



**MAHIDOL
UNIVERSITY**
Wisdom of the Land

Spatiotemporal Bayesian Networks for Malaria Prediction

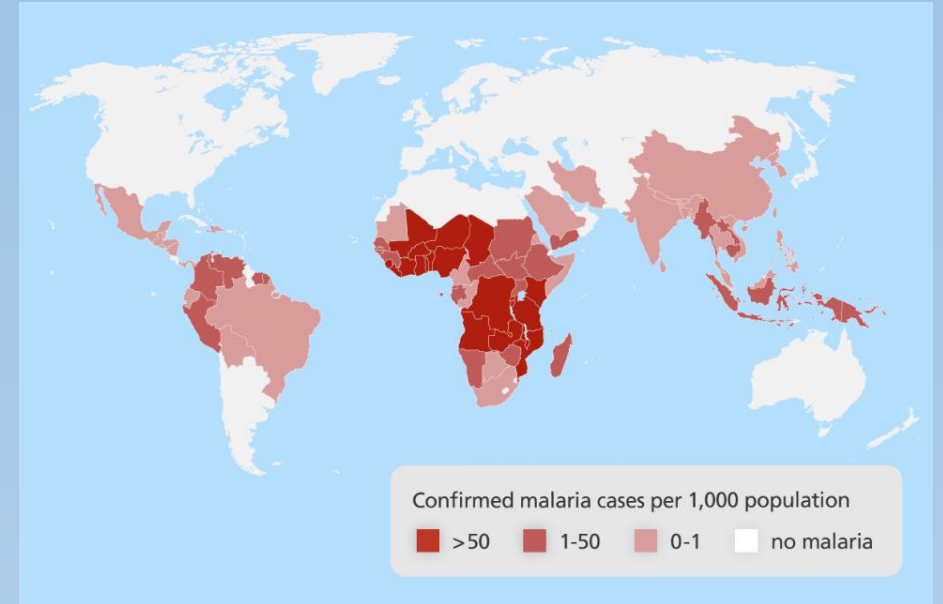
Prof. Dr. Peter Haddawy

Faculty of ICT

Mahidol University

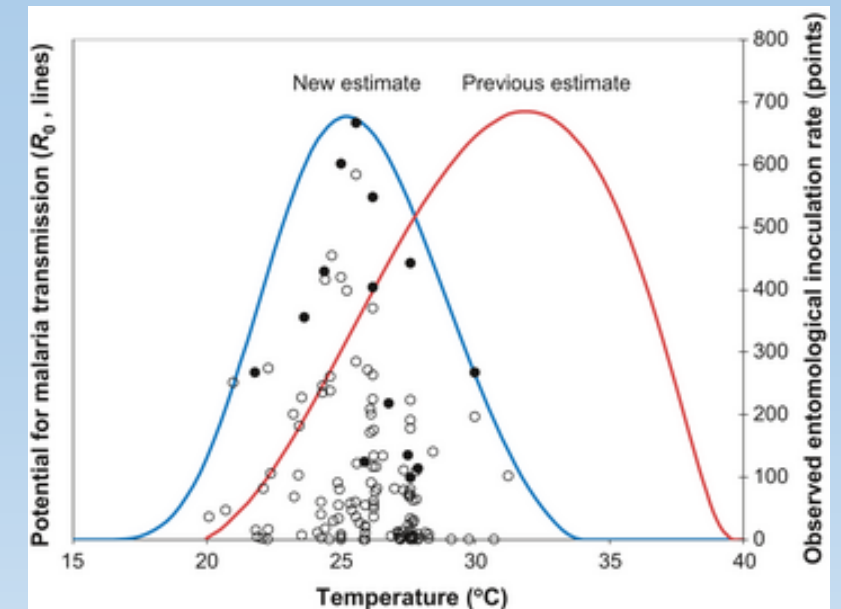
Malaria

- Global public health problem
 - 214 million cases globally in 2015
- Vector-borne disease
 - Parasite: *Plasmodium falciparum* (pf)
 - Vector: *Anopheles* mosquito
- Prevalent in less developed and more remote areas where public health resources are often scarce
 - Targeted intervention and allocation of clinical resources: where, when
- *Rapid detection* and *prediction* are essential



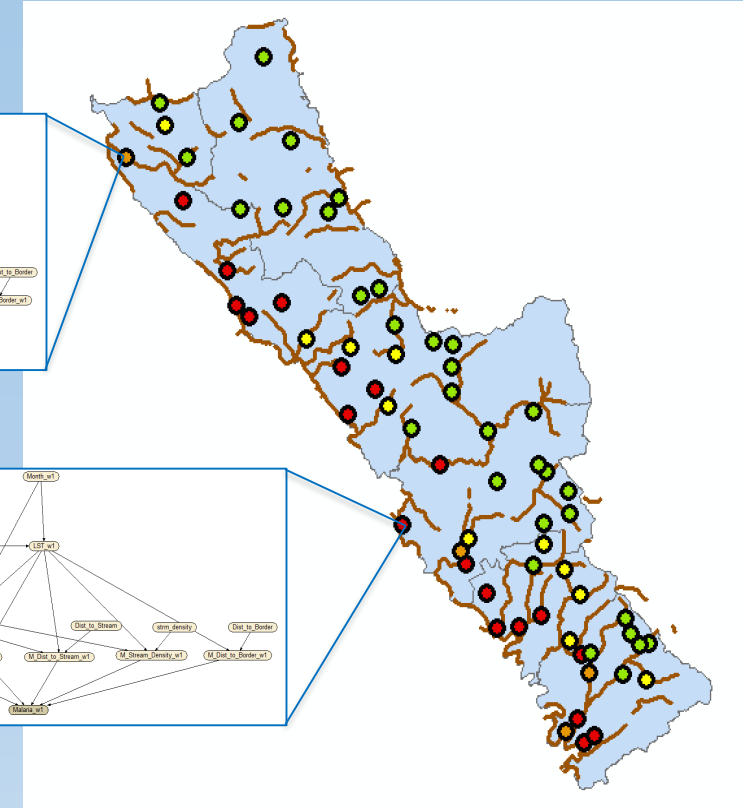
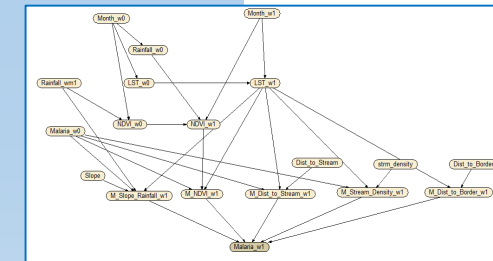
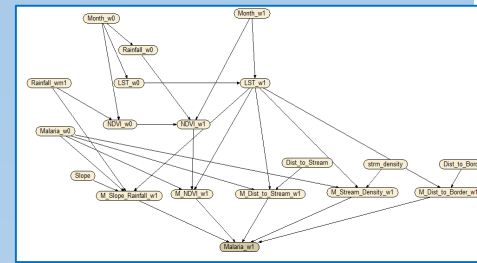
Challenges

- Disease transmission can exhibit spatial and temporal heterogeneity, spatial autocorrelation, and seasonal variation
- Environmental factors can affect vector population and disease transmission in complex and non-linear ways
 - Temperature – vector & parasite maturation, biting rate
 - Rainfall – flushing effect
- Representing uncertainty



