Spatio-temporal modelling of mosquitoes vector and its environmental drivers in Hong Kong

StanConnect 2022: Stan through Space and Time

Stan Yip

The Hong Kong Polytechnic University

31st October 2022

1. Introduction
2. Data
3. Statistical model
4. Results
5. Extension

Why Stan?

- Modularisation of complex structure is a solution for complex problem
- Easy to implement
- Easy to interpret
- Quicker development time (but slightly slower in running time)
- Language agnostic - interface for both Python and R

Aedes albopictus (1)

- An estimate shows that a total of 3.97 billion people living in 128 countries are at risk (Brady et al., 2012).
- Eggs can withstand drought very well
- Breed in small containers such as cans, discarded tyres, tree holes, rock pools and bamboo ends holding small amount of water
- Adults rest in shrubby area

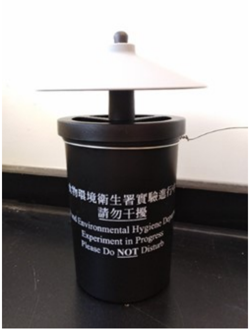


Figure 1: Oviposition Trap (Ovitrap)

Aedes albopictus (2)

- Adult has a white stripe on the dorsal surface of thorax and bands on legs
- Although exophilic, adults may also enter into houses
- Day biter (Peak : within 2 hours after dawn and before sunset)
- Weak flier (about 100m)
- Vector of dengue fever

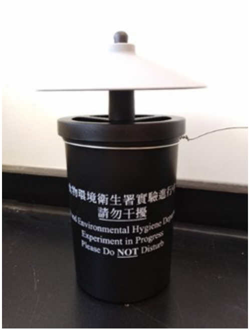


Figure 2: Oviposition Trap (Ovitrap)

Ovitrap/gravidtrap indices

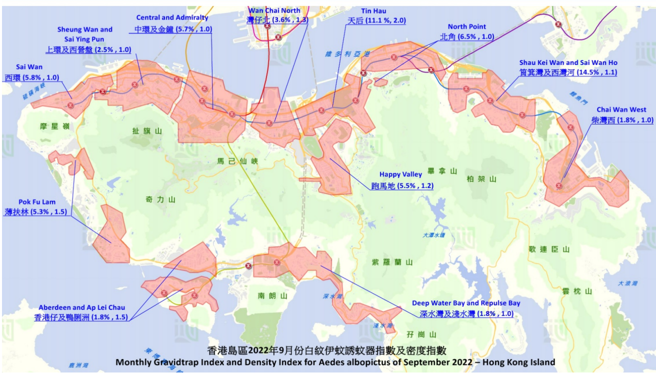
To study the distribution of Aedes albopictus, Gravidtraps are set in selected areas throughout the territory for monitoring the breeding of these mosquitoes. The percentage of the Gravidtraps found positive with breeding of these mosquitoes gives the Gravidtrap Index for Aedes albopictus. The value of Gravidtrap Index for Aedes albopictus indicates the extensiveness of the breeding of the vectors.

$$\text{Gravidtrap index} = \frac{\text{Number of Aedes positive Gravidtraps retrieved}}{\text{Total number of Gravidtraps retrieved from a particular area}}$$

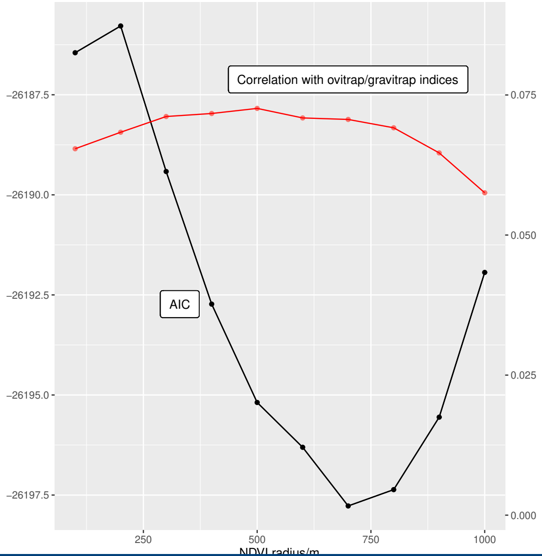
Data inventory

- Ovitrap/gravitrap
- Satellite data
- Streams and rivers shapefiles
- Local meteorological stations
- Synoptic scale climate data

