Title	Development and evaluation of the model for the spatial heterogeneity of disease burdens corresponding to disparities of the environmental context for the efficient healthcare resource deployment : a case of drug-resistant malaria issue in Cambodia
Sub Title	
Author	狼, 卓(Okami, Suguru) 神武, 直彦(Kotake, Naohiko)
Publisher	慶應義塾大学大学院システムデザイン・マネジメント研究科
Publication year	2015
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2015年度システムエンジニアリング学 第190号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40002001-00002015-0024

Master's Dissertation 2015

Development and Evaluation of the Model for the Spatial Heterogeneity of Disease Burdens Corresponding to Disparities of the Environmental Context for the Efficient Healthcare Resource Deployment: A Case of Drug-Resistant Malaria Issue in Cambodia

Suguru Okami

（Student ID Number：81433093）

Supervisor

Associate Professor Naohiko Kohtake

March 2016

Graduate School of System Design and Management,

Keio University

Major in System Design and Management

SUMMARY OF MASTER'S DISSERTATION

Student Identification Number	81433093	Name	Suguru Okami

Title

Development and Evaluation of the Model for the Spatial Heterogeneity of Disease Burdens Corresponding to Disparities of the Environmental Context for the Efficient Healthcare Resource Deployment: A Case of Drug-Resistant Malaria Issue in Cambodia

Abstract

Long lasting fight with Malaria has been entering in new situations. Still, Malaria is an important global health issue causing millions of deaths for many years. Substantial effort for the malaria containment action by many stakeholders has decreased the burden of this infectious disease in a number of endemic places. However, a number of remaining issues and emerging challenges still exist in the way toward malaria elimination. One emerging issue is the artemisinin resistance reported mostly in the Greater-Mekong subregions.

This thesis studies about the development and evaluation of the system for the modeling and computational simulation approach of the spatial heterogeneity of malaria disease burdens to visualize the effectiveness of containment actions for the purpose of optimal healthcare resource utilization in western Cambodia where the artemisinin resistance was previously reported. The mapping approach with malaria spatial risk distribution modeling is an effective tool and widely used for malaria containment actions, which was established through the long history of effort by many contributors. However, as the disease burden of malaria decreases along the way malaria elimination effort progresses, this approach needs some adjustment in accordance with the situational changes surrounding malaria elimination effort. Under the low-to-moderate transmission settings such as those in Cambodia, mathematical modeling methods and measurement approaches of malaria risk need reinvestigations.

To address this issue, we applied a mathematical modeling approach for standardized morbidity ratio (SMR) calculated by annual parasite incidence (API) using routinely aggregated surveillance reports, environmental data such as remote sensing data and non-environmental anthropogenic variables to create spatial risk distribution maps in fine-scale of two provinces (Pailin and Preah Vihear) in western Cambodia. Furthermore, we incorporated the combination of given containment status indicators into the model to demonstrate the regional heterogeneities of

the relationship between containment status and risks.

The estimated SMR by empirical Bayesian method (EBSMR) for each operational health district in western Cambodia was calculated using routinely aggregated surveillance reports. We then developed an explanatory mathematical model for EBSMR using environmental variables calculated from remote sensing data and non-environmental anthropogenic variables. Bayesian modeling frame was applied to estimate the uncertainty of the model and cross-scale predictions. Finally, we created maps visualizing the risks by interpolating the estimated SMR at each village and conducted computational simulations to demonstrate the relationship of expected outcomes and containment status indicators.

The explanatory model was fitted to estimate the SMR of each area (adjusted $R^2 = 0.774$, AIC = 149.423). Fine-scale maps were created by the inverse distance weighed method (IDW) and ordinal kriging interpolation of estimated SMR at each village. In comparison with geocoded case data, corresponding predicted values showed conformity [Spearman's rank correlation; r = 0.662 in IDW and 0.645 in ordinal kriging (95% confidence interval; 0.414 – 0.827 and 0.368 – 0.813, respectively), Welch's t-test; N.S.]. Computational simulations demonstrated visual representations of the different expected outcomes of interventions in respective areas.

We conclude the validity of the proposed approach by which, regional malaria risks can be well explained and fine-scale risk maps can be created under the low-to-moderate malaria transmission settings where reinvestigations of existing risk modeling approaches are needed. Moreover, different representations of simulated outcomes of containment status indicators for respective areas provide useful insights for the tailored interventional planning considering regional malaria endemicity.

Further studies are needed to demonstrate how this system will provide the effect in the Cambodian health information system. The implications from this study suggested the system would provide increased reciprocity of the information, by which the improved quality of reported data could be expected.

Key Word (5 words)

Mathematical modeling, Risk modeling, Risk mapping, Spatial epidemiology, Malaria

論　文　要　旨

学籍番号	81433093	氏名	狼　卓

論　文　題　目

効率的医療資源配置のための環境コンテキスト不均衡性に対応した空間的疾病負荷偏在性モデルの構築と評価 −カンボジアにおける薬剤耐性マラリア問題を事例として−

要　旨

長期にわたり続いてきたマラリアとの戦いは新たな局面に入っている。依然として、マラリアは何年もの間、数百万人もの死亡の原因となっている世界的な健康問題であるが、多くの関係者の多大な努力により、この感染性疾患の疾病負荷は多くの地域において減少した。しかしながら、幾つかの残された問題や新たな課題がマラリア撲滅を目指す途上に残されている。そのうちの一つが、メコン河流域で主に報告されているアルテミシニン耐性である。

本研究の目的は、過去にアルテミシニン耐性が報告されたカンボジア西部における、マラリア対策活動の有効性を可視化し、最適な医療資源配置の実現を目的としたマラリアの空間的疾病負荷偏在性モデリングとコンピュータシミュレーションの一連の流れをシステムとして取り扱い、その構築と評価を行うことである。マラリアの空間的なリスクの分布をモデル化し、可視化するアプローチはこれまで多くの研究者の貢献により確立されており、既存のマラリア対策活動の効果的な手法として幅広く用いられている。しかしながら、マラリア撲滅の取り組みが進行することによりマラリアの疾病負荷が減少した状況下においては上記の手法の修正が必要となる。カンボジアのような底-中等度の感染伝播状況下においてはマラリアの数理モデリングや測定方法について再度調査を実施する必要がある。これらの問題を解決するために、我々は常時実施されているサーベイランスにより集積された年間発生率(Annual Parasite Incidence; API)から算出した標準化罹患比及びリモートセンシングなどの環境データ、その他の人為的データを変数として用いた数理モデリング手法を適用し、カンボジア西部の二つの州(パイリン州およびプレアビヒア州)における空間的リスク分布を可視化した詳細スケールの地図を作成した。加えて、我々はマラリア対策の状況を示す指標をモデルに組み込むことによって、これらの指標とリスクの関係性に地理空間的偏在性があることを示した。

常時集積されたサーベイランス報告を用い、カンボジア西部地域のヘルスオペレーショナ

ルディストリクト各地における経験ベイズ法により推定された標準化罹患比(EBSMR)を算出した。その後、EBSMR を説明する数理モデルを、リモートセンシングデータを用いて算出した環境的変数及び非環境的な人為的変数を用いて構築し、複数の空間スケールの予測に伴うモデルの不確実性を推測するためにベイズモデルの枠組みを適用した。最後に各村レベルで推定された標準化罹患比の空間的補間を行うことによってリスクを可視化した地図を作成し、コンピュータシミュレーションにより、対策の状況を示す指標とそこから期待される結果の関係性を示した。

