UNIVERSITY OF TWENTE.

Space-time geostatistical modelling of malaria risk

DR. FRANK BADU OSEI
ASSISTANT PROFESSOR
UNIVERSITY OF TWENTE

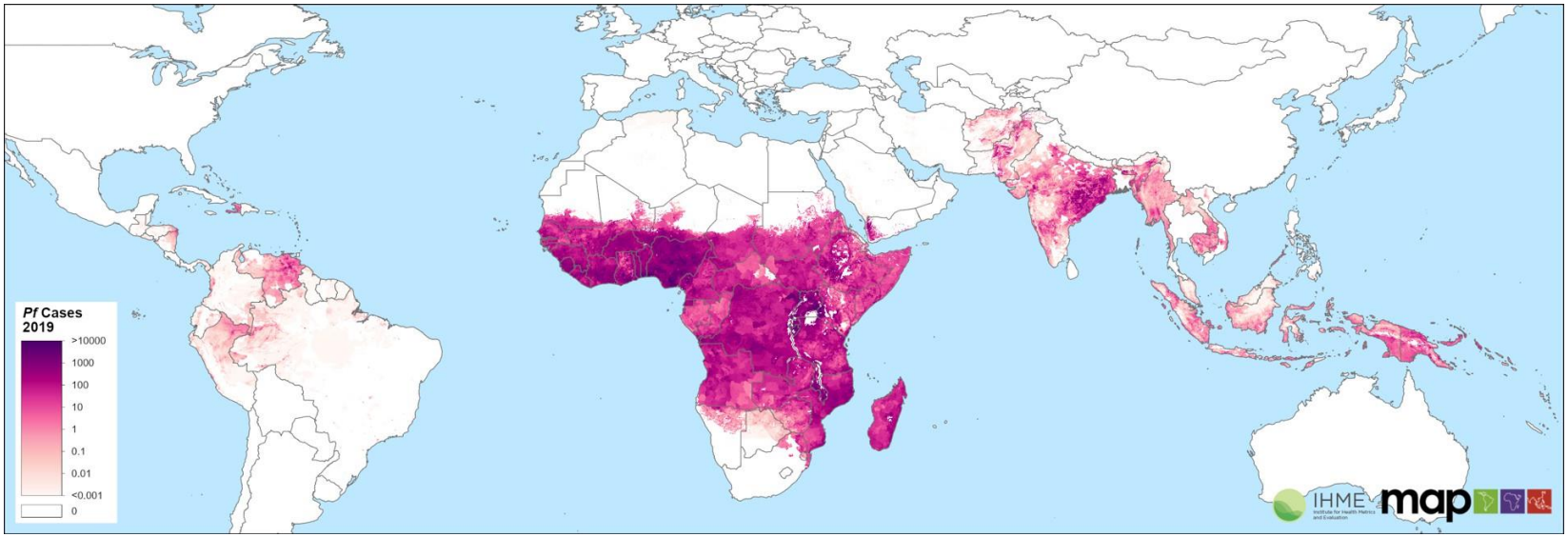
ITC FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

SPACE-TIME GEOSTATISTICAL MODELLING AND MAPPING OF OF MALARIA RISK

Global malaria picture

Malaria for instance is a global challenge which continuous to affect deprived communities especially in sub-Saharan Africa

PLASMODIUM FALCIPARUM CLINICAL CASES IN ALL AGE GROUPS IN 2019



Source: https://malariaatlas.org/trends/region/MAP/GLOBAL

MALARIA AS A GLOBAL CHALLENGE

- Maps have often been developed based on point reference data from demographic and health survey data
  - (Fobil et al, 2012)
  - (Yankson et al, 2019)
  - (Gemperli et al, 2006)

- Such data are however not routine, hence updating such maps is not feasible within shorter time periods
- They fail to examine the temporal trends

