

**INTEGRATING ARIMA AND SPATIOTEMPORAL BAYESIAN
NETWORKS FOR HIGH RESOLUTION MALARIA PREDICTION**



A. H. M. IMRUL HASAN

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2017**

Copyright by Mahidol University

COPYRIGHT OF MAHIDOL UNIVERSITY

INTEGRATING *ARIMA* AND SPATIOTEMPORAL *BAYESIAN* NETWORKS FOR HIGH RESOLUTION MALARIA PREDICTION

A. H. M. IMRUL HASAN 5738465 ITCS/M

M. Sc. (COMPUTER SCIENCE)

THESIS ADVISORY COMMITTEE: PETER HADDAWY, Ph.D. SARANATH LAWPOOLSRI, Ph.D. SONGSRI TANGSRIPAHOJ, Ph.D.

ABSTRACT

Since malaria is prevalent in less developed and more remote areas in which public health resources are often scarce, targeted intervention is essential in allocating resources for effective malaria control. To effectively support targeted intervention, predictive models must be not only accurate but they must also have high temporal and spatial resolution to help determine when and where to intervene.

This thesis developed a high resolution prediction model through the combination of Bayes nets and ARIMA. Bayes net and ARIMA have complementary strengths, with the Bayes nets better able to represent the effect of environmental variables and ARIMA better able to capture the characteristics of the time series of malaria cases. Also, experiments show that the Bayes net predicts malaria cases better in higher incidence regions whereas ARIMA performs better in lower and medium incidence regions. Leveraging these complementary strengths, two ensemble predictors were developed. Based on mean absolute error (MAE), the developed ensemble predictors had significantly better accuracy than either predictor alone. The impact of improvement in the problem domain by the designed ensemble models was also verified by conducting an outbreak detection experiment.

The models were built and tested with data from Tha Song Yang district in northern Thailand, creating village-level models with weekly temporal resolution.

KEY WORDS: BAYESIAN NETWORKS / ARIMA MODELS / TIME SERIES ANALYSIS / MALARIA MODELING

85 pages