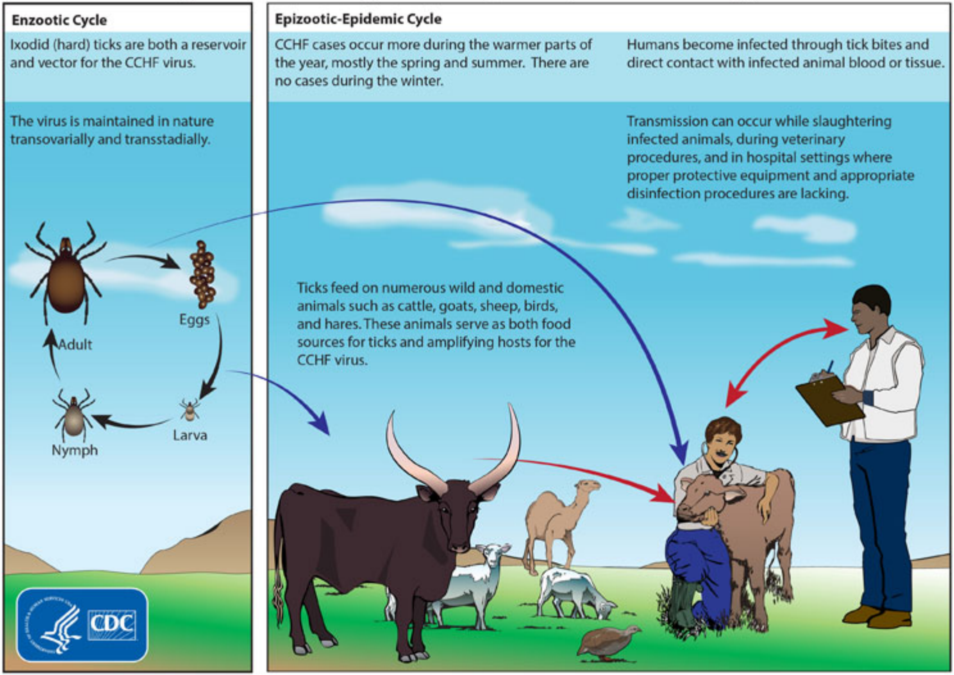
- Crimean-Congo Hemorrhagic Fever Incidents in Turkey
- A Dynamic Model for CCHF Incidents
- Estimation with Dynamically Weighted Particle Filter
- Concluding Remarks

Crimean-Congo Hemorrhagic Fever



- Caused by Crimean-Congo Hemorrhagic virus.
  $\Rightarrow$ 4-5 days of non-bleeding phase: headache, high fever, nausea, abdominal pain, muscle pain, diarrhea
  $\Rightarrow$ Bleeding in the eyes, the throat and in the stomach: hypotension, relative bradycardia, conjunctivitis, skin flushing and rash
  $\Rightarrow$ 10-20 days for convalescence
- No vaccine is available, and treatment is mostly supportive.

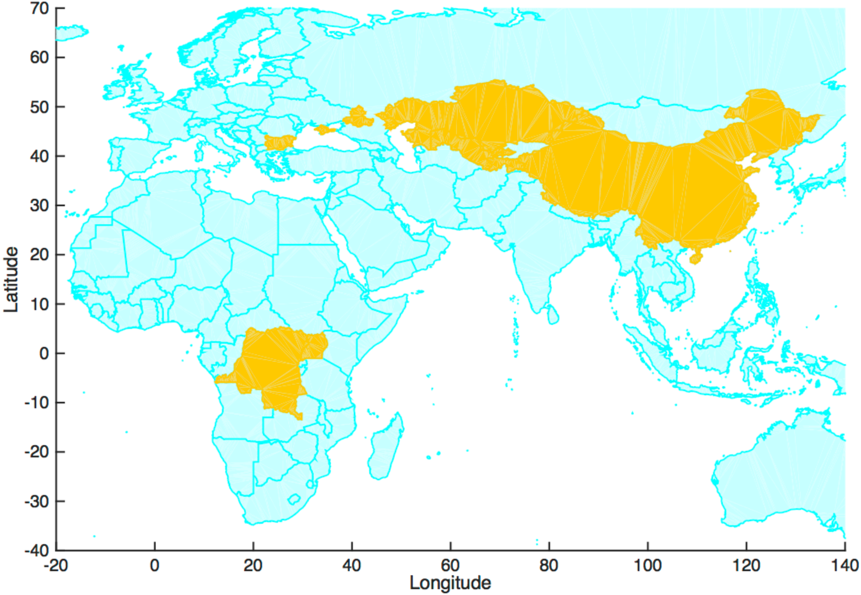
Crimean-Congo Hemorrhagic Fever (CCHF) Virus Ecology