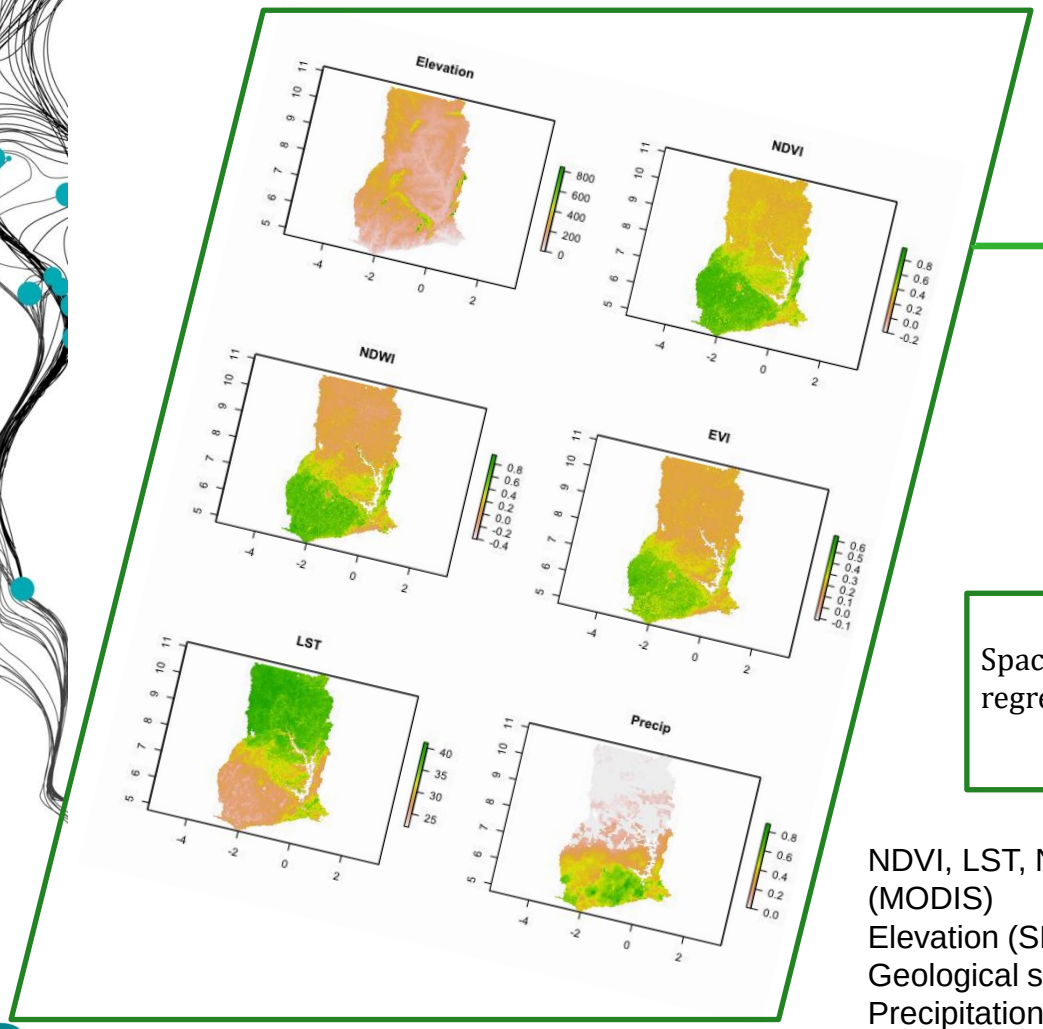
DHIMS DATA FOR SURVEILLANCE

- Routine surveillance managed through health information manage systems can rather be helpful
- Despite such data are routinely collected, they have poor spatial and temporal resolution.
- For instance the
  - **DHIMS** is currently deployed by many developing countries to store and manage aggregated morbidity data
  - Accessible spatial resolution is at the district levels
  - Accessible temporal resolution is monthly