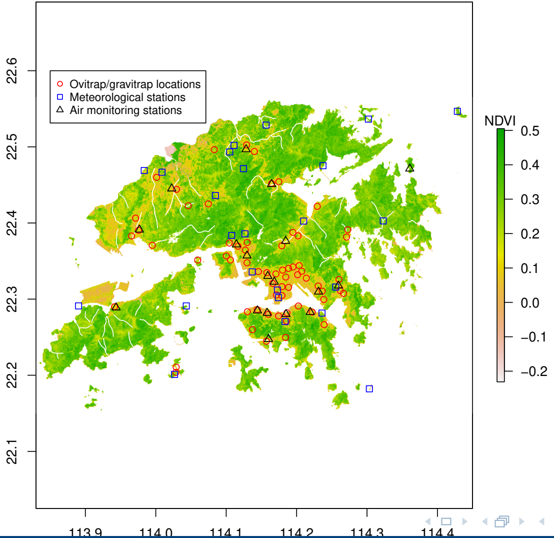
Hong Kong Island



Geography


Correlation with ovitrap/gravitrap indices
AIC
−26187.5
−26190.0
−26192.5
−26195.0
−26197.5
0.075
0.050
0.025
0.000
250
500
750
1000
NDVI radius/m

Geography



Ovitrap/gravitrap locations
Meteorological stations
Air monitoring stations
NDVI
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
22.6
22.5
22.4
22.3
22.2
22.1
113.9
114.0
114.1
114.2
114.3
114.4

Data (ovitrap/satellite)

- Oviposition trap (ovitrap) index is obtained from the Food and Environmental Hygiene Department of the Hong Kong government between January 2010 to March 2020. Gravitrap includes a sticky trap to collect adult mosquitoes with the same black plastic container of approximately 600ml.
- A 30m resolution normalized difference vegetation index (NDVI) from an environmental raster provided by Morgan and Guénard (2019) is used as a proxy of vegetation derived from satellite imagery. The NDVI values range from -1 to 1 with higher values correspond to denser vegetation. Mean NDVI are computed for radii from 100m to 2km surrounding the ovitrap/gravitrap survey areas' centroids.

Data (2)

- A geospatial vector data representing drainage provided by the Hong Kong Lands Department is used for calculating the shortest distances to rivers and streams from each ovitrap/gravitrap site.
- Surface meteorological measurements of air temperature and rainfall are recorded by the Hong Kong Observatory from the chosen 25 observing stations with elevation lower than 200m.
- The monthly Niño1+2, Niño3, Niño3.4 and Niño4 sea surface temperature (SST) indices are aggregated from the United States National Oceanic and Atmospheric Administration (NOAA) 1/4° daily optimum interpolation SST (Huang et al., 2021) are obtained from NOAA Climate Prediction Center.

Beta distribution

Consider a beta distribution parametrisation with a mean parameter μ and a shape parameter ϕ (Ferrari and Cribari-Neto, 2004),

$$f_{\text{Beta}}\left(y|\mu,\phi\right) = \frac{y^{(\mu\phi-1)}\left(1-y\right)^{[(1-\mu)\phi-1]}}{B\left(\mu\phi,\left(1-\mu\right)\phi\right)}, \tag{1}$$

where $B(\cdot)$ is the beta function. Note that the beta distribution has a support of $y_i \in (0,1)$ which does not include zero and one.

Beta distribution

The following commonly used transformation is applied to the ovitrap/gravitrap index (Smithson and Verkuilen, 2006):

$$y_i^* = \left(y_i\left(n-1\right)+0.5\right)/n, \tag{2}$$

where n is the number of ovitraps/gravitraps in a survey area. Regardless of the estimation method used, all parameters in a Matèrn family of correlation function, where exponential correlation is a special case of it, cannot be estimated consistently from observed data (Zhang, 2004).

Regression model

$$g(\mu_{st}) = X\beta + I_{\text{summer}}(t)\omega(s) + \epsilon,$$

where $g(\mu) = logit(\mu) = \log(\mu/(1-\mu))$, $\omega \sim N(\mathbf{0}, \Sigma)$, The covariance matrix $\Sigma_{ij} = \sigma_\omega^2 \exp(-\phi_s d(i,j))$, $d(i,j)$ is the geodesic distance between two locations.