結果、各地の標準化罹患比を推定する数理モデルが選択された(自由度調整済み決定係数 : 0.774, 赤池情報量基準 : 149.423)。詳細スケールの地図は逆距離加重法及び通常クリギング法を用い、村レベルで推定された標準化罹患比の空間的補間により作成された。また、過去に報告されたジオコーディングされた症例データと本研究により得られた地図から予測される同地点のリスクを比較したところ、一致性が確認された [スピアマンの順位相関係数; r = 0.662 (逆距離加重法) , 0.645 (通常クリギング), 95%信頼区間; 0.414 – 0.827(逆距離加重法), 0.368 – 0.813(通常クリギング), ウェルチのt検定; N.S. (逆距離加重法および通常クリギング)]。コンピュータシミュレーションにより対象地域によって異なる対策とその結果の可視化された関係性を得ることができた。

これらのことにより、常時集積されているサーベイランス報告とリモートセンシングデータや非環境的な人為的データを組み合わせることによりマラリアの各地のリスクはよく説明されていたと言える。構築されたモデルを用いることによって、詳細スケールのリスクマップを底-中等度の感染伝播状況下においても作成することができた。マラリア対策から得られる地域によって異なる期待される効果を示すことにより、マラリアの地域特性を考慮した上で介入を計画する際に有益な示唆の提供が可能となることが期待される。

本研究で評価されたシステムがカンボジアにおける医療情報システムの中でどのような効果をもたらすかについてはさらなる検討が必要である。今回の検討によりシステムは情報の返報性を高め、報告されるデータの質を改善させる可能性があることが示唆された。

キーワード (5 語)

数理モデリング、リスクモデリング、リスクマッピング、空間疫学、マラリア

Table of contents

List of figures

List of tables

ACKNOWLEDGEMENT

At first, I would like to express my deepest gratitude to Associate professor Naohiko Kohtake, whose advices and suggestions as a first supervisor comments gave me the direction of my research. Furthermore, he provided me with various opportunities to study about geospatial science. Through my experience in the Public health project of G-SPASE, I received not only a lot of instructive advices but also the support to conduct field study in Cambodia. This extremely valuable experience expanded my expertise and opened my eyes for the future. In addition, I would like to express my gratitude to Associate professor Makoto Ioki for reviewing my thesis as a secondary advisor and made a number of thoughtful suggestions for my research.

I would like to thank to Akihiko Nishino, Daisuke Takahata, Junichi Sasaki, Kaori Mizutani and Kinuko Iizawa for making a lot of advices and encouragement for my research. I would also like to thank all the laboratory members and the classmates of SDM, Keio University for sharing knowledge and discussions.

In addition, I would like to thank to Ms. Akemi Takahashi, Ms. Hiroko Oji, Ms. Mao Sugita and all the staff in Cambodia office of Foundation of International Development / Relief as collaborators of my research. They offer invaluable support and coordination for our field study in Cambodia. I would also like to express my gratitude to Hiroki Kojima as a colleague of the G-SPASE public health project.

This research has been conducted as a part of G-SPASE Program supported by Japanese Ministry of Education, Culture, Sports, Science and Technology and Space Application Promotion Program funded by NEC Corporation.

Finally, I would like to express my deepest gratitude to my wife for supporting me throughout years at SDM, Keio University. I can’t thank her enough and would like to start to reciprocate her dedicated support from now on.

1. INTRODUCTION

1.1 Background

1.1.1 Motivations

Spatial epidemiology is widely applied as an effective tool to understand the geospatial distributions of disease burdens and the level of health of people to reduce the risk threatening public health [1]. In recent years, a number of technologies related to this research area, such as the people mobility analysis using mobile phone log [2], the application of remote sensing technology [3], big data analysis and simulation technologies to predict the effectiveness of interventions, have made significant progresses, which have already been piloted or applied to a number of healthcare projects [4]. Still, further opportunities can be found for the improvements in providing more practical and visualized information for the appropriate drug use with these advanced technologies.

1.1.2 Malaria problem

Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected mosquitoes. It is a public health issue causing millions of deaths for many years. In 2013, malaria caused 584,000 deaths (Uncertainty range 367,000 – 755,000) [5]. Furthermore, malaria is the 5th biggest cause of death in children except neonatal causes (Figure 1) [6].

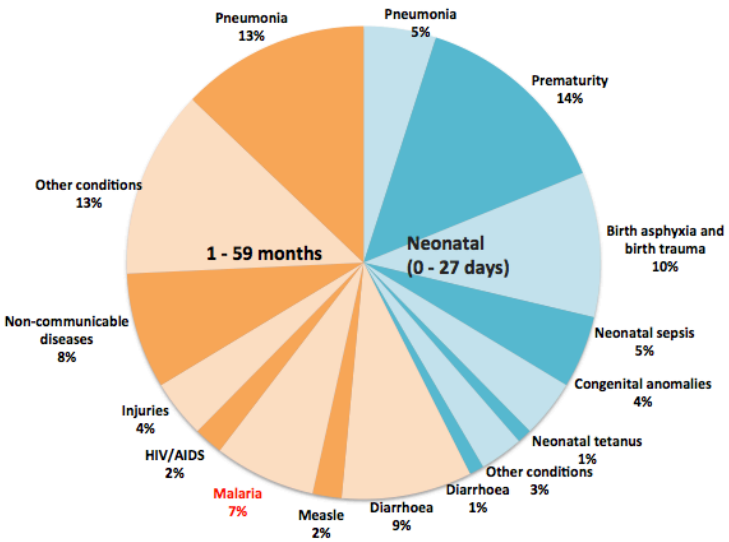


Figure 1 Cause of deaths among children under 5 years, 2011 [6]

Plasmodium falciparum, *Plasmodium vivax*, *Plasmodium ovale* and *Prasmodium malariae* are the parasite species that can transmit to humans. Malaria damages the body through a number of pathways. Malaria causes the red blood cell distruction leading to anemia. The waves of parasites are bursting red blood cells, which can be a trigger for the classic cycles of fever and chills. The changes of adhesive properties of infected redblood cells can block the blood flow in the vessels causing tissue hypoxia. If this tissue hypoixia is happen in brain, it can cause cerebral malaria, which is often fatal.

Once people get infected through the bites of vectors, infected *Anopheles* mosquitoes, they get acute febrile illness and several patients are in the severe condition such as hypoglycemia, anemia, respiratory distress and cerebral malaria, which can lead to the fatal conditions. Malaria also has chronic effects such as anemia, neurologic cognitive and developmental disfunction which lead to the impaired growth and development causes malnutrition as well as infant mortality and impaired productivity.

1.1.3 Situational changes in the malaria problem

Long lasting fight with Malaria has been entering in new situations. Still, Malaria is an important global health issue causing millions of deaths for many years. Substantial effort for malaria containment actions by many stakeholders has decreased the burden of this infectious disease in a number of endemic places (Figure 2).