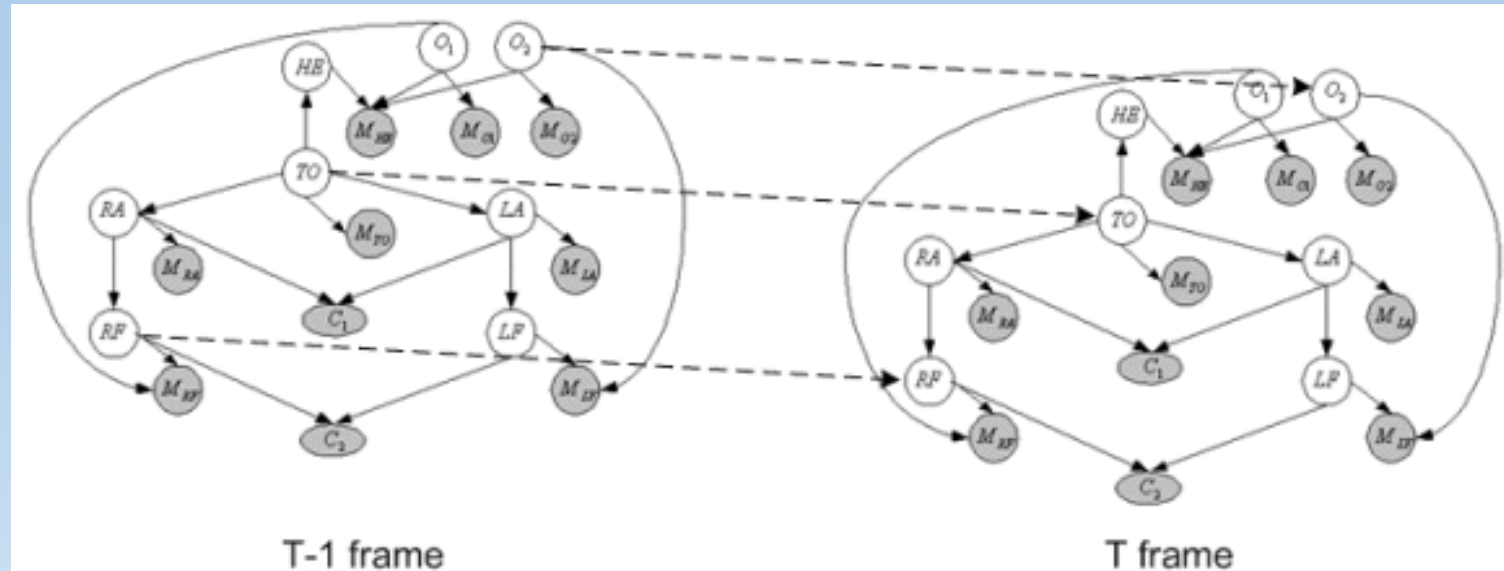
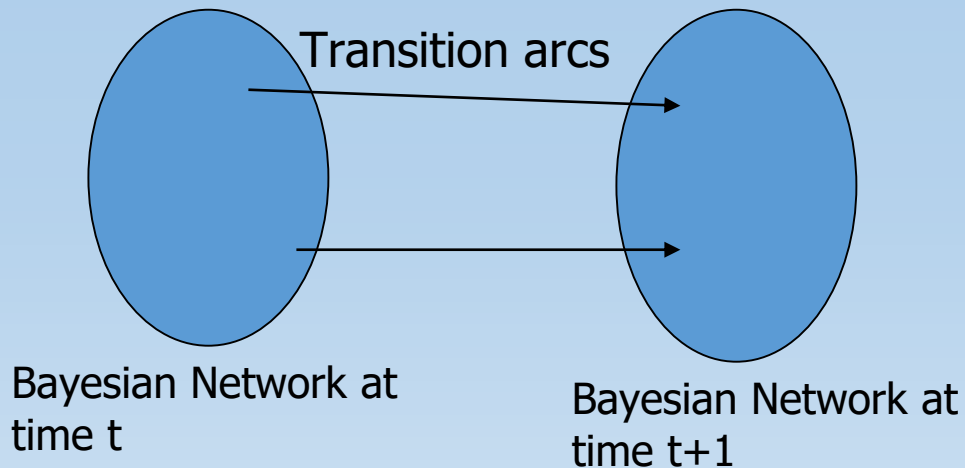
Bayesian Networks

- Spatial and Temporal dimensions of disease transmission
- Non-linear relations
- Explicit representation of uncertainty
- Can be used in a decision-theoretic framework to support decision making
- Incorporate expert judgement and data
- Models can provide explanations



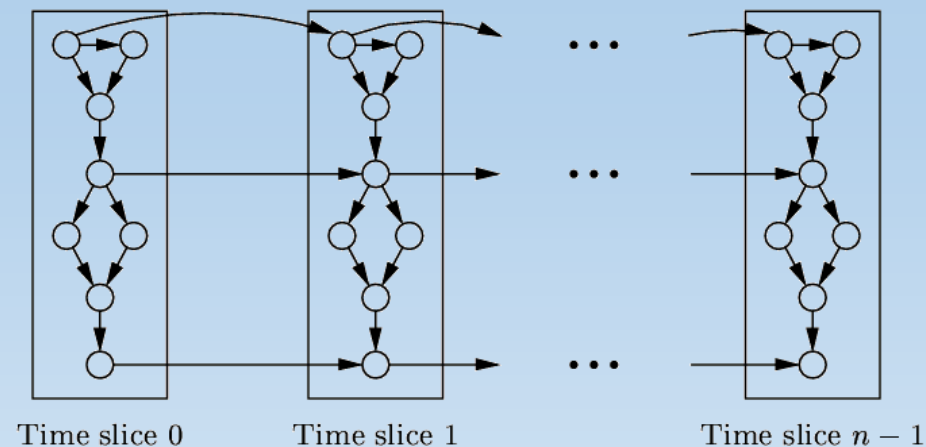
Dynamic Bayesian Networks (DBNs)

- A way to represent Markov models: HMM's, Kalman filters, others
- Nodes organized into time slices
- States: Collections of nodes in a time slice and relations among them
- Transitions: Links between time slices



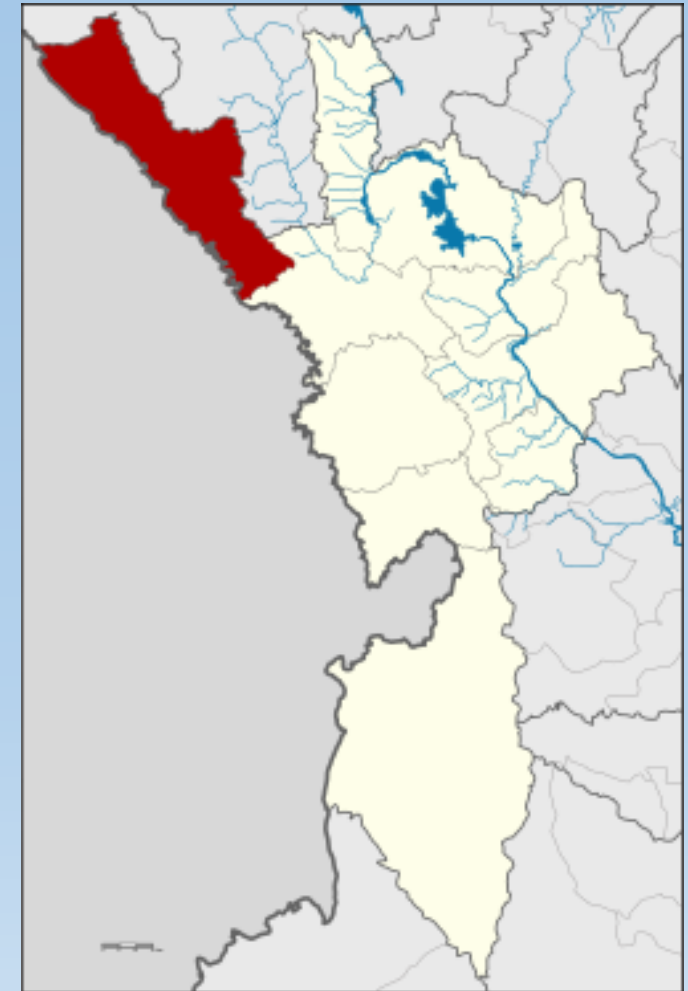
DBN Model

- Edge means how tight the coupling is between nodes
- Effect is immediate \rightarrow edge within same time slice
- Effect is delayed \rightarrow edge between time slices
 - Edge between same node = persistence
 - Edge between different nodes = time-lagged effect
- The larger the time scale, the more links within a time slice and the fewer between time slices



Tha Song Yang (ท่าสองยาง)