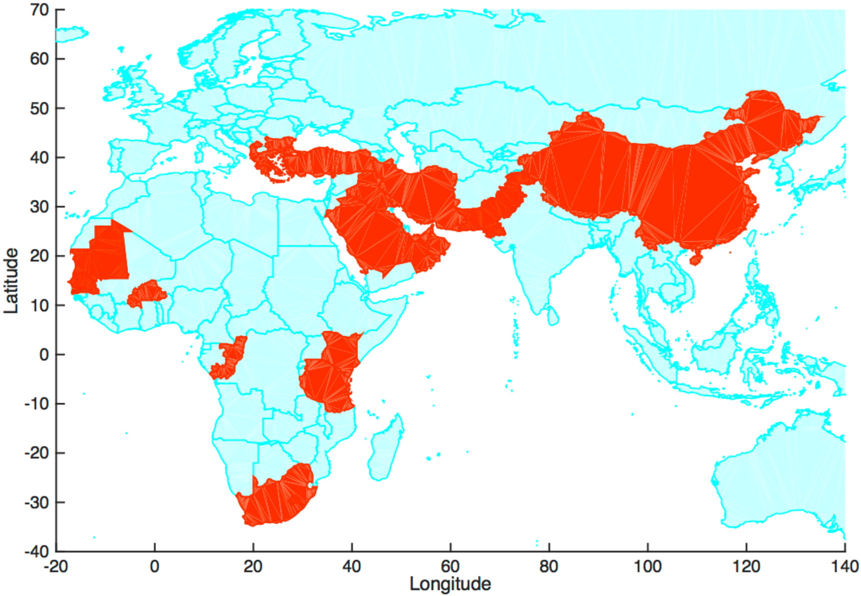
CCHF Characteristics

- CCHF is a fatal viral infection disease with the fatality rate of up to 30%.
- The virus is transmitted to human from ticks and livestock animals, as well as human.
- The great majority of incidents are from agriculture and/or husbandry:
  - Ergönül [1] reported that almost 90% of the cases around year 2007 outbreak in Turkey were farmers.
- CCHF has been found in parts of Africa, Asia, Eastern Europe and Middle East.
- The region of prevalent CCHF is growing as the following.
  - Before 1970: Soviet Union, Zaire, Uganda and China.
  - 1970-2000s: Africa, Eastern Europe, Middle East and China.

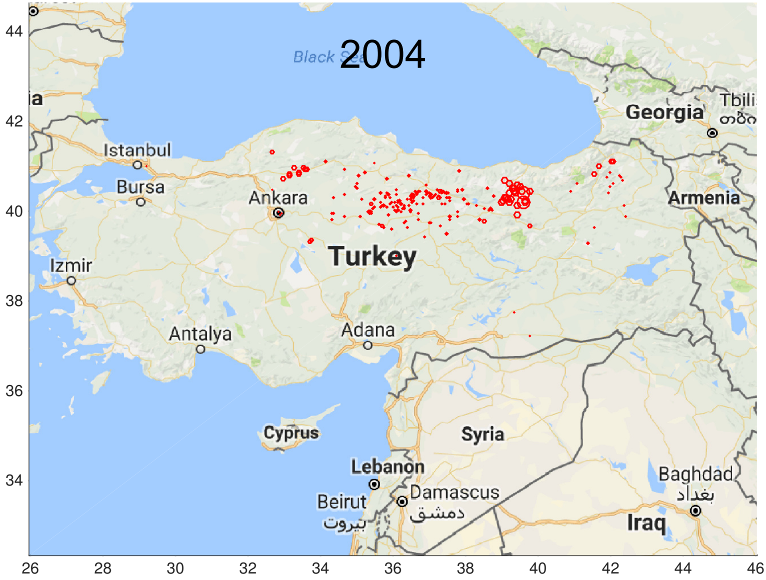
CCHF Reported before 1970



CCHF Reported between 1970-2000s



CCHF Incidents in Turkey



CCHF Incidents in Turkey

- From year 2004 to year 2012, there were 7040 CCHF incidents observed at 3514 locations in 73 cities.
- We consider CCHF incidents per 500,000 population of each city in Turkey:
  - Incidents are nonnegative integers.
  - Incidents are time dependent.
  - Incidents from neighboring locations are associated with each other.
- We analyze the incidents of 9 months from March to November in each year.
  - There is no incident during the winter season.
  - Incidents show a periodic pattern from March to November.
  - Incidents also show a nonlinear pattern over cities.