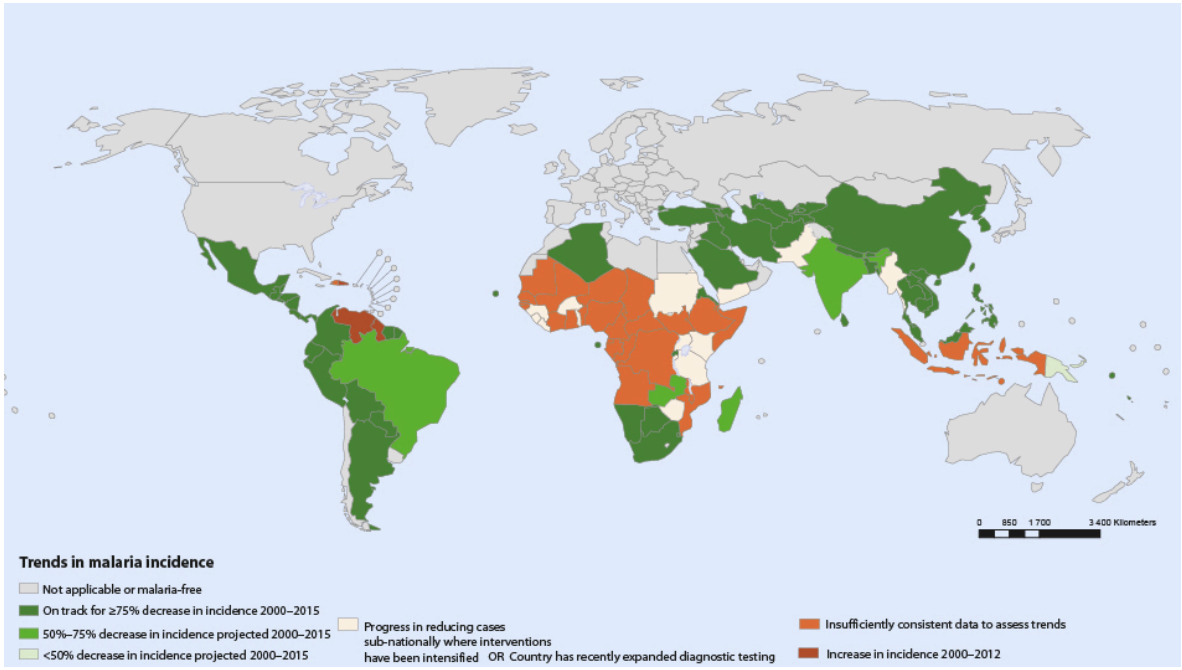


Figure 2 Trends in malaria incidence [7]

Green colored area represents more than 50% decrease in malaria incidence.

Some decades after the past effort of the global malaria eradication program, malaria elimination features again in global health agenda [8]. In recent years, an increasing number of countries with low-to-moderate transmission areas have moved into the actions toward the malaria elimination from their entire territories [9]. In Cambodia, this target is placed on 2025 [10]. Recent activities have decreased the incidence of malaria in Cambodia to less than a half compared with those in early 2000s [11]. Nowadays, about a half of people are living in malaria-free or low-transmission settings [12]. However, a number of remaining issues and emerging challenges still exist in the way toward malaria elimination.

1.1.4 Emerging artemisinin resisitance in the Greater Mekong subregions

One of the most potentially hazardous issues is the artemisinin resistance reported mostly in the Greater-Mekong subregions [13]. Artemisinin is a potent and rapidly acting blood schizontocide, effective for all plasmodium species, which is used in combination with other antimalarials (Figure 3) [14]. In *P. falciparum* malaria, artemisinin also kills the gametocytes – including the stage 4 gametocytes, which are otherwise sensitive only to primaquine. This drug should be given as combination therapy to protect from drug resistance because of its short half-life in the body.

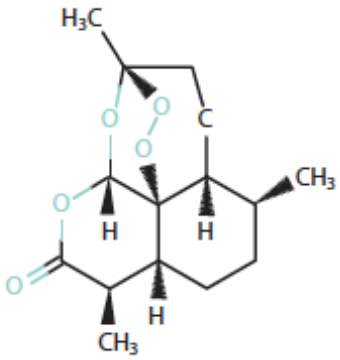


Figure 3 Chemical structure of artemisinin [14]

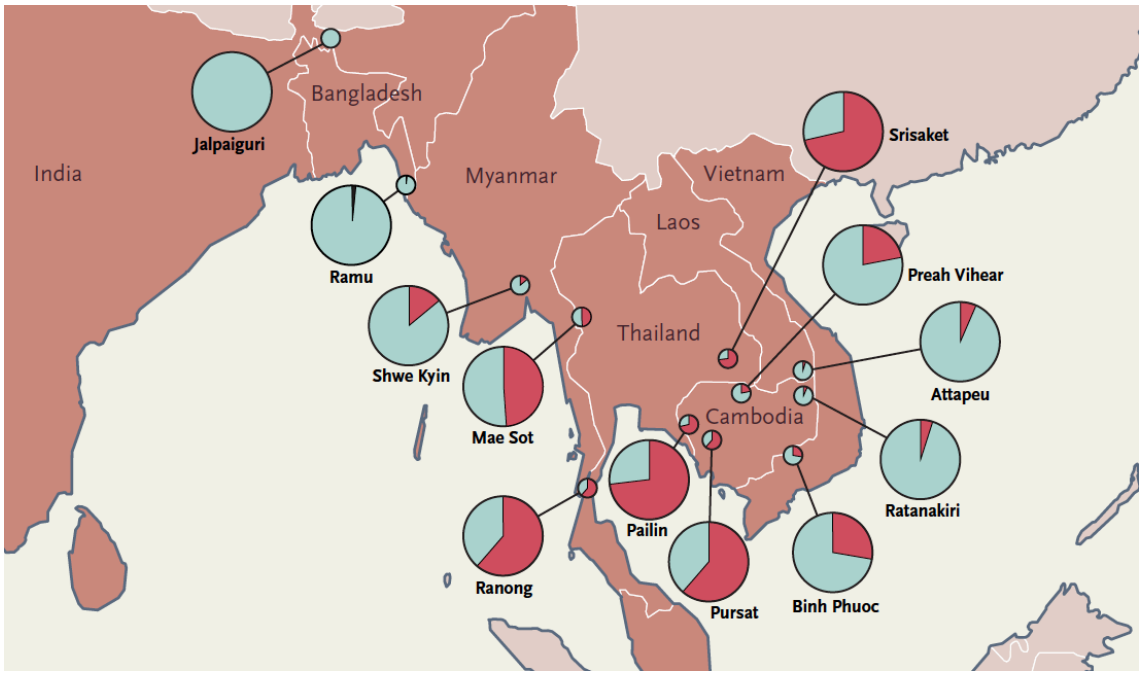


Figure 4 Extent of artemisinin resistance in the Greater-Mekong subregions [17]

The blue color represents the proportion of parasite clearance half-life is equal or less than 5hrs and the red color represents the proportion of parasite clearance half-life is more than 5hrs or *kelch 13* polymorphism at or beyond amino acid position 441.

The delayed parasite clearance, i.e. resistance to artemisinin, has so far been detected in five Southeast Asian countries; Cambodia, the Lao People's Democratic Republic, Myanmar, Thailand and Viet Nam and there has been a number of report of delayed parasite clearance in patients taking artemisinin in western Cambodia [15-19] (Figure 4). As no alternative effective antimalarial treatment can be used at present, the public consequence could be dire, if resistance spreads to wide geographical regions [15]. In areas along the Cambodia–Thailand border, *P. falciparum* has become resistant and multi-drug resistance is a current major concern [16]. Usually in such areas, mobility of people is high, which contains the potential dangers of spreading the multi-drug resistance to larger geographical areas. One recent report showed that the artemisinin resistant malaria has infectious potential to vectors in other geographical regions [20].