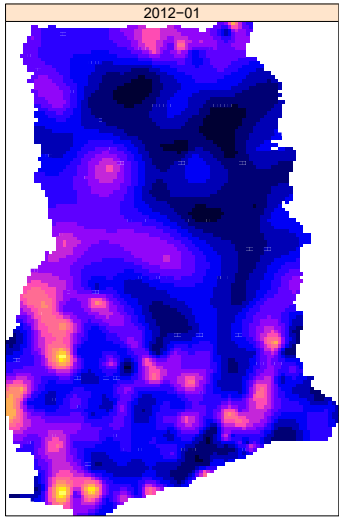
LOCAL ACTION GAPS

- Malaria
  - They remain a major public health burden in developing countries, especially in sub-Saharan Africa
  - Most studies
    - either focus on the spatial patterns at a particular point in time
    - or the temporal patterns for an entire geographic area
  - Why?
    - data challenges and/or
    - unavailable easy to implement statistical methods

STATIC AND DYNAMIC REMOTELY SENSED COVARIATES (1 KM X 1 KM)



Space-time Poisson regression kriging



NDVI, LST, NDWI, EVI (MODIS)
Elevation (SRTM, US Geological survey)
Precipitation (Climate hazard group infrared precipitation)

THE POISSON DATA GENERATING PROCESS

- We consider malaria counts and population data
  - $\{y_{it}, n_{it}\}$
- As spatio-temporal outcomes of malaria and population data disaggregated by
  - districts $i = 1, \ldots, m = 216$
  - months $t = 1, \ldots, T = 55$ (from Jan 2012 to July 2016)

- Such sampling models are typically realizations from the Poisson process
  - $y_{it}|r_{it} \sim Poi(\lambda_{it})$
  - $\log \lambda_{it} = \log n_{it} + \log r_{it}$
  - $\log r_{it} = \beta_0 + \sum_{k=1}^{K} \beta_{kt} x_{ikt} + \sum_{p=1}^{P} \omega_k z_{ik}$
  - x_{ikt} represent the dynamic covariates/predictors
  - z_{ik} represent the static covariates/predictors

STANDARD SEMIVARIANCE ESTIMATION

- Here, the maximum likelihood estimate of the residual risk is
  - $\tau_{it} = \frac{(y_{it} - \lambda_{it})}{n_{it}}$
- Assuming the risk was not a ratio variable, then
- The spatio-temporal variogram represents the semi-variance between any pair of risks estimates which are separated by spatial lag h and/or temporal lag u:
  - $\gamma(\mathrm{h}, \mathrm{u}) = \frac{1}{2}\mathbb{E}\left(\tau_{it} - \tau_{i+h,t+u}\right)^2$
  - $\gamma^*(\mathrm{h}, \mathrm{u}) = \frac{1}{2}\sum_{N(\mathrm{h},\mathrm{u})}\left[\tau_{it} - \tau_{i+h,t+u}\right]^2$

MODIFIED SEMI-VARIANCE ESTIMATION

- The standard nonparametric geostatistical methods for semivariogram estimation used for continuous data are not directly applicable for the current model due to the
  - Heteroscedasticity (varying variance inversely proportional to the population)
  - non-stationarity of the process of counts