Dynamic Model for CCHF Incidents

- The CCHFs are sequentially observed at 79 months over 73 cities in Turkey.

- At a given time, incidents show a nonlinear relationship with location and take nonnegative integer values.

  $\Rightarrow$ Response Model: Poisson regression model and log-link function with the radial basis function (RBF) network.

- Some variables and parameters in the response model change by time.

  $\Rightarrow$ Transition Model: Linear or nonlinear models for state variables and time-varying parameters.

Response Model

- For a given time t, let Y_t denote the number of incidents at a location $\boldsymbol{X}_t = (\text{latitude}, \text{longitude})_t$ with $E(Y_t | \boldsymbol{X}_t) = \psi_t$.
- Then, we may model Y_t by the following generalized linear mixed model

$$
\begin{aligned}
Y_t &\sim \text{Poisson}(\psi_t) \\
L_t &= \log \psi_t + \delta_t = f_t(\boldsymbol{X}_t) + \delta_t,
\end{aligned}
$$

  where

  - $f_t(\boldsymbol{X}_t)$ is the RBF network,
  - $\delta_t \sim N(0, \sigma_t^2)$ is an independent random error,
  - L_t is a latent variable.

RBF network

- Let $\boldsymbol{X}_t^m$ denote the set of all terms of m-degree polynomial, e.g., for $\boldsymbol{X}_t = (X_{t1}, X_{t2})$, $\boldsymbol{X}_t^2 = (X_{t1}^2, X_{t1}X_{t2}, X_{t2}^2)$.
- Consider K_t knots $\boldsymbol{\mu}_t = (\boldsymbol{\mu}_{t,1}, \ldots, \boldsymbol{\mu}_{t,K_t})$ and let $z_{t,k}$ denote the distance from $\boldsymbol{X}_t$ to $\boldsymbol{\mu}_{t,k}$, e.g., $z_{t,k} = \| \boldsymbol{X}_t - \boldsymbol{\mu}_{t,k} \|$.
- Then, for a radial basis function ϕ on $z_{t,k}$ as in Holmes [2], RBF network is defined as a linear combination with smoothing penalty γ_t

$$f_t(\boldsymbol{X}_t) = \beta_{t,0} + \sum_{m=1}^{M} \boldsymbol{X}_t^m \boldsymbol{\beta}_{t,m} + \sum_{k=1}^{K_t} \phi(z_{t,k}) \beta_{t,M_b+k} = \boldsymbol{d}_t \boldsymbol{\beta}_t,$$

where

  - $\boldsymbol{d}_t = [1, X_{t1}, \ldots, X_{t2}^m, \ldots, \phi(z_{t,K_t})]$ is a design row vector,
  - $\boldsymbol{\beta}_t = [\beta_{t,0}, \beta_{t,1}, \ldots, \beta_{t,M_b}, \ldots, \beta_{t,M_b+K_t}]$ is a vector of regression coefficients, for $M_b = 1 + M(M+3)/2$.

Transition Model

- Consider normal random errors ϵ_{t_0}, ϵ_{t_1}, ϵ_{t_2} and ϵ_{t_3}.
- From time t to time $t+1$, all nuisance parameters in the response model $\boldsymbol{\lambda}_t = (\sigma_t^2, \gamma_t, K_t, \boldsymbol{\mu}_t)$ are transited by:

$$\log \sigma_{t+1}^2 = C_1 \log \sigma_t^2 + \epsilon_{t_1}, \quad \log \gamma_{t+1}^{-1} = C_2 \log \gamma_t^{-1} + \epsilon_{t_2},$$

$$\pi(K_{t+1}|K_t) = \begin{cases} 1/3, & \text{for } K_{t+1} = K_t \\ 1/3, & \text{for } K_{t+1} = K_t + 1 \\ 1/3, & \text{for } K_{t+1} = K_t - 1 \end{cases},$$

$$\boldsymbol{\mu}_{t+1} = \begin{cases} \boldsymbol{\mu}_t + \epsilon_{t_3}, & \text{for } K_{t+1} = K_t \\ (\boldsymbol{\mu}_t, \mu_{t+1,K_{t+1}}) + \epsilon_{t_3}, & \text{for } K_{t+1} = K_t + 1 \\ (\boldsymbol{\mu}_t)_{[-i]} + \epsilon_{t_3}, & \text{for } K_{t+1} = K_t - 1 \end{cases}.$$