In relation to this, there are some issues in antimalarial drug use such as spreading availability of the artemisinin monotherapy, poor quality counterfeit medicines and unregulated antimalarial use [17, 21]. An important driver is said to be the use of oral artemisinins alone as monotherapy [15].

The reported treatment failure in western Cambodia varies on conditions [17,

22-24]. However, even considering the variance in these aggregated reports, all of them strongly suggest urgent needs for addressing this issue. Under the current situations, the appropriate medication use is undoubtedly important. The right patients should be screened and treated by the right way. In areas such as those in western Cambodian country border, this approach, in occasions, needs the intensive care and monitoring for the patients. In order to attain the required outcome, several studies such as focused screening and treatment [25], community based surveillance [26] and mass drug administration [27] have been piloted. What is common to these interventions are intensive support and monitoring for local practitioners were critical for obtaining desired outcomes.

1.1.5 Problem structure of aretemisinin resistance in Cambodia

An example of the problem structure of artemisinin resistance is shown in Figure 5. This problem structure was drawn through the interview results with the director of provincial health center (Table 1) and the literature search. Each element in this structure is connected by relevant causal relationship, indicating that multiplicative contributions made by multiple issues are consisting of this problem. The middle layer of the structure indicates two types of resource related issues, insufficiency and quality issues.

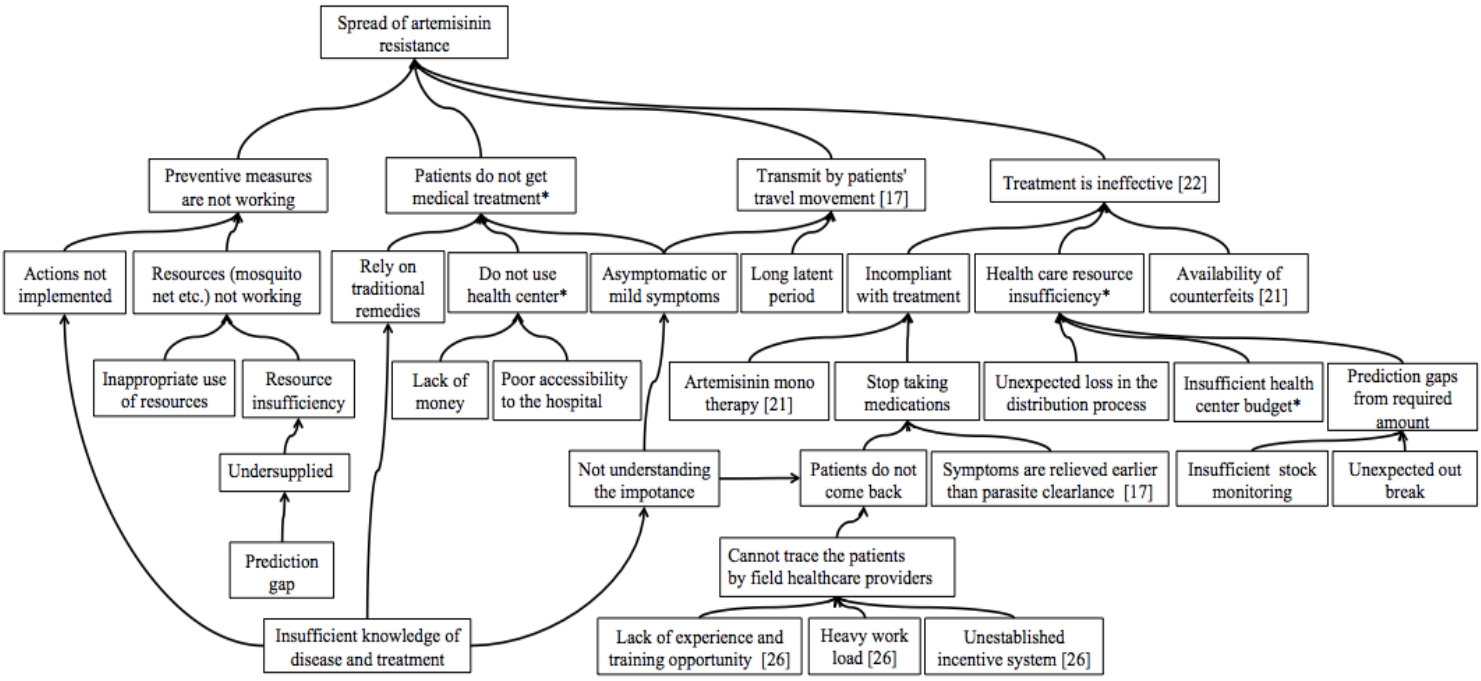


Figure 5 An example of the problem structure related to artemisinin resistance in Cambodia

* Represents findings from the interview

Another factor, seen in this structure is the patient related issues. Patients are in the various circumstances and have respective reasons for not reaching appropriate treatment. Not only the accessibility to the healthcare but also the insufficient knowledge and awareness can be expected for the reasons. Especially for these patient related issues, the effort of local healthcare providers such as village malaria workers, in other words the intervention by field healthcare providers is the last stand to address these problems.

Table 1 Questions and responses at the interview with the local health center staff

No	Questions	Responses
1	How many patients have come to this center? (Today/ This week/ This month)	50 / 350 / 1,500
2	What kind of disease do patients have most?	Fever and Gastroenteritis
3	How about malaria patients, how much is the proportion?	Not a serious issue
4	Can you spare enough time for each patient?	Yes
5	Once patients are treated at this hospital do they come back here to check to ensure they are fully recovered?	Yes
6	For how long do you prescribe antimalarials is it 3 Day, 6 Day or 1 week?	3 days
7	Are health resources and staffs enough to cover all the patients?	Yes
8	What kind of information do you need to provide enough care?	Medical consultation
9	Are you cooperating with village malaria worker well?	Yes
10	Do they have enough knowledge, experience and skills for treating patients appropriately?	Not enough
11	What is the issue in cooperating with village malaria worker?	Need more training and budget
12	How many patients are going to private health care facilities such as private pharmacy or clinics?	Almost 50%
13	Do you need more budget for providing quality care for patients?	Yes
14	Do you think patients will adhere to the treatment more and keep taking medications up to prescribed terms if they are followed up more intimately?	Yes
15	Do you have any other things to worry about?	Worry about some people move their home to other living and education is low.

Interviewee: Mr. Phin Chanmonin (Director of provincial health center in Kampon Chhnang province)

As of February 2015

Figure 6 Provincial health center visited for fieldwork

If we can measure the risk and present it on the map, for areas not covered by current healthcare system sufficiently, the required amount of healthcare resource delivery and staff deployment can be facilitated. Furthermore, for the cross-border transmissions, efficient cross-border screening can be targeted to the high-risk area. As healthcare resources cannot be used inexhaustibly, identification of the target hotspot of malaria endemic area, delivery of the sufficient stockpile of resources and intimate support for field healthcare providers are essential, especially for remote endemic places where the accessibility cannot be retained over a year.

To summarize, it is essential for drug resistant malaria containment activities to let the patients continue the standard quality treatment based on the sufficient health resource distribution. However, the level of the availability of healthcare resources and the support for field healthcare providers differs by area. The widespread mobility of the people in the area is also attributing to this issue. Through the utilization of information from predicted regional disease burdens and the predicted effectiveness of interventions, further steps closer to the effective healthcare resource distributions and the support for the field healthcare providers can be attained.