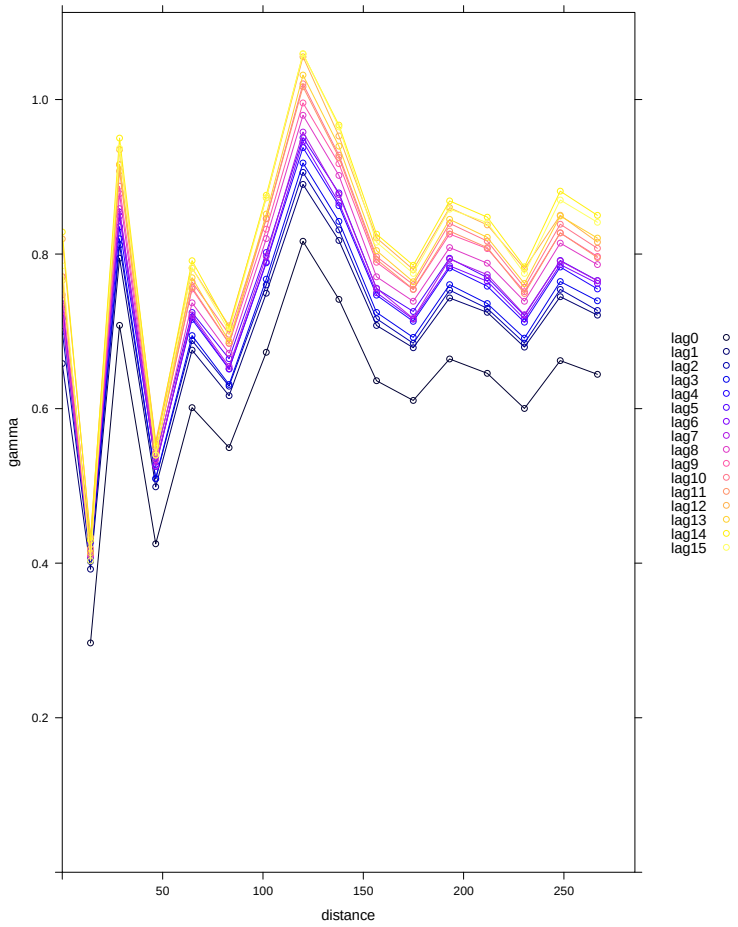
- $$\gamma^*(\mathrm{h},\mathrm{u}) = \frac{\sum_{N(\mathrm{h},\mathrm{u})}\left\{\frac{n_{it}\cdot n_{i+\mathrm{h},t+\mathrm{u}}}{n_{it}+n_{i+h,t+u}}\left[\tau_{it}-\tau_{i+\mathrm{h},t+\mathrm{u}}\right]^2-\bar{\tau}\right\}}{2\sum_{N(\mathrm{h},\mathrm{u})}\frac{n_{it}\cdot n_{i+\mathrm{h},t+\mathrm{u}}}{n_{it}+n_{i+\mathrm{h},t+\mathrm{u}}}}$$

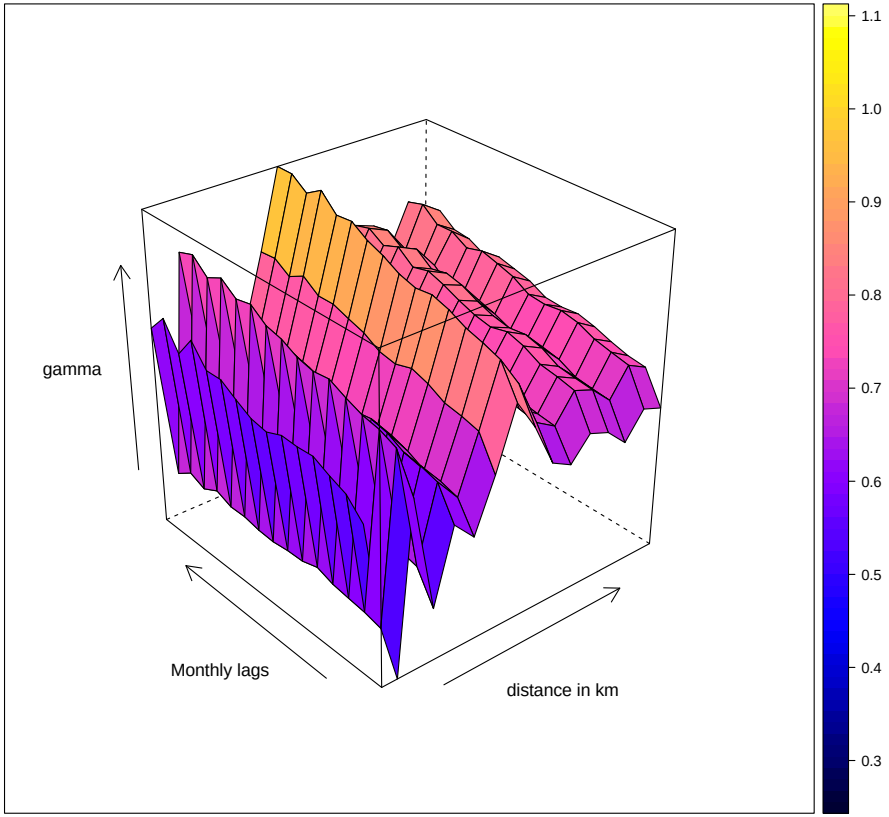
- Where the population weighted average risk is