- The latent variable is also transited by:

$$L_{t+1} = L_t + \epsilon_{t_0}.$$

Directed Acyclic Graph for Model

$\boldsymbol{\mu}_t$ knot locations

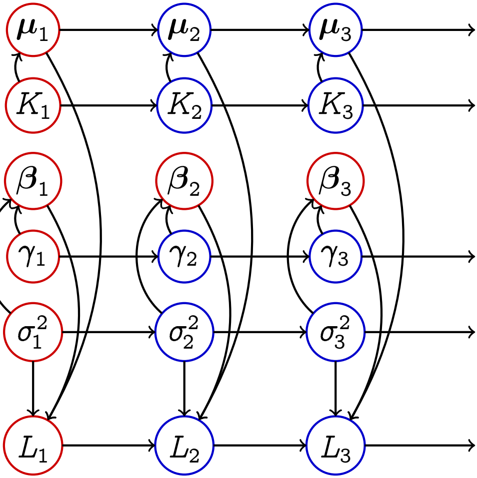
K_t no. of knots

$\boldsymbol{\beta}_t$ regression coeffs

γ_t smoothness

σ_t^2 error variance

L_t latent variable



Prior Distributions

- Only for time $t = 1$, assign prior distributions for nuisance parameters $\boldsymbol{\lambda}_1$:
  - Number of knots: $K_1 \sim$ Uniform,
  - Knot locations: $\boldsymbol{\mu}_1 | \boldsymbol{X}_1 \sim N(\boldsymbol{X}_1, \boldsymbol{I}_2)$,
  - Nuisance parameters: $\sigma_1^2 \sim IG(A_y, B_y)$, $\gamma_1 \sim G(A_\gamma, B_\gamma)$.
- For $t > 1$, the transition model projects $\boldsymbol{\lambda}_t$ based on $\boldsymbol{\lambda}_{t-1}$.
- For all times, we assign a conjugate Gaussian prior distribution for regression coefficients $\boldsymbol{\beta}_t$:

$$\begin{aligned}(\beta_{t,0}, \boldsymbol{\beta}_{t,1}^T, \ldots, \boldsymbol{\beta}_{t,M}^T) &\propto 1, \\ (\beta_{t,M_b+1}, \ldots, \beta_{t,M_b+K_t}) &\sim N\left(0, \frac{\sigma_t^2}{\gamma_t} \boldsymbol{I}_{K_t}^*\right).\end{aligned}$$

Full Conditional Distributions

- Regression coefficients:

$$\boldsymbol{\beta}_t | \cdot \sim N(\boldsymbol{V}_t^{-1}\boldsymbol{D}_t^T\boldsymbol{L}_t, \sigma_t^2\boldsymbol{V}_t^{-1}), \quad t \geq 1,$$

where $\boldsymbol{D}_t$ is a design matrix that accumulate row vectors $[\mathbf{1}, \boldsymbol{X}_t^m, \phi(z_{t,1}), \ldots, \phi(z_{t,K_t})]$ corresponding to all locations, $\boldsymbol{L}_t$ is a vector of latent variables and $\boldsymbol{V}_t = \boldsymbol{D}_t^T\boldsymbol{D}_t + \gamma_t\boldsymbol{I}^*$.

- Nuisance parameters:

$$\sigma_1^2 | \cdot \sim IG[A_y + (n+K_1)/2, \{B_y^{-1} + (\mathrm{RSS}_1 + \gamma_1\boldsymbol{\beta}_1^T\boldsymbol{I}^*\boldsymbol{\beta}_1)/2\}^{-1}],$$
$$\gamma_1 | \cdot \sim G[A_\gamma + K_1/2, \{B_\gamma^{-1} + (\boldsymbol{\beta}_1^T\boldsymbol{I}^*\boldsymbol{\beta}_1)/(2\sigma_1^2)\}^{-1}],$$

where $\mathrm{RSS}_1 = (\boldsymbol{L}_1 - \boldsymbol{D}_1\boldsymbol{\beta}_1)^T(\boldsymbol{L}_1 - \boldsymbol{D}_1\boldsymbol{\beta}_1)$.