1.2 Research objective

This thesis studies about the development and evaluation of the system for the modeling and computational simulation approach of the spatial heterogeneity of malaria disease burdens to visualize the effectiveness of containment actions for the purpose of optimal healthcare resource utilization in western Cambodia where the artemisinin resistance was previously reported. The mapping approach with malaria spatial risk distribution modeling is an effective tool and widely used for malaria containment actions, which was established through the long journey of effort by many contributors. However, as the disease burden of malaria decreases along the way malaria elimination effort progresses, this approach needs some adjustment in accordance with the situational changes surrounding malaria elimination effort.

The specific objectives of this study were as follows:

(1) *To develop the spatial risk distribution modeling, adjusted for the current endemic situations, i.e. low-to-moderate malaria transmission settings.*

(2) *To identify the areas at high malaria endemicity where intensive monitoring and support is needed by creating fine-scale risk maps using cross-scale prediction with Bayesian modeling frame.*

(3) *To demonstrate the computational simulations of expected outcomes from given combinations of containment status indicators for evaluating tailored action planning considering regional malaria endemicity.*

1.3 Contributions of this work

Our original contributions are; (1) the demonstration of the mathematical modeling approach which utilizes the reliable measure under the low-to-moderate transmission settings from routine aggregated surveillance reports, (2) cross-scale prediction by modeling framework corresponding to environmental context disparities to create malaria risk maps in fine-scale under the low-to-moderate transmission settings, where a number of countries are or will soon be facing, and (3) the demonstration of different representations of computational simulations from given containment status indicators, which can provide the insight for tailored planning of action alternatives considering regional malaria endemicity.

In spite of numbers of previous publications in this research area, few studies demonstrated the practical applications of developed risk modeling frame. This study covers not only the spatial risk distribution modeling using available dataset but also the application of developed modeling frame by predicting expected outcomes using computational simulations. Figure 7 represents the typical program phase of malaria elimination. The low-to-moderate transmission setting is the situation where the many of the countries aiming for malaria elimination will soon be facing and thus an important step toward the goal. At this stage, the program reorientation is needed for moving the program effort forward. The approach and implications described here will provide more efficiency for countries under such situations.

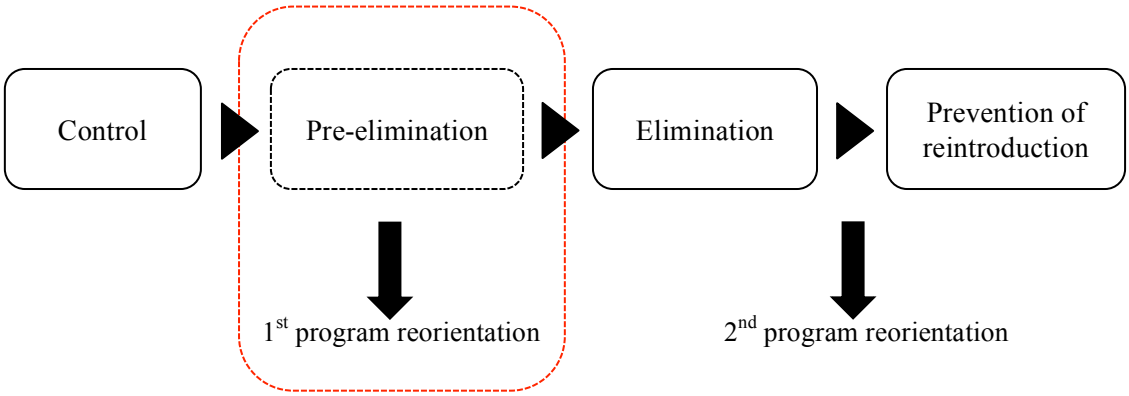


Figure 7 Program phase and major interventions in the malaria elimination program [28]

1.4 Chapter organization

This thesis consists of following chapter organizations:

Chapter 2 reviews related studies in the area of the mapping approach of malaria disease burdens with spatial risk distribution modeling. Besides, the author explains the limitations of these approaches under the current situational changes in the malaria problem and proposes modified method to address these issues.

Chapter 3 describes the design processes and products of the system of interest. Author provides the requirement analysis and the system architecture using the diagrams developed through this system design process.

Chapter 4 presents the development and evaluation of the spatial risk distributions model using Bayesian modeling frame corresponding to the environmental context disparities for cross-scale prediction.

Chapter 5 presents the fine-scale mapping with the spatial risk distributions model developed in chapter 4. This approach was evaluated through the comparison with previously reported maps and geocoded case data.

Chapter 6 presents the demonstrations of the computational simulations under the given combinations of containment status indicators. Author proposes the examples of how these results can be utilized for the malaria problems.

Chapter 7 reviews the verifications and validation of the system of interest. Based on the requirement developed discussed in chapter 3, author summarizes the Requirement Verification Traceability Matrix of the system. Furthermore, the stakeholders' review for the system validation is provided.

Chapter 8 discusses the results and findings throughout this study. Author provides the implications from development and evaluation processes for future applications of study results. Furthermore, several limitations of this research are discussed.

Chapter 9 concludes this thesis. This chapter summarizes contributions of this study and describes the future study directions.

2. LITERATURE REVIEW AND THE PROPORSED APPROACH ADDRESSING EXISTING ISSUES

2.1 Spatial prediction map of malaria

Recent efforts for the quantification of risk burden and the creation of the spatial prediction map of malaria risk made substantial contributions to address the needs for targeted interventions for malaria [29-30].

2.1.1 The framework for geospatial science applied to malaria elimination

Figure 8 shows the framework for geospatial science applied to malaria elimination [30]. Various kinds of data can be used such as the intervention coverage and infrastructure and target residences in view of operations. Also the remote sensing data, meteorological data from environmental side as well as malaria surveillance, survey and entomogical data can be used for the malaria quantifications. By integrating these data, the geographical information systems (GIS) can be developed, then, the spatial statistical analysis can be conducted so that these data can be used for malaria containment activities as outputs. By predicting areas at risk and examining the effectiveness of interventions from estimated risk at the target hotspot, some more steps closer to the efficient resource allocation can be attained. Among the outputs described in this frame, our focus was providing spatial prediction map. Moreover, model-based simulation approach was demonstrated, which is one of our original contributions.

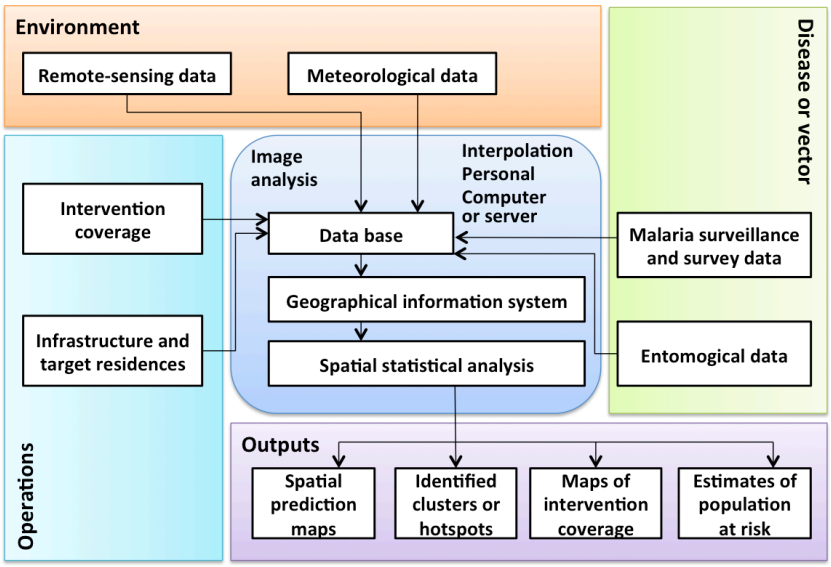


Figure 8 Framework for geospatial science applied to malaria elimination [30]

2.1.2 The world map of malaria endemicity

In that context, world map of *P. falciparum* malaria endemicity was published using parasite rate (PR) surveillance report and the model-based geostatistical approach (Figure 9-10) [31-32]. These procedures were implemented within a Bayesian statistical framework to represent the uncertainty in the unknown map while retaining robustness of these predictions [33].