- 1,920 km² district in Tak province
- 66 villages near border with Myanmar
- Malaria is endemic
- Imported cases from Myanmar



Malaria Case Data

- Two years (2012 – 2013) of weekly clinically confirmed malaria case data from 66 villages
- 6,579 records with 12,800 total cases (PF, PV)
 - Cases: Min 0, Max 82, Mean 2.1

Environmental Data

temporal

- NDVI
 - Monthly from MODIS (MOD11A3)
 - Temporal interpolation used to fill in missing values
- Rainfall
 - Daily at 10 km resolution
 - From JAXA Global Rainfall Watch
- Land Surface Temperature (LST)
 - Monthly at 5 km resolution from MODIS (MOD11C3)
 - Spatial and temporal interpolation to fill in missing values

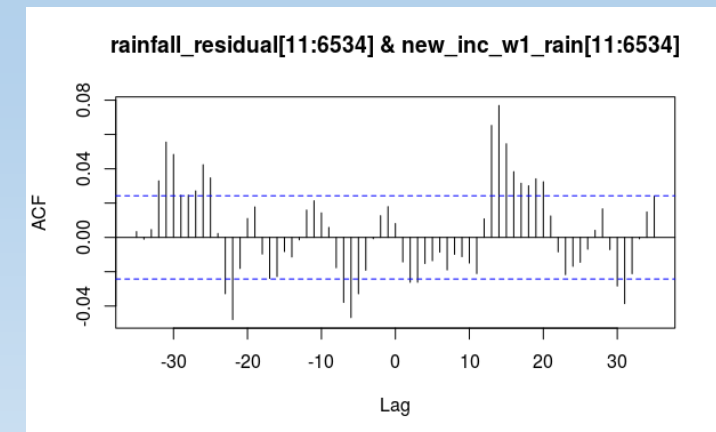
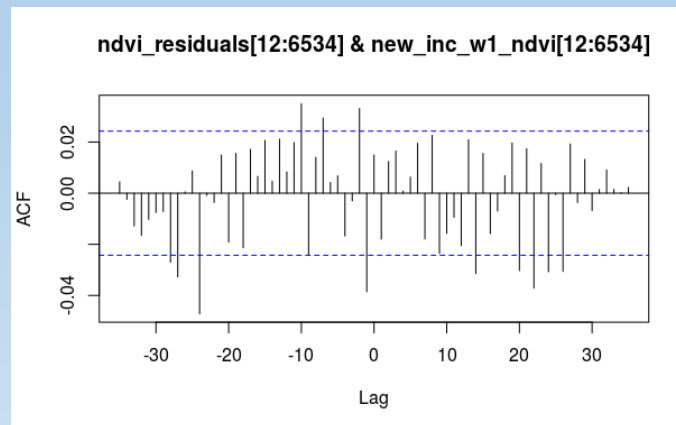
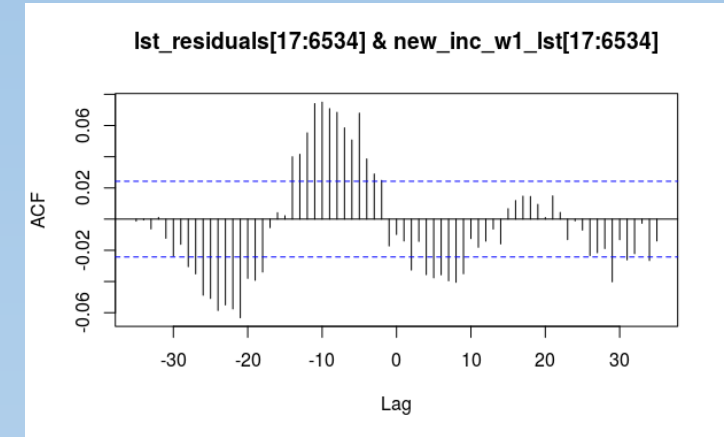


non temporal

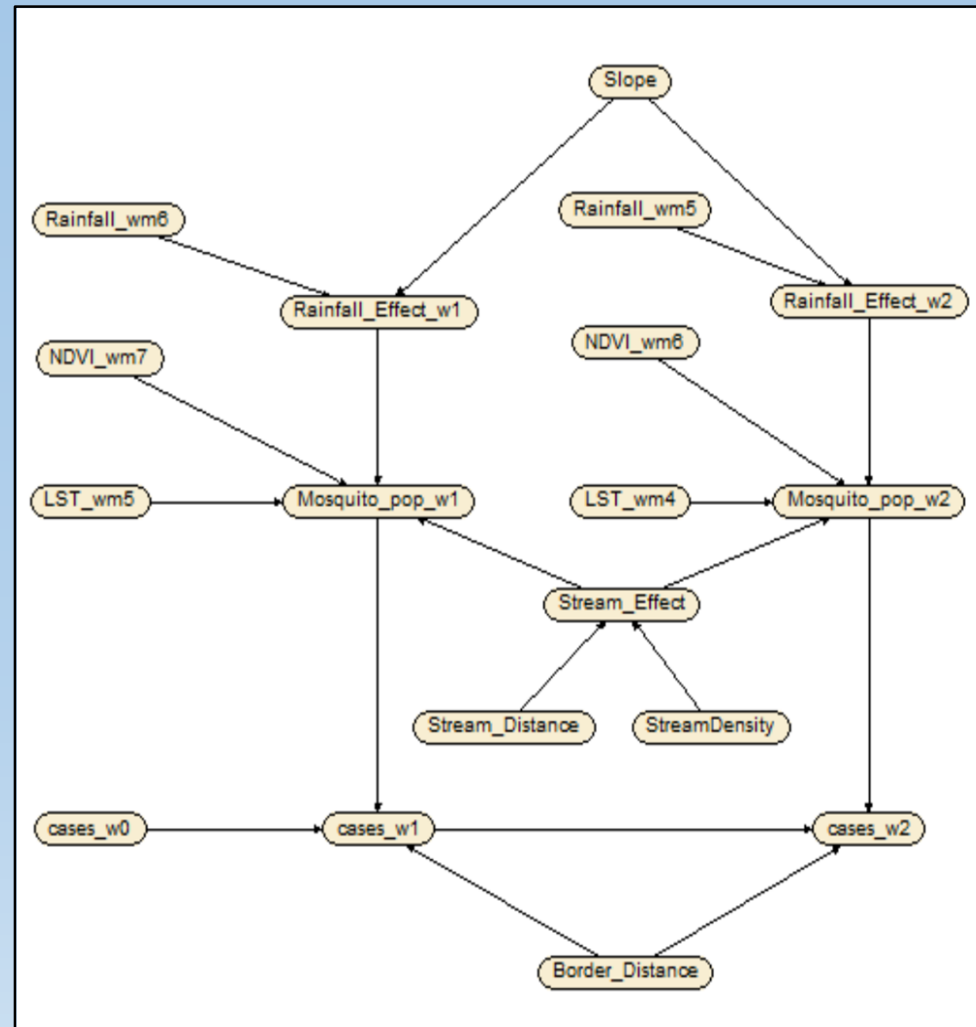
- Streams
 - Distance to closest stream
 - Stream density: Total stream length in 4 km buffer
- Slope
 - Average in 1 km buffer (from elevation data)
- Distance to border: proxy for imported cases