- $$\bar{\tau} = \frac{\sum n_{it}\cdot\tau_{it}}{\sum n_{it}}$$

EMPIRICAL SPACE-TIME SEMI-VARIOGRAM

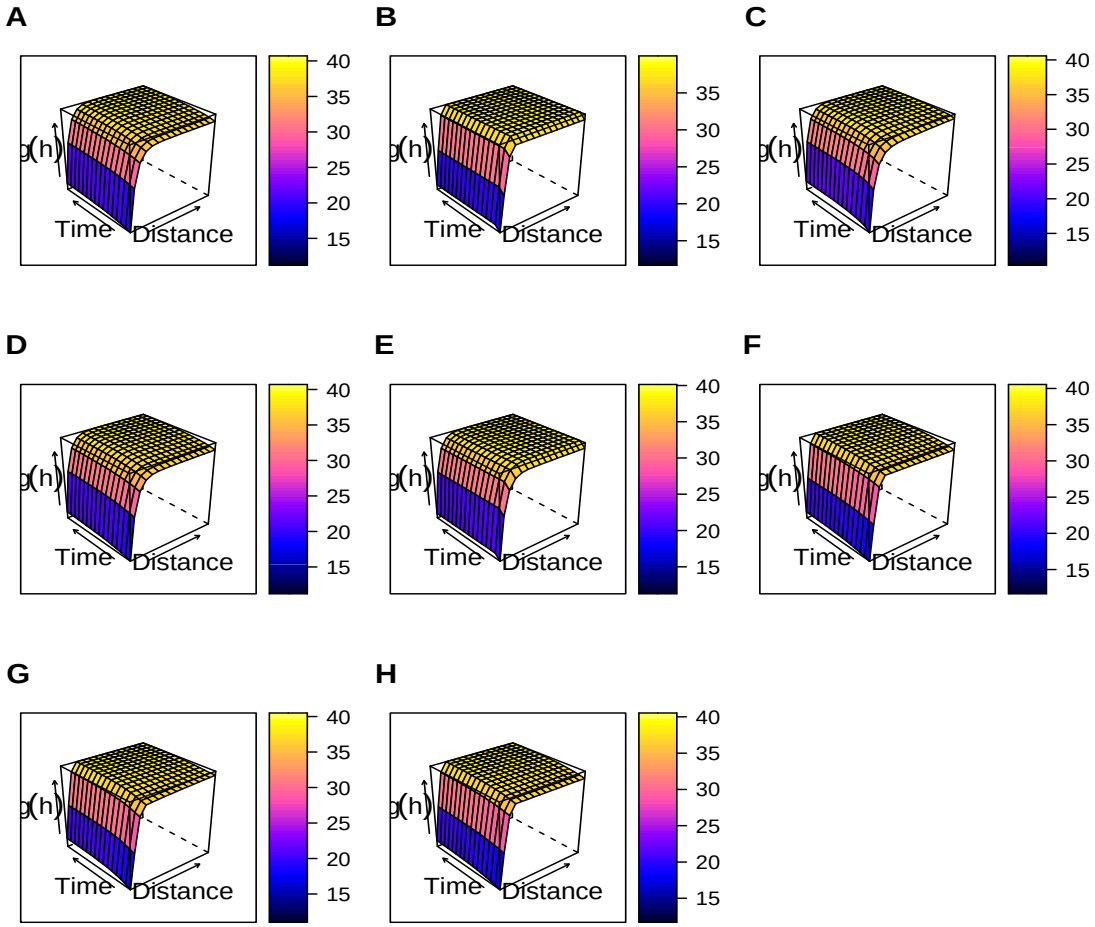
The space-time variogram clearly shows indication of space-time correlation