These mapping products and methodologies for spatial risk distribution modeling are provided by Malaria Atlas Project, which is aiming to disseminate free, accurate and up-to-date information on malaria. The team in the university of Oxford is receiving designation as a World Health Organization (WHO) Collaborating Centre in geospatial disease modeling.

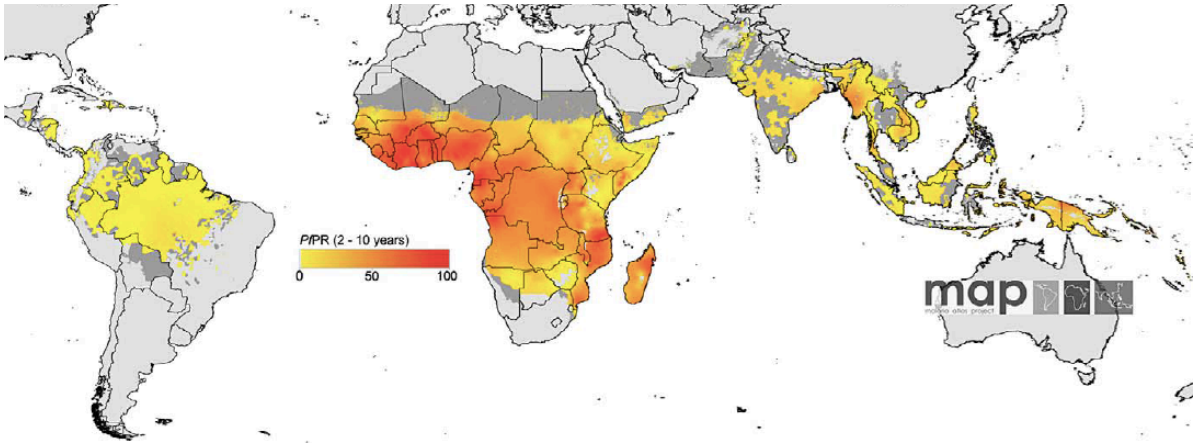


Figure 9 The Spatial distribution of *P. falciparum* malaria endemicity by Malaria Atlas Project [31]

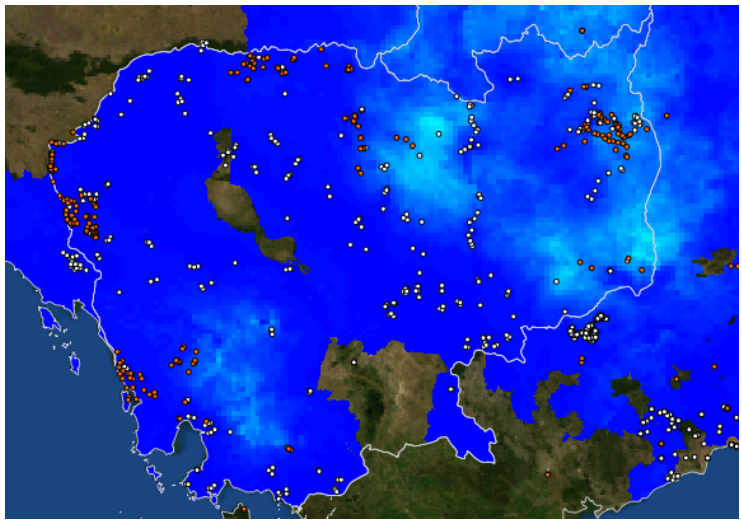


Figure 10 The Spatial distribution map of *P. falciparum* in Cambodia [31]

2.1.3 Application of remote sensing data from space satellites

The remote sensing techniques are the powerful tools to identify the hotspot and investigating the malaria epidemiology [3]. Several environment related indices calculated from remote sensing data, such as normalized difference vegetation index (NDVI), normalized difference water index (NDWI) and topological wetness index (TWI) were utilized to predict the regional malaria endemicity [34-37]. Climate is also closely related to the risk of malaria [38-39]. Cohen et al. created fine-scale risk maps of both high endemic and low endemic seasons from routine aggregated case report using meteorological and remote sensing data (Figure 11) [40].

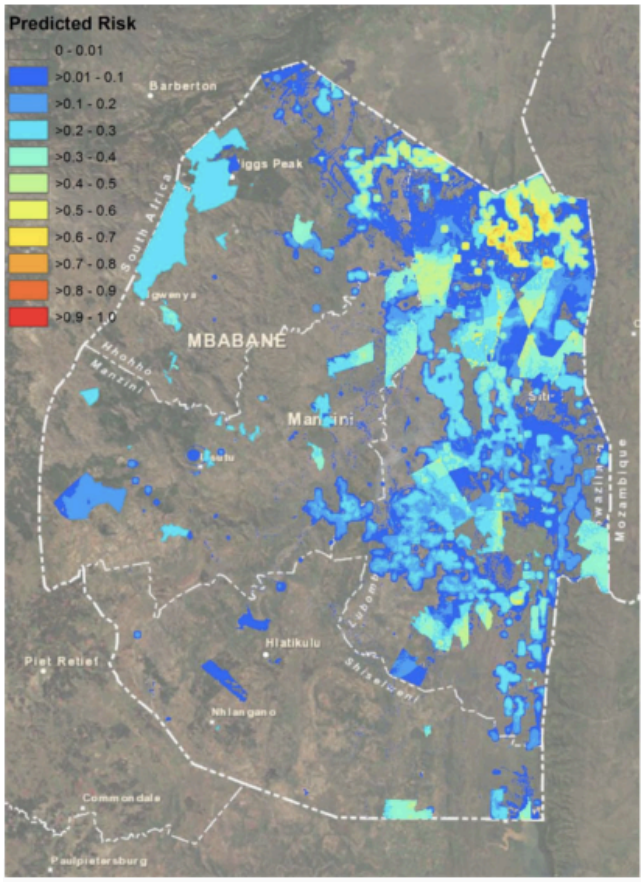


Figure 11 Predicted probability map for malaria cases in Swaziland [40]

2.1.4 Human interactions with surrounding environment

It is important to take human interactions with environment into the model when describing the risk as malaria is transmitted through the interaction with vectors lurking

in the surrounding environment. In fact, several behavioral factors and human reactivity to this disease are incorporated in the mathematical model for the prediction of malaria transmission [41-42]. From this perspective, the map represents the model incorporating non-environmental anthropogenic factors were reported [43]. Also with related to this topic, Ellis et al. proposed the concept of anthropogenic of the world (Figure 12) [44]. This concept is, in short, the conceptual thinking of "putting people in the map", meaning environmental factor together with anthropogenic factor such as population density may explain how the people use or interact with surrounding environment, which is expected to relate to the geographically observed phenomenon.