Model Building: Time Lags

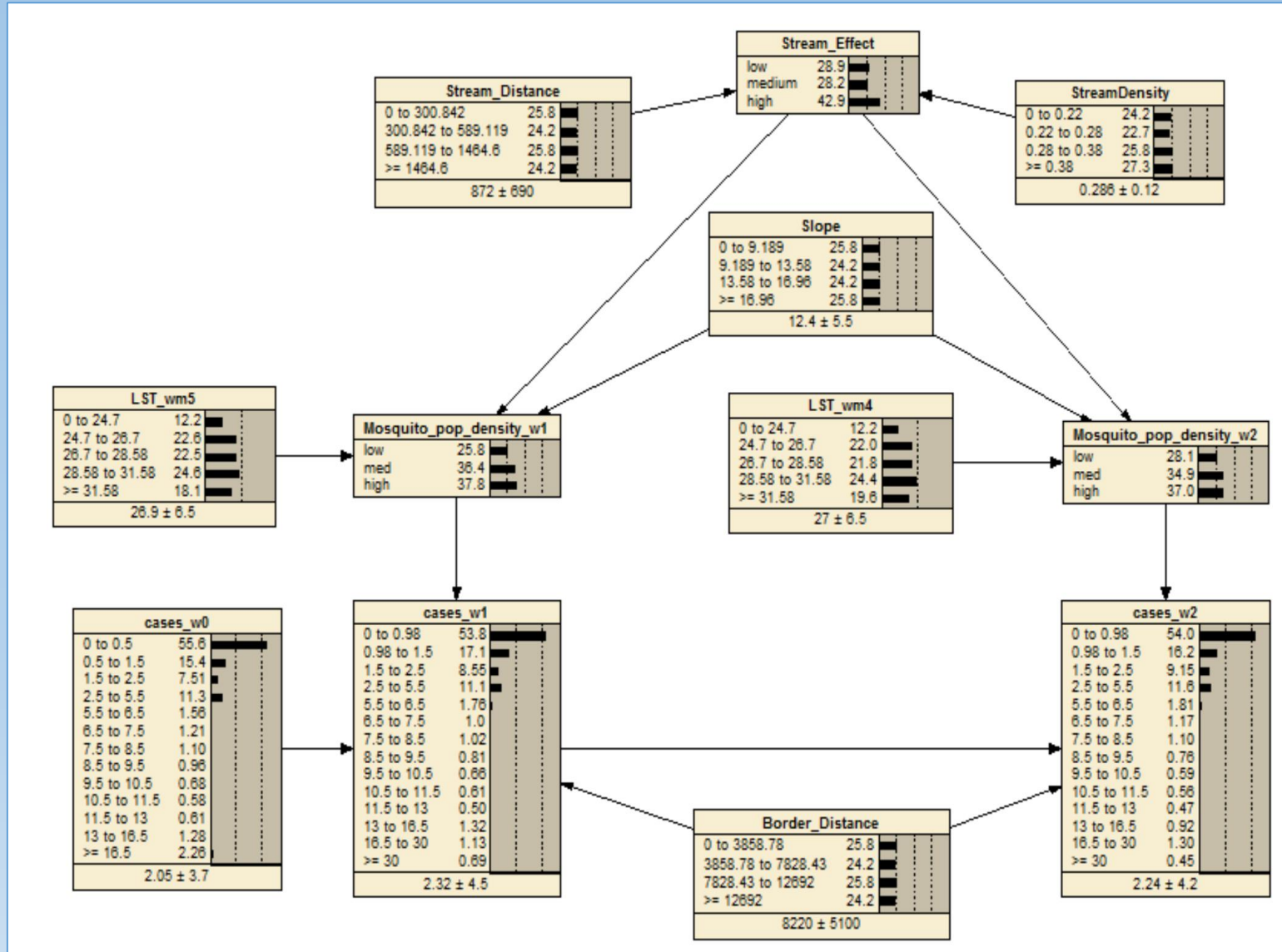
- Environmental variables effect different parts of the vector and parasite cycle
- Time lags determined using cross-correlation with pre-whitening + model fitting
 - Fit ARIMA model to independent variable X
 - Use to filter dependent variable Y
 - Calculate cross-correlation on residuals for X and filtered Y
- LST: 6 weeks
- Rainfall: 7 weeks
- NDVI: 8 weeks



Initial Village-Level Bayes Net Model

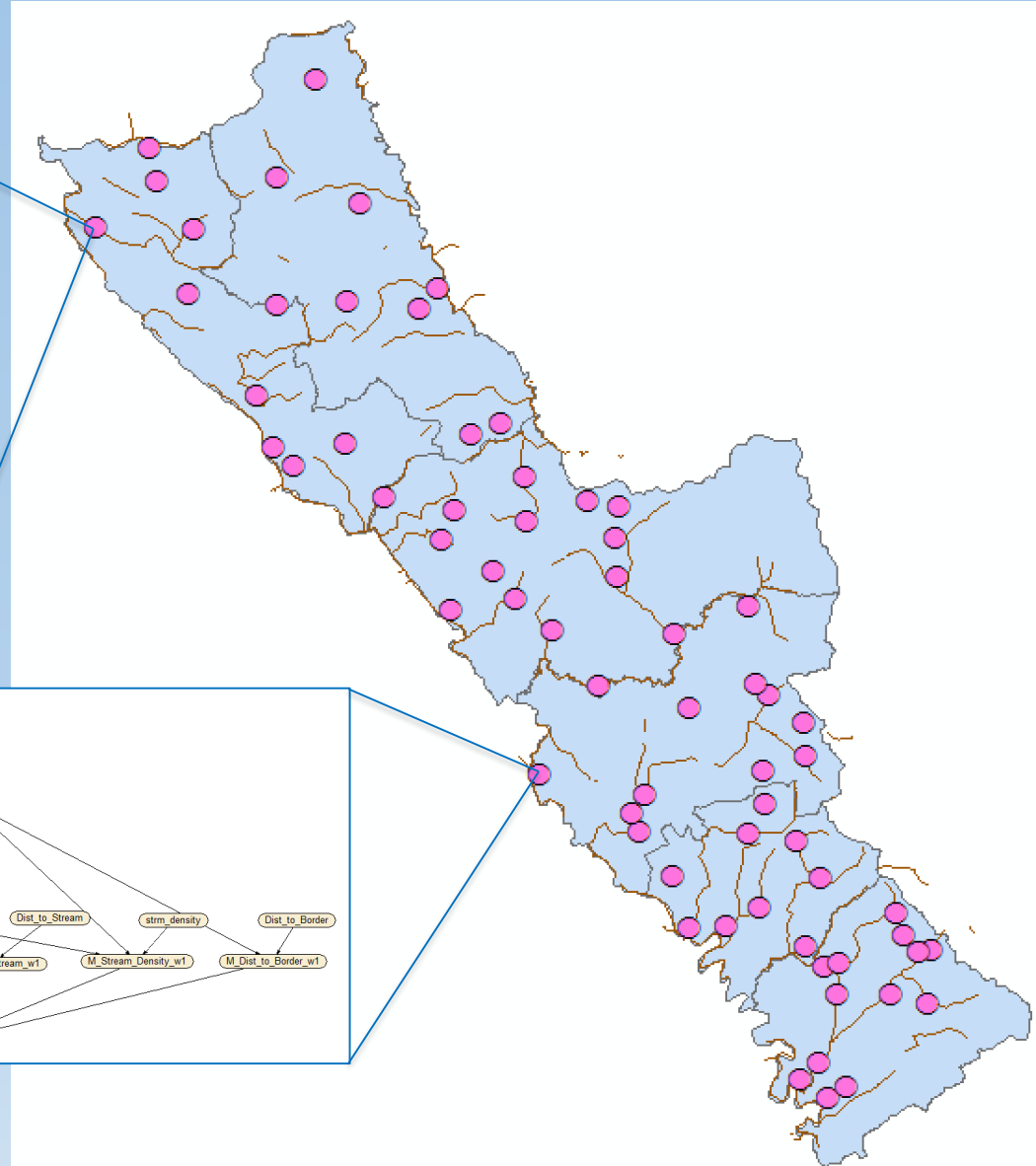
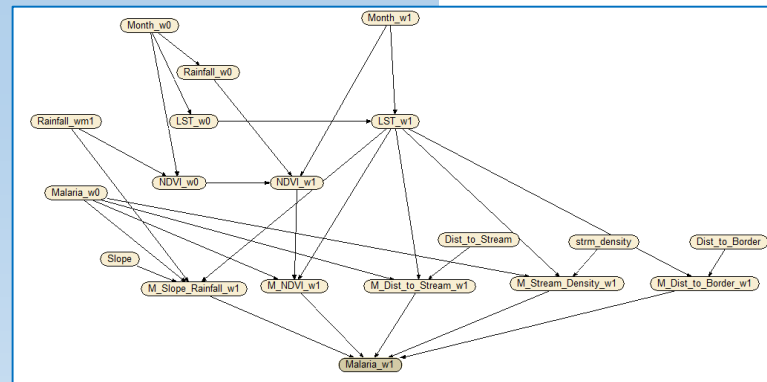
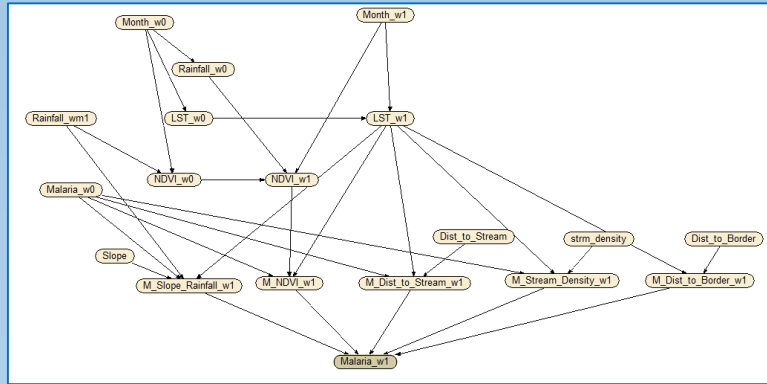