SPACE-TIME SEMI-VARIOGRAM MODELS

The combination of exponential, spherical and Gaussian space-time variogram models



SPACE-TIME SEMI-VARIOGRAM MODEL PARAMETERS AND ASSOCIATED SSErr

	Estimates							
Parameters	EEE	SSS	ESE	EES	ESS	SES	SSE	SEE
Parameters	A	B	C	D	E	F	G	H
τ_s^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
τ_t^2	0.08	1.48	0.51	0.08	0.53	0.06	0.02	0.00
τ_{st}^2	13.05	13.44	12.28	13.05	13.05	13.44	12.83	13.46
σ_s^2	23.75	23.27	23.55	23.75	23.75	23.27	23.06	23.26
σ_t^2	2.13	0.06	0.01	2.13	1.03	2.13	0.02	2.14
σ_{st}^2	0.00	0.00	2.37	0.00	0.00	0.00	2.80	0.00
ϕ_s	8.86	22.33	8.87	8.86	8.86	22.33	22.25	22.38
ϕ_t	5.44	0.4	0.4	5.45	0.4	5.45	0.4	5.13
ϕ_{st}	1.69	1.56	4.49	1.48	1.54	1.58	5.54	6.92
SSErr	8.18	7.08	8.23	8.18	8.41	**6.85**	**6.87**	**6.85**

A (spatial: Exponential, temporal: Exponential, joint: Exponential); B (spatial: spherical, temporal: spherical, joint: spherical); C (spatial: exponential, temporal: spherical, joint: exponential); D (spatial: exponential, temporal: exponential, joint: spherical); E (spatial: exponential, temporal: spherical, joint: spherical): **F (spatial: spherical, temporal: exponential, joint: spherical); G (spatial: spherical, temporal: spherical, joint: exponential); H (spatial: spherical, temporal: exponential, joint: exponential)**