- We use latent variables L_1 instead of Poisson responses Y_1:

$$L_1 | \cdot \propto p(Y_1 | L_1) N(\boldsymbol{d}_1\boldsymbol{\beta}_1, \sigma_1^2)$$

Projection by Transition Model

- For $t > 1$, we can generate $\boldsymbol{\lambda}_t = (\gamma_t, \sigma_t^2, K_t, \boldsymbol{\mu}_t)$ with a data assimilation method.

- Defining $\boldsymbol{\lambda}_0$ by the hyperparameters, transition model bring the target distribution for $\boldsymbol{\lambda}_t$:

$$p\{\boldsymbol{\lambda}_{0:t} \big| \boldsymbol{X}_{1:t}, \boldsymbol{L}_{1:t}\}$$

Data Assimilation Methods

- Kalman Filters:
  - Extended Kalman Filter (EKF)
  - Ensemble Kalman Filter (EnKF)
- Particle Filters:
  - Sequential Importance Sampling (SIS)
  - Rejection Control (RC)
  - Dynamically Weighted Importance Sampling (DWIS)

Kalman Filter

- Dynamic Model: Y_t is observed sequentially on state and time-varying parameters X_t.

$$\begin{aligned} Y_t &= B_t X_t + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_e), \\ X_t &= A_t X_{t-1} + U_t, \quad U_t \sim N(0, \Sigma_u). \end{aligned}$$

- Forecast (predict): Project state variable,

$$\widetilde{X}_t = A_t \widehat{X}_{t-1} \sim N(m_t, S_t).$$

- Update (match): Correct the projection

$$\widehat{X}_t = \widetilde{X}_t + K_t(Y_t - B_t \widetilde{X}_t) \sim N(\mu_t, \Sigma_t),$$

where

$$\begin{aligned} K_t &= S_t B_t^T (B_t S_t B_t^T + \Sigma_e)^{-1}, \\ \mu_t &= E(\widetilde{X}_t | Y_t = B_t m_t) = m_t + K_t(Y_t - B_t m_t), \\ \Sigma_t &= \text{cov}(\widetilde{X}_t | \widetilde{Y}_t) = S_t - K_t B_t S_t. \end{aligned}$$

Particle Filter

- Dynamic Model: We relax the linearity and normality by considering nonlinear f_t and g_t and non-Gaussian ϵ_t and U_t

$$\begin{aligned} Y_t &= f_t(X_t) + \epsilon_t, \\ X_t &= g_t(X_{t-1}) + U_t. \end{aligned}$$

- Approximate the target (filtering) distribution $\pi(X_{1:t}|Y_{1:t})$ by a weighted set of particles, $\{\widetilde{X}_{1:t}^{(1)}, \dots, \widetilde{X}_{1:t}^{(N_t)}\}$, through importance sampling for a stream of particle i:

$$\begin{aligned} \text{Sample :} \quad & \widetilde{X}_{1:t}^{(i)} \text{ from } q(X_{1:t}|Y_{1:t}), \\ \text{Weight :} \quad & w_t^{(i)} = \frac{\pi(X_{1:t}|Y_{1:t})}{q(X_{1:t}|Y_{1:t})}, \end{aligned}$$

where q is a proposal distribution.

Sampling Methods for Particle Filters

- Sequential Importance Sampling:
  - Apply the decomposition to the target density π and proposal density q.

$$
\begin{aligned}
\pi(X_{1:t}|\,Y_{1:t}) &= \pi(X_t|X_{1:t-1},\,Y_{1:t})\,\pi(X_{1:t-1}|\,Y_{1:t}) \\
&= \pi(X_1|\,Y_1)\prod_{k=1}^{t}\pi(X_k|X_{1:k-1},\,Y_{1:k})
\end{aligned}
$$

  - Collects only one-step ahead samples with established previous samples.

- Rejection Control: If a stream of samples has less weight than the threshold, send it back to $t = 1$.

- Dynamically Weighted Importance Sampling: At each time investigate the streams and then prune and enrich them.