Final Village Level Bayes Net Model



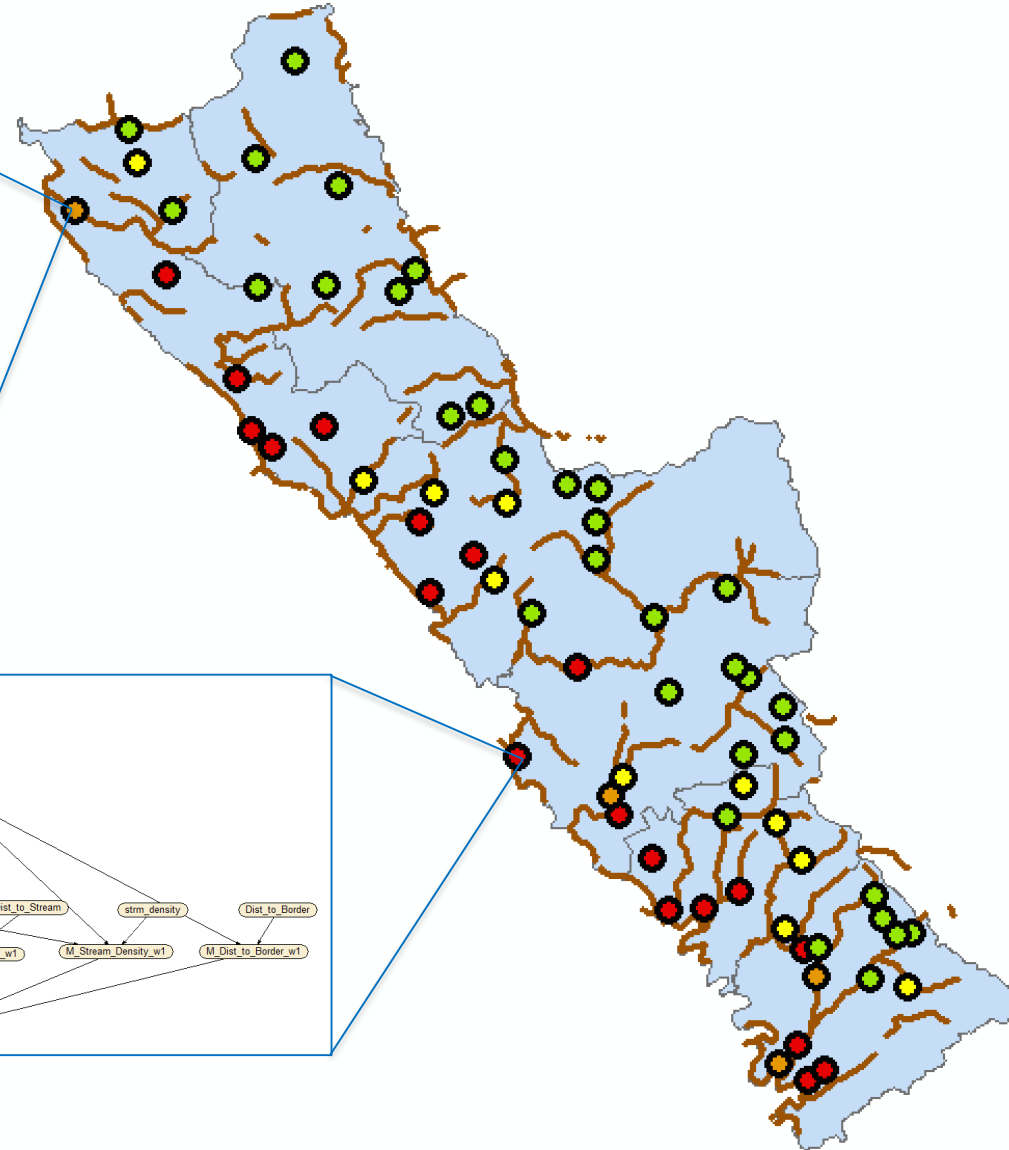
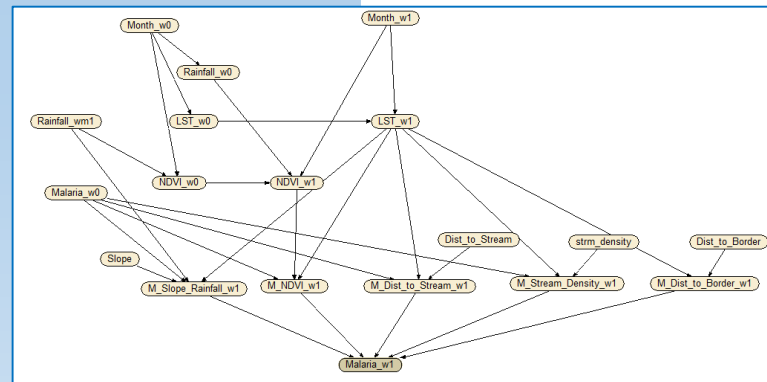
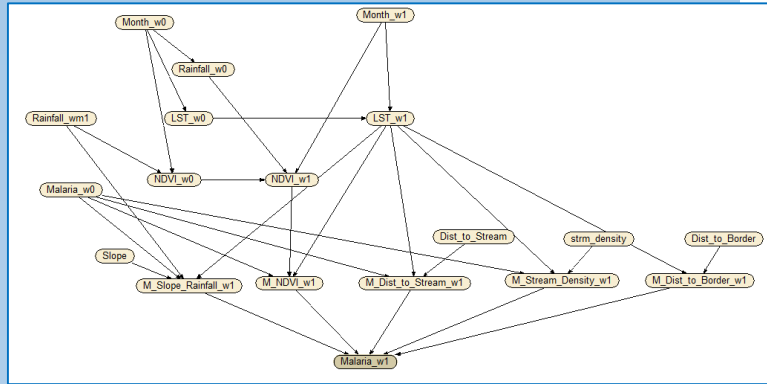
Linking Bayes Net to GIS

One Bayes net per village



Linking Bayes Net to GIS

One Bayes net per village



Prediction Accuracy (MAE)

Subsets	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
13 high	3.387	3.404	3.454	3.504	3.43	3.457
13 med	1.886	2.073	2.129	2.319	2.299	2.411
14 low	0.305	0.368	0.422	0.463	0.519	0.538
All 66	1.415	1.501	1.557	1.644	1.657	1.729

BN outperforms
ARIMA,
ARIMAX,
Linear
Regression,
Poisson
Regression

- High incidence {Min: 0, Max: 82, Ave: 7.43},
- Medium incidence {Min: 0, Max: 16, Ave: 1.91}
- Low incidence {Min: 0, Max: 3, Ave: 0.099}