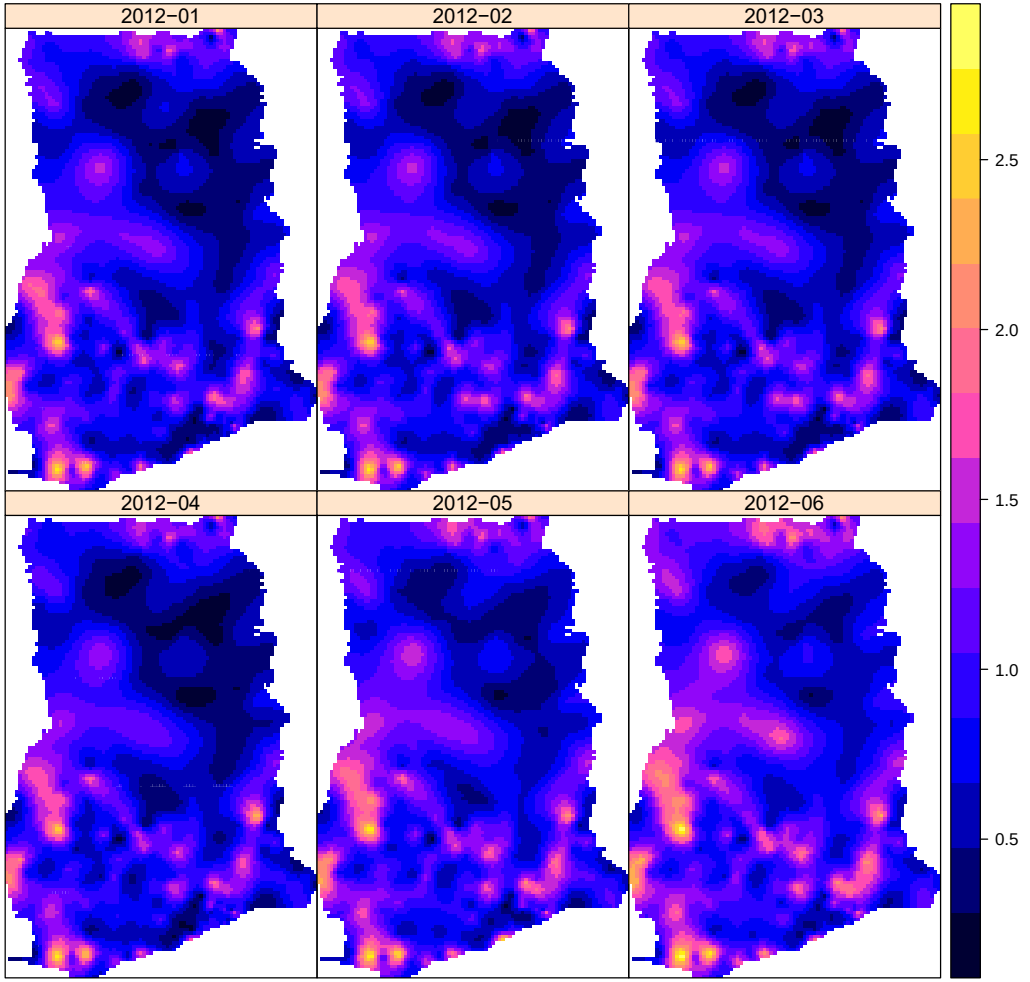
SPACE-TIME PREDICTION

- The prediction equation is a weighted average of neighboring risks
  - $\hat{\tau}_{io,t0} = \sum_{it}^{mT} \lambda_{it} \tau_{it}$
- Therefore
  - $\hat{r}_{i0,t0} = \exp\left(\beta_0 + \sum_{k=1}^{K} \hat{\beta}_{kt} x_{i0,kt} + \sum_{p=1}^{P} \hat{\omega}_k z_{i0,k}\right) + \hat{\tau}_{io,t0}$
- The kriging weights express the strength of the association between observation locations and the prediction location, estimated as
  - $\lambda_{it} = \gamma_0^T \gamma^{-1}$
- The kriging weights takes into consideration the the spatial covariances between all data points as well as the covariances between the prediction locations and observation location