Dynamically Weighted Particle Filter

- For CCHF, we need to sample $\boldsymbol{\lambda}_{1:t}$ from the target distribution $p\left\{\boldsymbol{\lambda}_{0:t} \middle| \boldsymbol{X}_{1:t}, \boldsymbol{L}_{1:t}\right\}$.
- Assuming a Markovian structure:

$$p\{\boldsymbol{\lambda}_{0:t}|\boldsymbol{X}_{1:t},\boldsymbol{L}_{1:t}\} \propto p\{\boldsymbol{\lambda}_{0:t-1}|\boldsymbol{X}_{1:t-1},\boldsymbol{L}_{1:t-1}\}\, p(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1},\boldsymbol{X}_t,\boldsymbol{L}_t),$$

where $p(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1},\boldsymbol{X}_t,\boldsymbol{L}_t)$ is the marginalized over $\boldsymbol{\beta}_t$

$$\begin{aligned} p(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1},\boldsymbol{X}_t,\boldsymbol{L}_t) &\propto \int p(\boldsymbol{X}_t,\boldsymbol{L}_t|\boldsymbol{\lambda}_t,\boldsymbol{\beta}_t)\pi(\boldsymbol{\beta}_t|\boldsymbol{\lambda}_t)\pi(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1})\,d\boldsymbol{\beta}_t \\ &\propto \pi(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1})\,\exp\left\{-\mathrm{RSS}_t/2\sigma_t^2\right\}. \end{aligned}$$

- As an effective sampling scheme, as Liang [3] and Ryu [4] used, DWPF is a combination of SIS and DWIS algorithms.
- DWIS consists of dynamic weighting and population control scheme.

Dynamic Weighting

1. Update the weight upper and lower bounds:

$$(W_{lt}, W_{ut}) \leftarrow \begin{cases} (W_{lt}/a, W_{ut}/a), & \text{if } N_t < N_{low}, \\ (aW_{lt}, aW_{ut}), & \text{if } N_t > N_{up}, \\ (W_{lt}, W_{ut}), & \text{otherwise.} \end{cases}$$

for $i = 1, \ldots, N_t$ do

2. Draw $\boldsymbol{\lambda}_*^{(i)}$ from a proposal distribution $q(\boldsymbol{\lambda}|\boldsymbol{\lambda}_t^{(i)})$:

$$r_t^{(i)} = w_t^{(i)} \frac{p(\boldsymbol{\lambda}_*^{(i)}|\boldsymbol{\lambda}_{0:t-1}^{(i)})\, q(\boldsymbol{\lambda}_t^{(i)}|\boldsymbol{\lambda}_*^{(i)})}{p(\boldsymbol{\lambda}_t^{(i)}|\boldsymbol{\lambda}_{0:t-1}^{(i)})\, q(\boldsymbol{\lambda}_*^{(i)}|\boldsymbol{\lambda}_t^{(i)})}.$$

3. Update the weight with $\delta_t = \left[c_1 + c_2 W_{u,t}^{1+c_3}\right]^{-1}$:

$$\hat{\boldsymbol{\lambda}}_t^{(i)} = \boldsymbol{\lambda}_*^{(i)}, \quad \hat{w}_t^{(i)} = (1+\delta_t) r_t^{(i)}$$

endfor

Population Control Scheme

- Adaptive pruned-enriched population control:

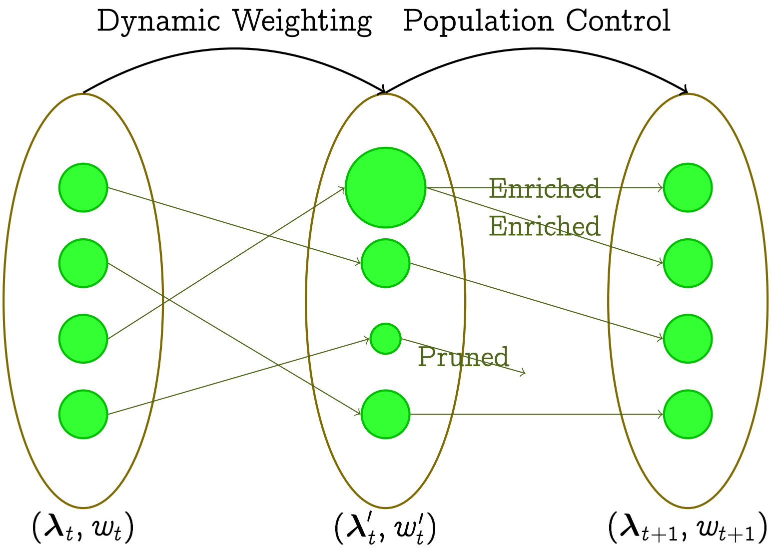
$$(\widehat{\boldsymbol{\lambda}}_t^{(i)}, \widehat{w}_t^{(i)}) \ \Rightarrow \ (\boldsymbol{\lambda}_t^{(i')}, w_t^{(i')})$$