Comparison of BN Relative to ARIMA

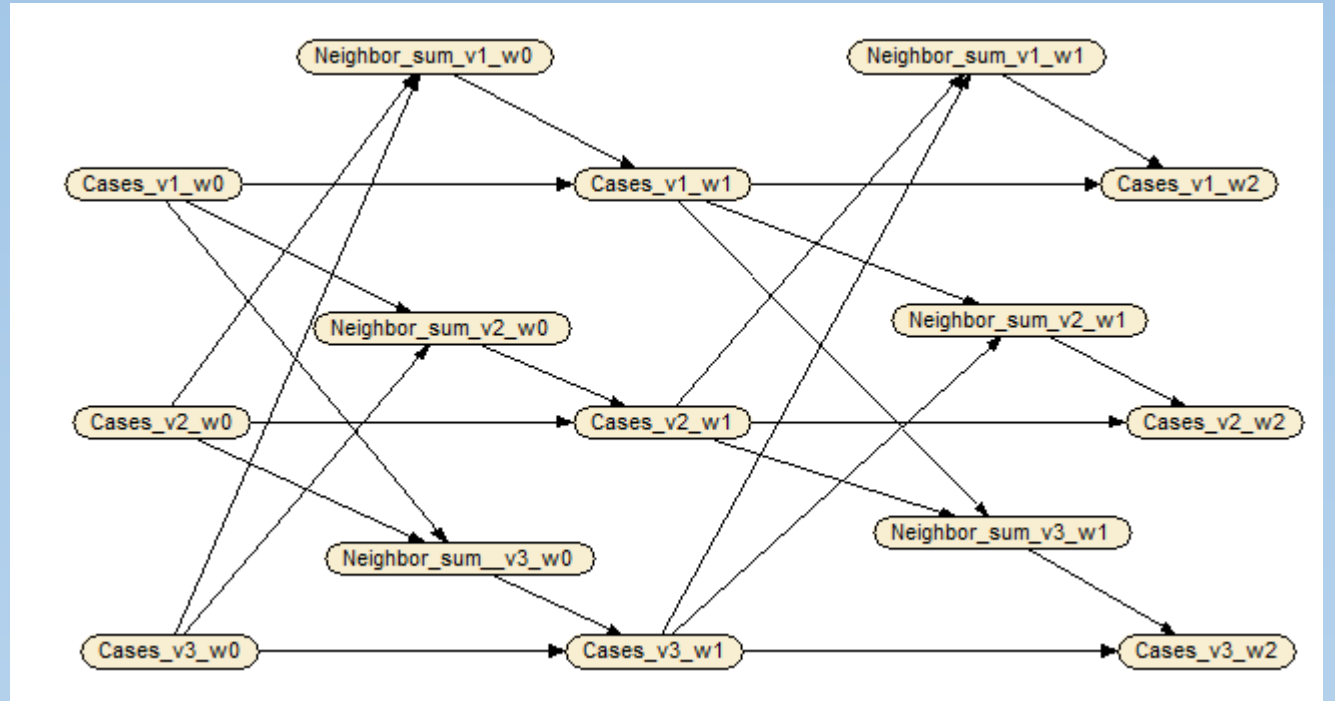
- ARIMA outperforms BN for medium and low incidence villages
- BN outperforms ARIMA for high incidence villages for all time horizons past 1 week. The longer the horizon, the larger the difference.

Subset	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
13 high	-3.8%	6.4%	9.3%	15.0%	20.5% *	26.9% *
13 med	-1.7%	-4.3%	-3.6%	-10.0%	-8.0%	-9.8%
14 low	-102.0% *	-145.3% *	-158.9% *	-178.9% *	-268.1% *	-304.5% *
All 66	-6.1%	-3.1%	-4.4%	-4.9%	-4.6%	-2.1%

* difference statistically significant (two-tailed t test $p < 0.05$).

Modeling Spatial Autocorrelation

- Individual village models can be linked together
- Sum of cases in nearby villages (Neighbor_sum) influences incidence in current village
- But models become too large and complex to build by hand



Mutual influence among 3 neighboring villages

Solution: Knowledge-Based Model Construction

- Store model fragments in a library
- Represent as rules using probability logic

$$\forall x P(\text{Flies}(x) \mid \text{Bird}(x)) = 0.9$$

- Automatically construct models tailored to data in GIS

Model Rules

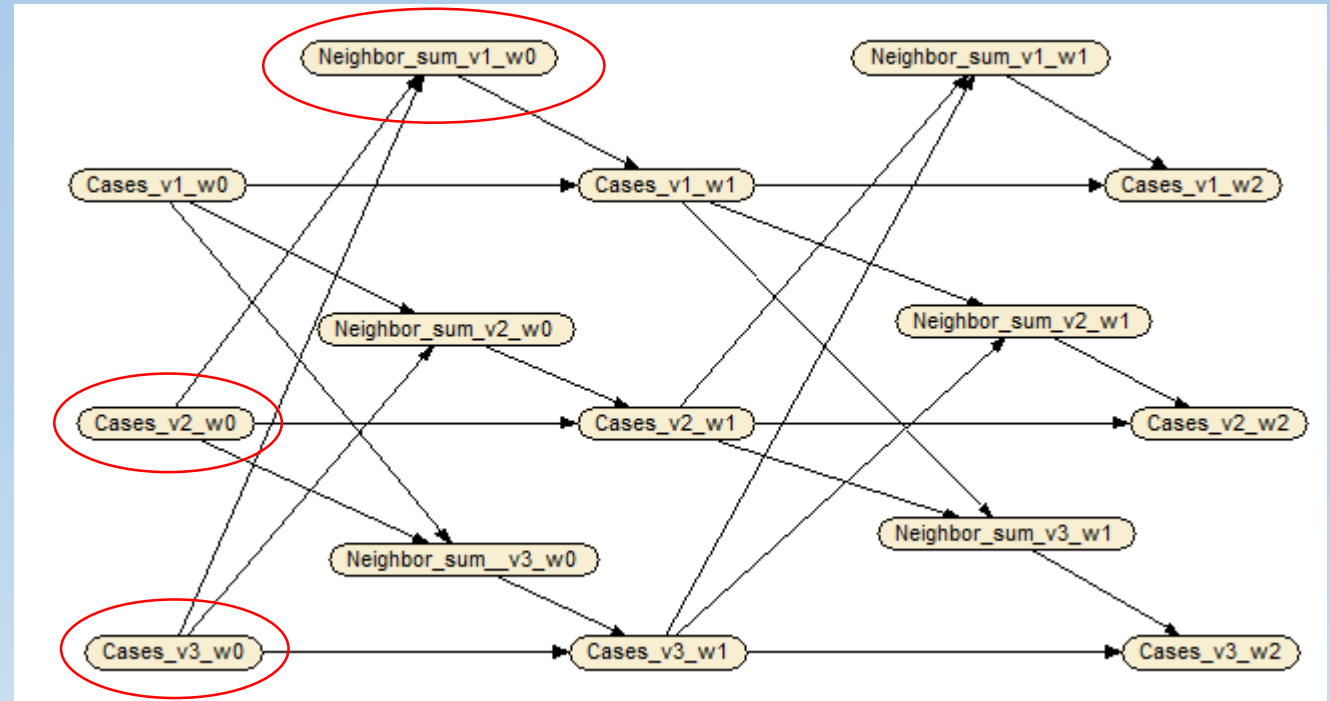
If villages are close together, malaria may spread from one to another

FOR(t:TIME)(x1:LOCATION)(x2:LOCATION) WHERE $\text{dist}(x1,x2) \leq 3000$ AND $(x1 \neq x2)$

PARENT **NeighborSum** AT x1,t IS **Cases** AT x2,t

CPT AUTO_SUM;

CPT AUTO_SUM is a function that constructs the CPT to represent a sum function depending on the number of parents.