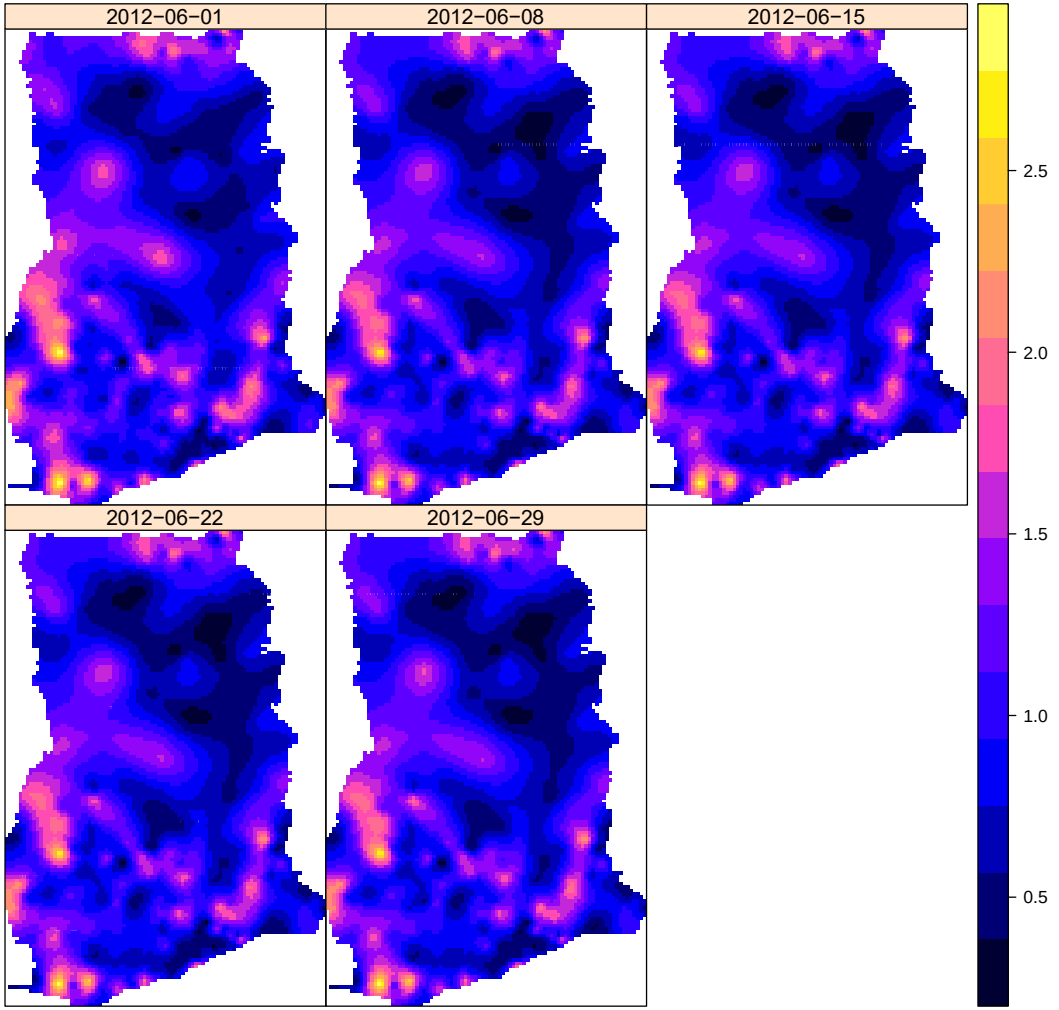
UNCERTAINTIES AND CROSS VALIDATION

- Prediction uncertainties were assessed through the kriging variance
  - $\sigma^2(s_0, t_0) = \gamma_0^T \gamma^{-1} \gamma_0$
- We further used the leave-one-out cross validation (LOOCV) to assess the prediction accuracy.
- The Root Mean Square Error (RMSE) is then used to assess the accuracy of the kriging predictions.
  - $RMSE = \sqrt{\frac{1}{m \times T} \sum_{s=1}^{m} \sum_{t=1}^{T} (\tau_{it} - \hat{\tau}_{it})^2}$

RESULTS: JAN-JUNE-2012



RESULTS: WEEKS IN JUNE 2012



CONTRIBUTIONS/CONCLUSIONS

- **Methodology**
  - Developed a spatial statistical framework for space-time prediction of malaria
  - The method is transferable to other diseases with similar data structure
- **Epidemiologically**
  - Observed spatial correlation of malaria risk, reflecting spatial clustering
  - The spatial clustering is heterogenous across times/different months
  - The spatial patterns and clusters are somehow spatially persistent
- **Further/Future work**
  - Account for over-dispersion since the Poisson assumption has no variance parameter
  - Improve the prediction by incorporating high temporal resolution external information for the covariates

CONCLUSIONS

- We have shown here that spatial statistics can be employed for analyzing and modelling geographically referenced health data
  - Either point pattern, lattice, or geostatistical data in terms of the underlying data generating process
  - Points and area data in terms of the feature representation
  - The methods demonstrated in the sample studies are only few. With more challenging data structure, different relevant methods can be adapted and/or emerge
  - This affirms that the beautiful marriage between spatial statistics and public health conceives GeoHealth

THANK YOU

- **Q $ A**