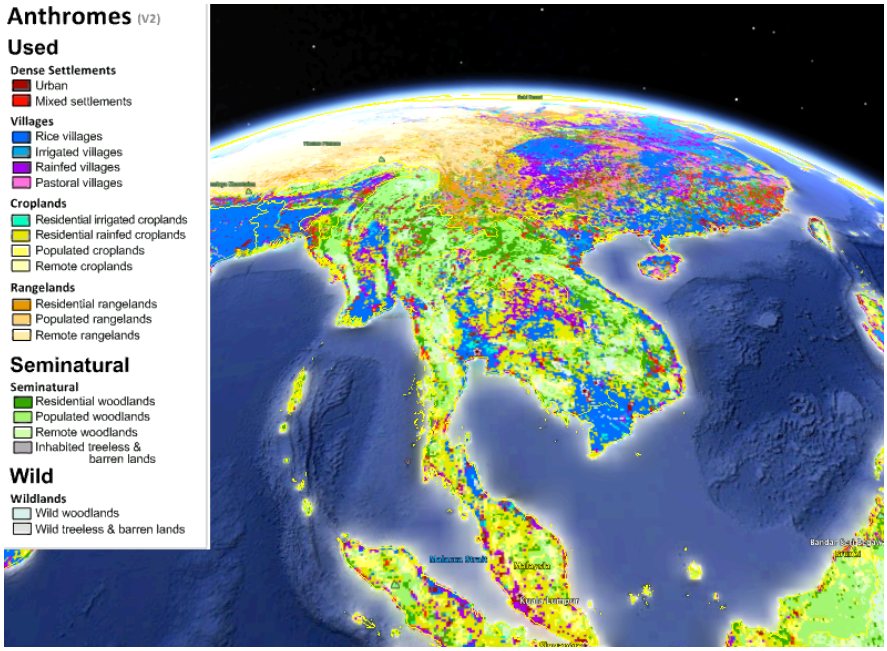


Figure 12 Anthropogenic biome visualized in the Google Earth© [44]

2.1.5 Quantifications of the impact of human mobility

It is also important to consider the influence of the human mobility for the malaria transmissions. Using spatially explicit mobile phone data and malaria prevalence information, Wesolowski et al. projected the source and sink of malaria parasites in Kenya (Figure 13). By that, they could identify the dynamics of human carriers that may drive parasite importation between regions [2].

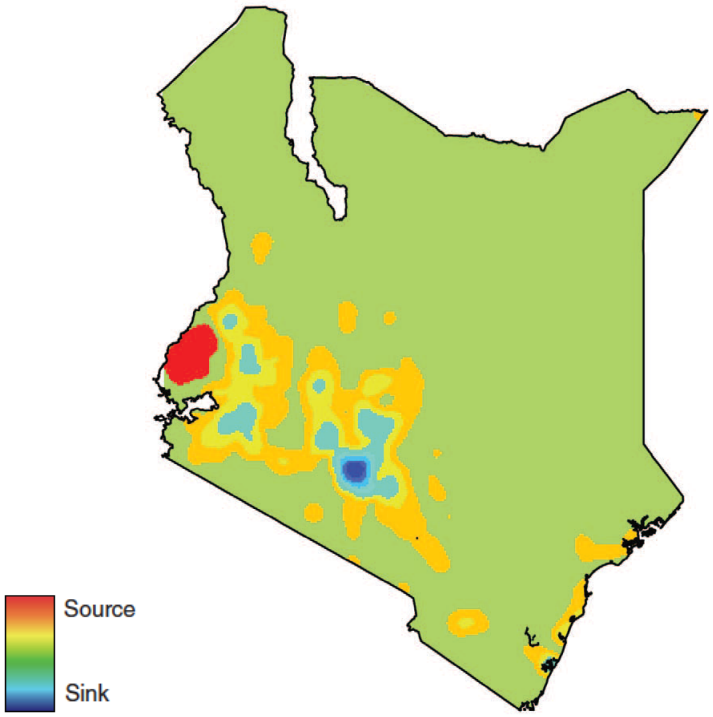


Figure 13 Projected source and sink of parasites in Kenya [2]

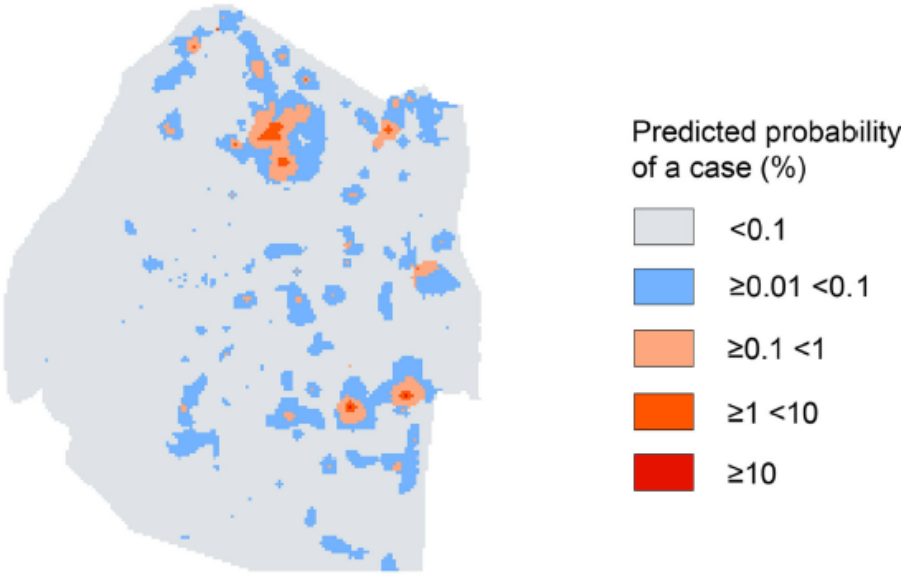


Figure 14 Fine-scale risk map created by cross-scale prediction in Swaziland [49]

2.2 Limitations of existing malaria mapping and measurement approach

Despite these advancement, as the prevalence of malaria decreases along the way malaria elimination effort progresses, this risk mapping approach needs some adjustment in accordance with the situational changes.

2.2.1 Measurement of the malaria risk

In terms of the measurement of malaria disease burden, when malaria becomes rare, it becomes increasingly difficult to detect ongoing transmission monitoring by PR [45]. Since, these situations are the important steps toward malaria elimination, there is an important need for examining the modeling method of disease burden under low to moderate transmission settings. Annual Parasite Incidence (API) can be a reliable measure for reporting new malaria infections under these settings supported by good reporting systems [46]. However, intensive focused screening method indicated that, in low-transmission settings, not a few malaria cases are asymptomatic, which makes it difficult to identify all the cases by passive surveillance systems [47-48].

2.2.2 Accessibility for fine-scale data

Under low transmission settings, because of few infectious cases reported, the sample size required for both estimation and spatial prediction of infection prevalence becomes very large and such information, usually, cannot be obtained in fine-scale. Instead of using such data, cross-scale prediction with the data collected in the coarser scale is used while managing the spuriousness and unreliability of prediction by the Bayesian modeling framework (Figure 14) [49].