- Algorithm: for $i = 1, \ldots, N_t$
  1. If $\widehat{w}_t^{(i)} < W_{lt}$, prune with probability $1 - \frac{\widehat{w}_t^{(i)}}{W_{lt}}$, or keep the particle w.p. $\frac{\widehat{w}_t^{(i)}}{W_{lt}}$ and set the weight to W_{lt}.
  2. If $\widehat{w}_t^{(i)} > W_{ut}$, enrich the particle with $h_t = \lfloor \frac{\widehat{w}_t^{(i)}}{W_{ut}} + 1 \rfloor$ replications of particles and the adjusted weight $\frac{\widehat{w}_t^{(i)}}{h_t}$.
- Assess: If $N_t' \notin (N_{\min}, N_{\max})$, adjust weight bounds:

$$(W_{lt}, W_{ut}) \leftarrow \begin{cases} (aW_{lt}, aW_{ut}), & \text{if } N_t' > N_{\max}, \\ (W_{lt}/a, W_{ut}/a), & \text{if } N_t' < N_{\min}, \end{cases}$$

and keep the process until $N_t' \in (N_{\min}, N_{\max})$.

Dynamically Weighted Importance Sampling



DWPF Procedure - Stage 1

1. Sample: Sample $\widehat{\boldsymbol{\lambda}}_1^{(i)}$ from $p(\boldsymbol{\lambda}_1|\boldsymbol{X}_1, \boldsymbol{L}_1)$, and set $\widehat{w}_1^{(i)} = 1$ for $i = 1, \ldots, N_0$. These form the initial population $(\widehat{\boldsymbol{\lambda}}_1, \widehat{\boldsymbol{w}}_1)$, and N_0 is called the initial population size.
2. DWIS: Generate $(\boldsymbol{\lambda}_1, \boldsymbol{w}_1)$ from $(\widehat{\boldsymbol{\lambda}}_1, \widehat{\boldsymbol{w}}_1)$ using DWIS, with $p(\boldsymbol{\lambda}_1|\boldsymbol{X}_1, \boldsymbol{L}_1)$ as the target distribution.

DWPF Procedure - Stage 2

1. Extrapolation: Generate $\widehat{\boldsymbol{\lambda}}_2^{(i)}$ from $\boldsymbol{\lambda}_1^{(i)}$, with the extrapolation operator $q(\boldsymbol{\lambda}_2 | \boldsymbol{\lambda}_1^{(i)}, \boldsymbol{X}_{1:2}, \boldsymbol{L}_{1:2})$, and set

$$\widehat{w}_2^{(i)} = w_1^{(i)} \frac{p(\boldsymbol{\lambda}_1^{(i)}, \widehat{\boldsymbol{\lambda}}_2^{(i)} | \boldsymbol{X}_{1:2}, \boldsymbol{L}_{1:2})}{p(\lambda_1^{(i)} | L_1) q(\widehat{\boldsymbol{\lambda}}_2^{(i)} | \boldsymbol{\lambda}_1^{(i)}, \boldsymbol{X}_{1:2}, \boldsymbol{L}_{1:2})}$$

   for each $i = 1, 2, \ldots, N_1$.

2. DWIS: Generate $(\boldsymbol{\lambda}_2, \boldsymbol{w}_2)$ from $(\widehat{\boldsymbol{\lambda}}_2, \widehat{\boldsymbol{w}}_2)$ using DWIS, with $p(\boldsymbol{\lambda}_{1:2} | \boldsymbol{X}_{1:2}, \boldsymbol{L}_{1:2})$ as the target distribution.

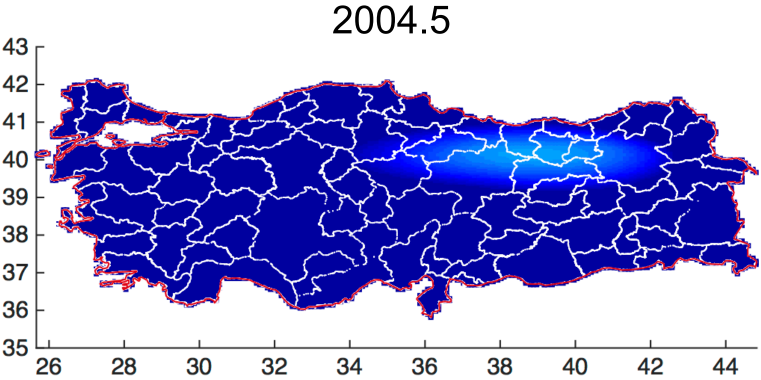
1. Extrapolation: Generate $\widehat{\boldsymbol{\lambda}}_t^{(i)}$ from $\boldsymbol{\lambda}_{t-1}^{(i)}$, with the extrapolation operator $q(\boldsymbol{\lambda}_t | \boldsymbol{\lambda}_{1:t-1}^{(i)}, \boldsymbol{X}_{1:t}, \boldsymbol{L}_{1:t})$ and set