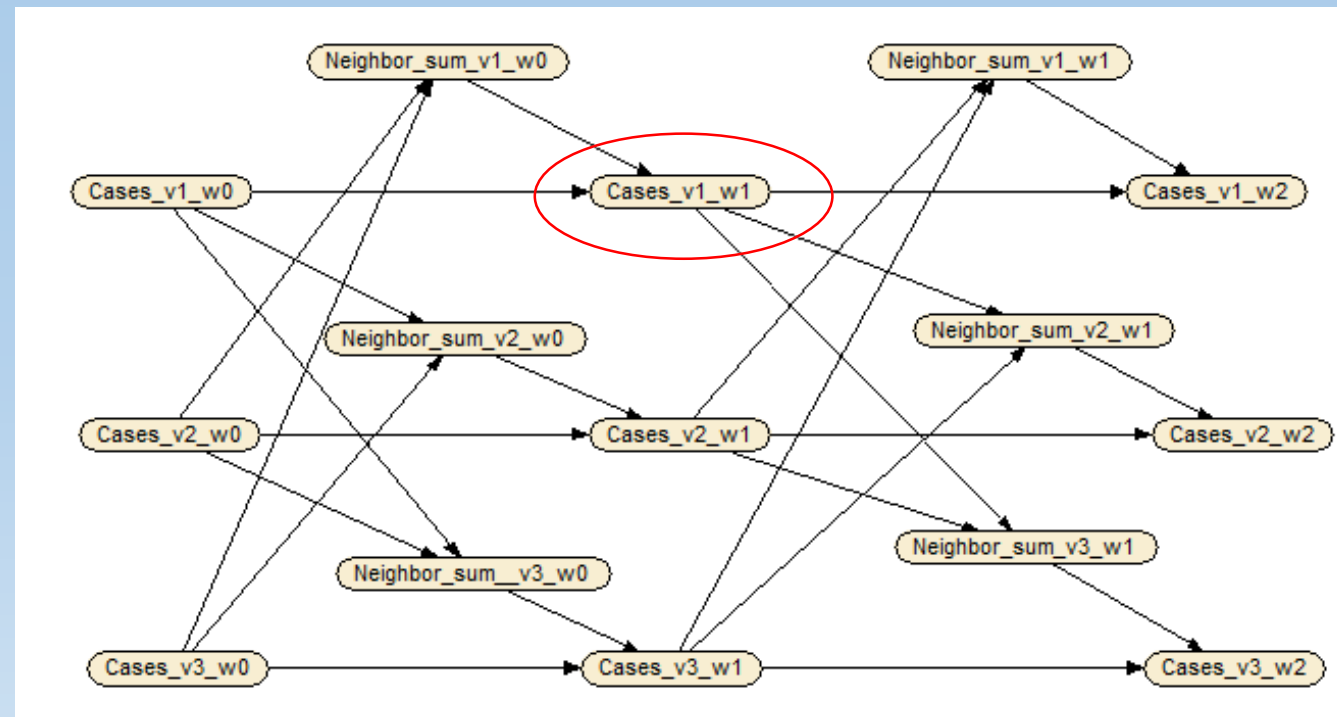
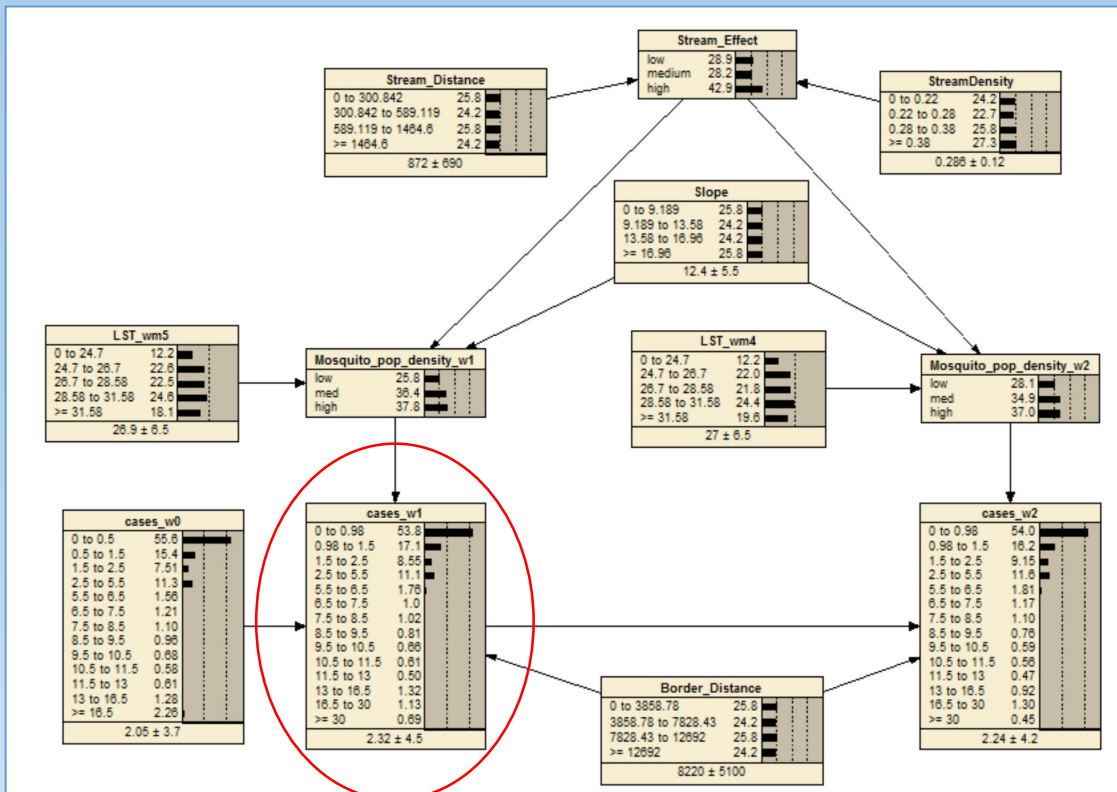


Model Rules

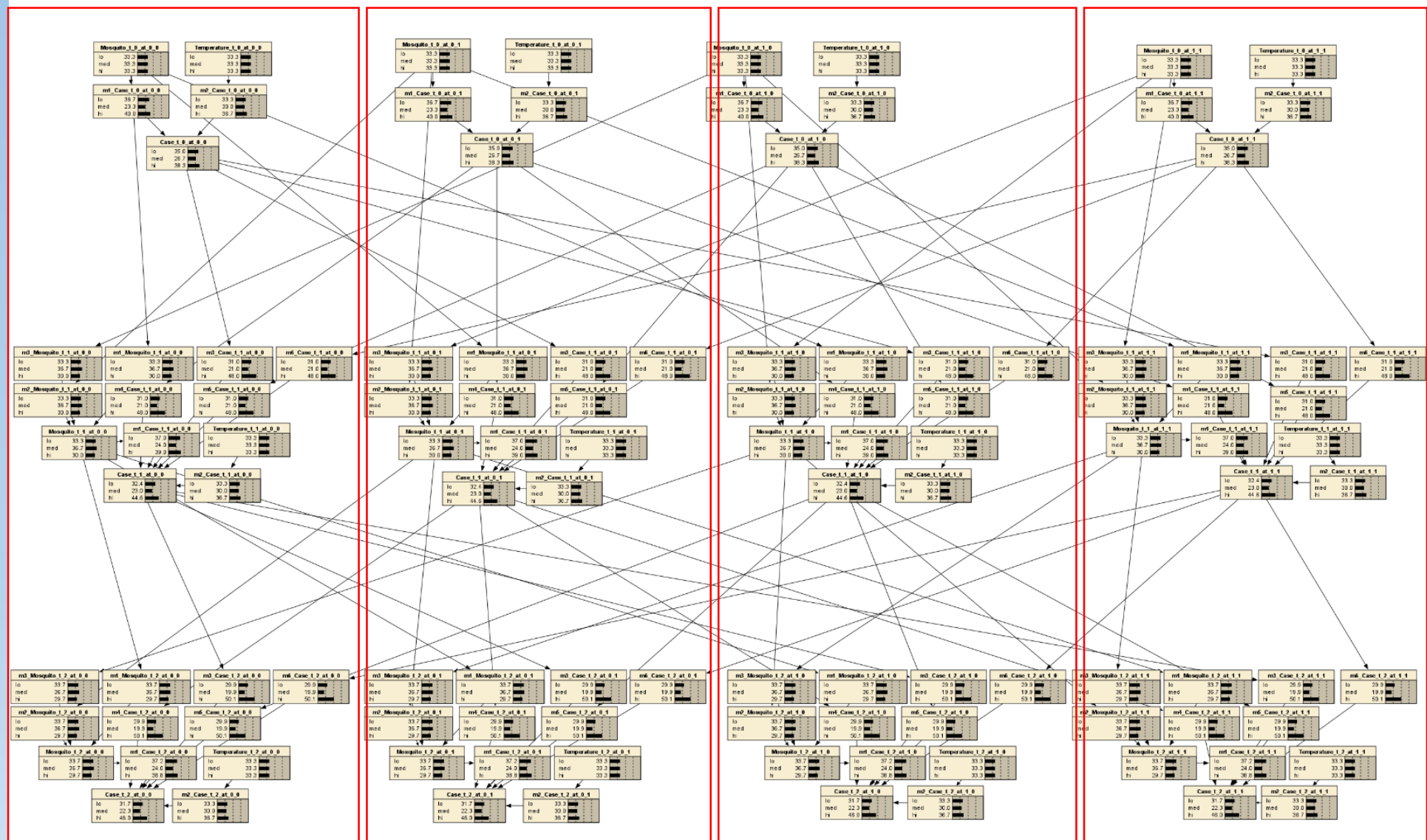
Cases has parents BorderDistance, previous week's Cases, MosquitoPop, and NearestSum