Regression model

```
transformed parameters {
vector[N] mu;
     for (i in 1:N) {
          mu[i] = inv_logit(X[i,] * betas +
          summer[i] * sigmao * omega[areano[i]] +
          sigma * epsilon[i]);
     }
}

model {
     y ~ beta_proportion(mu,kappa);^^I
     kappa ~ gamma(2,1);
     sigmao ~ gamma(2,1);
     sigmao ~ gamma(2,1);
     omega ~ multi_normal_cholesky(rep0,L);
     epsilon ~ normal(0,1);
}
```

1. Introduction
2. Data
3. Statistical model
4. **Results**
5. Extension

Model

Model	looic	pLOO
(A) Gaussian	-18836.7	17.9
(B) Beta	-26499.0	12.7
(C) Beta + Seasonality	-29405.5	28.2
(D) Beta + Seasonality + Spatial	-29701.2	107.9

Table 1: Model performance by the LOO information criterion (looic, Vehtari et al. (2017)), where pLOO is the estimated effective number of parameters of the model.

Model

Table 2: Parameter estimates of RR for the Model D

		Credible Interval	
	RR	2.5%	97.5%
ndvi700m(0.01unit)	1.0255	1.0193	1.0321
t	0.9987	0.9981	0.9993
distw	0.9843	0.9787	0.9901
lag (Temp, 1)	1.0346	1.017	1.0523
lag (Rainfall, 1) (100mm)	1.0569	1.0428	1.0714
lag (Ozone, 1) (10ppb)	1.0161	0.9991	1.0326
lag (Ozone, 2) (10ppb)	1.0346	1.0183	1.0515
lag (Rainfall, 1) $\times$ (ndvi700m $>$ 0.16)	0.9727	0.9493	0.9957
lag (Temp, 1) $\times$ (ndvi700m $>$ 0.16)	0.9989	0.9955	1.0023
Ovitrap	0.8794	0.8242	0.9394
lag (Niño4, 1) (10ppb)	0.963	0.9287	0.9991
lag (Niño12, 1) (10ppb)	0.9828	0.9636	1.0018

Hurdle model

- 46.3% zeros, many zeros but no one.

Hong Kong Island (Map)														
District	Survey Area		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Central/ Western	Central and Admiralty	AGI	0.0%	0.0%	0.0%	3.8%	2.9%	15.5%	6.5%	5.7%	5.7%	1.0%		
		ADI	N/A	N/A	N/A	1.0	2.3	2.0	1.1	1.0	1.0	1.0		
	Sheung Wan and Sai Ying Pun	AGI	0.0%	0.0%	0.0%	0.0%	9.0%	31.3%	22.2%	10.3%	2.5%			
		ADI	N/A	N/A	N/A	N/A	1.1	2.3	1.5	1.6	1.0			
	Sai Wan	AGI	0.0%	0.0%	0.0%	0.9%	23.4%	31.1%	8.9%	11.2%	5.8%			
		ADI	N/A	N/A	N/A	1.0	1.7	1.3	1.6	1.3	1.0			
Eastern	Chai Wan West	AGI	0.0%	0.0%	0.0%	2.6%	9.6%	15.8%	5.4%	4.4%	1.8%	0.9%		
		ADI	N/A	N/A	N/A	1.0	1.0	1.0	1.0	1.0	1.0	1.0		
	Shau Kei Wan and Sai Wan Ho	AGI	1.8%	2.5%	1.7%	5.1%	34.8%	26.3%	11.7%	6.8%	14.5%	8.1%		
		ADI	1.0	1.0	1.0	1.0	1.7	1.5	1.1	1.0	1.1	1.2		
	North Point	AGI	0.0%	0.0%	0.0%	1.8%	13.6%	16.4%	8.5%	5.6%	6.5%	2.8%		
		ADI	N/A	N/A	N/A	1.0	1.2	1.4	1.1	1.0	1.0	1.0		

Hurdle model and other extension

Andrew Heiss (Georgia State University) blogpost suggested the following:

- A logistic regression model that predicts if an outcome is 0 or not, defined by α.
- A beta regression model that predicts if an outcome is between 0 and 1 if its not zero, defined by μ and ϕ.
- Regression on ϕ is also worth looking at.

Brady, O. J., Gething, P. W., Bhatt, S., Messina, J. P., Brownstein, J. S., Hoen, A. G., Moyes, C. L., Farlow, A. W., Scott, T. W., and Hay, S. I. (2012). Refining the global spatial limits of dengue virus transmission by evidence-based consensus.

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7):799–815.

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T., and Zhang, H.-M. (2021). Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1. *Journal of Climate*, 34(8):2923–2939.

Morgan, B. and Guénard, B. (2019). New 30 m resolution hong kong climate, vegetation, and topography rasters indicate greater spatial variation than global grids within an urban mosaic. *Earth System Science Data*, 11(3):1083–1098.

Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, 11(1):54.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27(5):1413–1432.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.