$$\widehat{w}_t^{(i)} = w_{t-1}^{(i)} \frac{p(\boldsymbol{\lambda}_{1:t-1}^{(i)}, \widehat{\boldsymbol{\lambda}}_t^{(i)} | \boldsymbol{X}_{1:t}, \boldsymbol{L}_{1:t})}{p(\boldsymbol{\lambda}_{1:t-1}^{(i)} | \boldsymbol{X}_{1:t-1}, \boldsymbol{L}_{1:t-1}) q(\widehat{\boldsymbol{\lambda}}_t^{(i)} | \boldsymbol{\lambda}_{1:t-1}^{(i)}, \boldsymbol{X}_{1:t}, \boldsymbol{L}_{1:t})}$$

   for each $i = 1, 2, \ldots, N_{t-1}$.

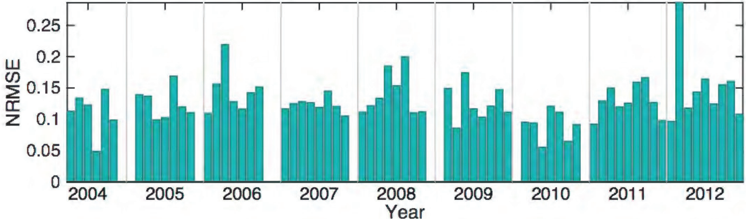
2. DWIS: Generate $(\boldsymbol{\lambda}_t, \boldsymbol{w}_t)$ from $(\widehat{\boldsymbol{\lambda}}_t, \widehat{\boldsymbol{w}}_t)$ using DWIS, with $p(\boldsymbol{\lambda}_{1:t} | \boldsymbol{X}_{1:t}, \boldsymbol{L}_{1:t})$ as the target distribution.

Predicted Prevalence of CCHF

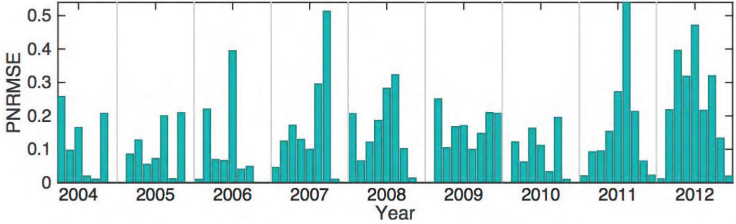


Prediction Performance

$$\text{NRMSE}_t = \sqrt{\frac{1}{73}\sum_{i=1}^{73}(Y_{it} - \widehat{\psi}_{it})^2}/\text{range}(Y_{1t}, \ldots, Y_{73t})$$



$$\text{PNRMSE}_t = \sqrt{\frac{1}{23}\sum_{i=1}^{23}(T_{it} - \widehat{\varphi}_{it})^2}/\text{range}(Y_{1t}, \ldots, Y_{73t})$$



Concluding Remarks

- The proposed model is suitable for modeling and mapping relative risk of CCHF incidents in cities of Turkey.
- It also delivers the results in a timely manner using an effective computation method DWPF.
- Estimated CCHF propagation reveals:
  - Birds and population of wild pigs in the region are suspected for a disease to move in the direction of north to south.
  - Mostly farmers, interact and share the same habitant and living space with those animals.
  - Due to economic reasons, farmers give up on precaution and do not pay attention health information.
- A timely control of the tick population and immunization of the livestock animals are also highly recommended.

References

Ergonul, O., and Whitehouse, C. A.
Introduction. Crimean Congo Hemorrhagic Fever: A Global Perspective.
Springer, Dordrecht, the Netherlands, 2007.

Holmes, C. C., and Mallick, B. K.
Bayesian radial basis functions of variable dimension.
Neural Computation 10, 5 (1998), 1217–1233.

Liang, F.
Dynamically weighted importance sampling in monte carlo computation.
Journal of the American Statistical Association 97, 459 (2002), 807–821.

Ryu, D., Liang, F., and Mallick, B. K.
Sea surface temperature modeling using radial basis function networks with a dynamically weighted particle filter.
Journal of the American Statistical Association 108, 501 (2013), 111–123.