FOR(t1:TIME) (t2:TIME) (x:LOCATION) WHERE t2=t1+1

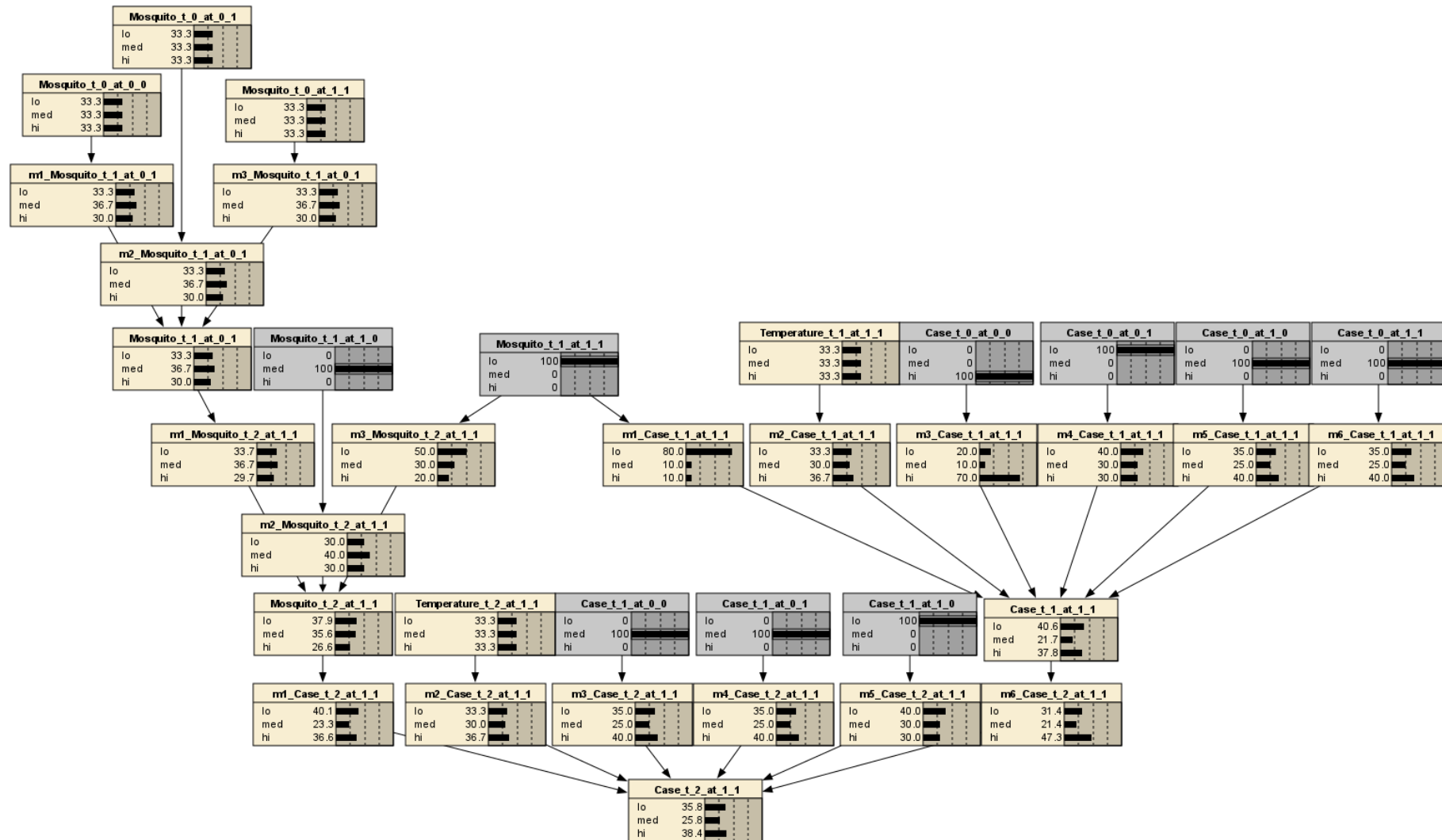
PARENT **Cases** AT x,t2 IS **BorderDistance** AT x,t2 AND **Cases** AT x,t1 AND **MosquitoPop** AT x,t2 AND **NeighborSum** AT x,t1 CPT (...);



Model for entire problem space for a subset of villages



Model to answer a query given evidence



Improvement using Autocorrelation Model

Village No.	Total Incidence	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
109	2672	17.8% *	15.6% *	11.8%	8.92%	3.34%	5.78%
201	412	1.18%	7.75%	8.52%	4.24%	7.66%	5.38%
208	274	0.88%	3.34%	11.3% *	9.77%	10.6%	10.4%
205	83	12.9% *	22.5% *	22.2% *	18.2%	12.7%	9.05%
410	79	5.03%	4.06%	6.84%	9.15%	11.0%	8.84%
107	51	4.02%	6.17% *	12.5%	16.2%	21.6% *	20.9%

* difference statistically significant (two-tailed T test $p < 0.05$)

Future Work

- Malaria may not spread between villages that are close due to separation by geographic features like hills or rivers and it may spread between villages that are more distant because they are linked by roads.
- Investigate whether prediction accuracy could be improved by a more refined model of spatial autocorrelation taking into account geographic features.
- Incorporate breeding site information from drone images

Further reading

P. Haddawy, A.H.M. Imrul Hasan, R. Kasantikul, S. Lawpoolsri, P. Sa-angchai, J. Kaewkungwal, P. Singhasivanon, Spatiotemporal Bayesian Networks for Malaria Prediction, *Artificial Intelligence in Medicine*, 84, pp 127-138, 2018.

P. Haddawy, M. Su Yin, T. Wisanrakkit, R. Limsupavanich, P. Promrat, S. Lawpoolsri and P. Sa-angchai, Complexity-Based Spatial Hierarchical Clustering for Malaria Prediction, *Journal of Healthcare Informatics Research*, 2(4), pp 423-447, Dec 2018.

A.H.M. I. Hasan and P. Haddawy, S. Lawpoolsri, A Comparative Analysis of Bayesian Network Approaches to Malaria Outbreak Prediction, *Proc. 13th Int'l Conf. on Computing and Information Technology* (IC2IT 2017), Bangkok, July 2017.

A.H.M. I. Hasan and P. Haddawy, Integrating ARIMA and Spatiotemporal Bayesian Networks for High Resolution Malaria Prediction, *Frontiers in Artificial Intelligence and Applications*, vol 285: *Proc. European Conference on Artificial Intelligence (ECAI 2016)*, pp 1783 – 1790, Aug 2016.