2.2.3 Assessment of the effectiveness of combinations of interventional measures

At present, there are a few examples of investigations for the benefits of combining different vector control measures, but further studies are needed about the assessment of the effectiveness of using these combined approaches. This issue has become increasingly important, as “One-size fits all” approach is no longer applicable to the areas under the low-to-moderate transmission intensity. All interventions should be reviewed carefully and should be tailored for regional circumstances in an ongoing way to ensure that they remain fully effective and cost-effective.

2.3 Hypothetical questions and proposed approach

Based on the current situations and findings from previous studies we propose following hypothetical questions:

(1) *Can we use standardized morbidity ratio (SMR) calculated by API in the spatial risk prediction model as an appropriate measure of disease burdens?*

(2) *Can we predict the fine-scale risk better if human interactions with surrounding environment, i.e. environmental context, are considered?*

(3) *Can we demonstrate the expected outcome of interventional measures by incorporating containment status indicators into the risk prediction model?*

Here, we applied a mathematical modeling approach for SMR calculated by API using routinely aggregated surveillance reports and variables related to human interactions to surrounding environment to create spatial risk distribution maps in fine-scale of two provinces (Pailin and Preah Vihear) in western Cambodia where the artemisinin resistance was previously reported. In addition, we incorporated the combinations of containment status indicators into the model, by which the regional heterogeneities of the relationship between containment status and risk can be visually represented for the efficient healthcare resource allocations and intervention planning considering temporal descriptions of regional malaria endemicity.

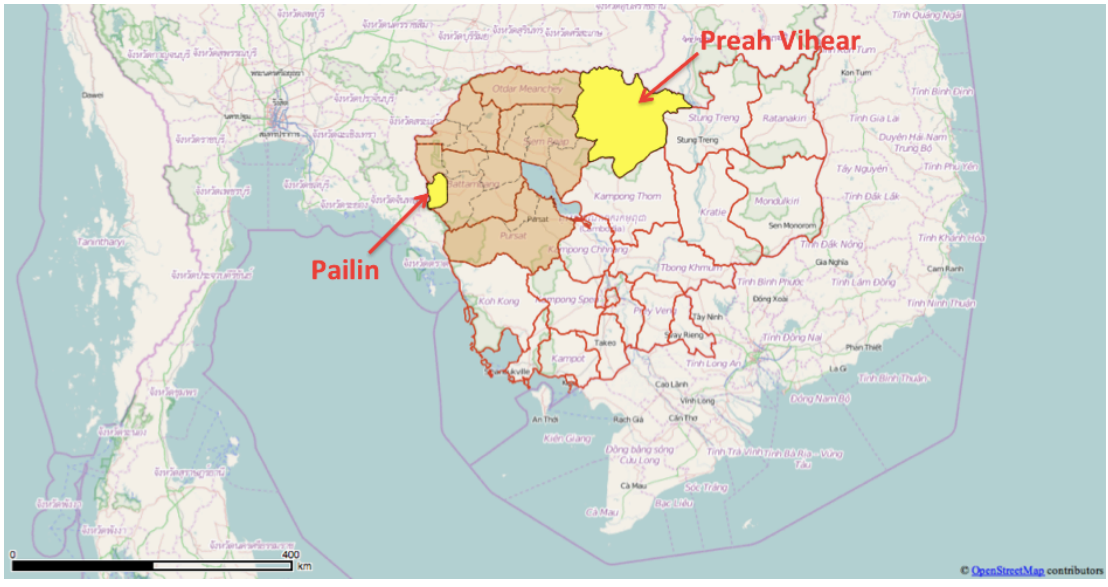


Figure 15 A map of the research area

Open Street Map© was used to create this map

2.4 The environmental context disparities

Human interactions with surrounding environment are the important factors affecting the risk especially for communicable diseases transmitted through infectious vectors. Figure 16 shows an applied case of the epidemiologic triad of disease causation [50] for the malaria case. This triad consists of an external agent (Malaria), a host (Human) and an environment in which host and agent are brought together, causing disease. Vector plays as the transmission carrier of malaria parasite that does not present malaria symptoms.

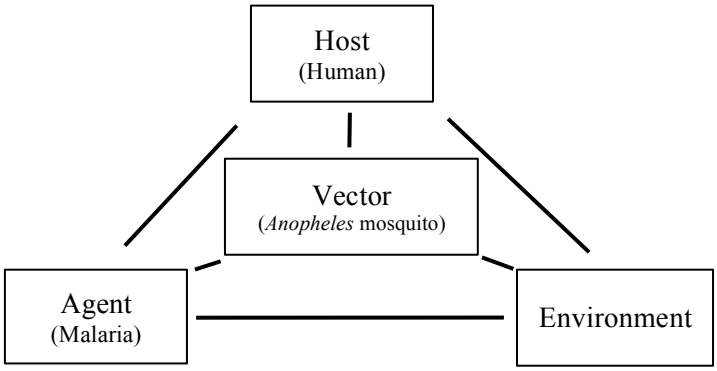


Figure 16 An applied case of the epidemiologic triad of disease causation for malaria

Based on this conceptual thinking, the environmental context surrounding human communities, i.e. how the environmental features exist and interact with people, is an important factor affecting the risk of this infectious disease. Thus, disparities in environmental context among communities could explain the extent of disease burden across areas. In this study, we defined this concept as "**The environmental context disparities**" as a key factor for the development of malaria risk model.

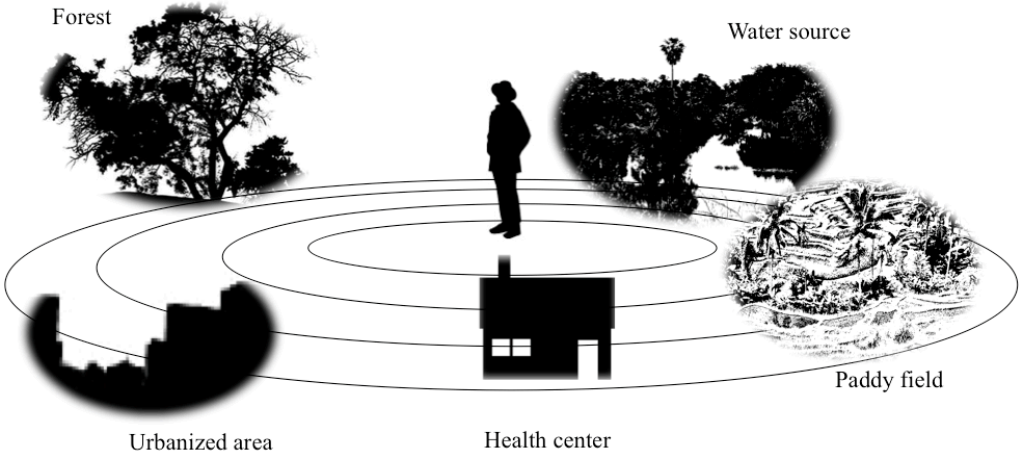


Figure 17 An example of environmental context

3. SYSTEM DESIGN FOR THE MALARIA RISK MODELING APPROACH

In this chapter, we designed the system for the modeling approach of the spatial heterogeneity of malaria disease burdens and its implementation called “Malaria risk modeling and simulation system”. The purpose of this chapter is to examine the key stakeholders, the context, the behavior, the requirement and the architecture of the system. Based on that, the system was verified and validated for evaluations. At the same time, early design products could be validated through the fieldwork and the interview with key local stakeholders.

3.1 Use case analysis

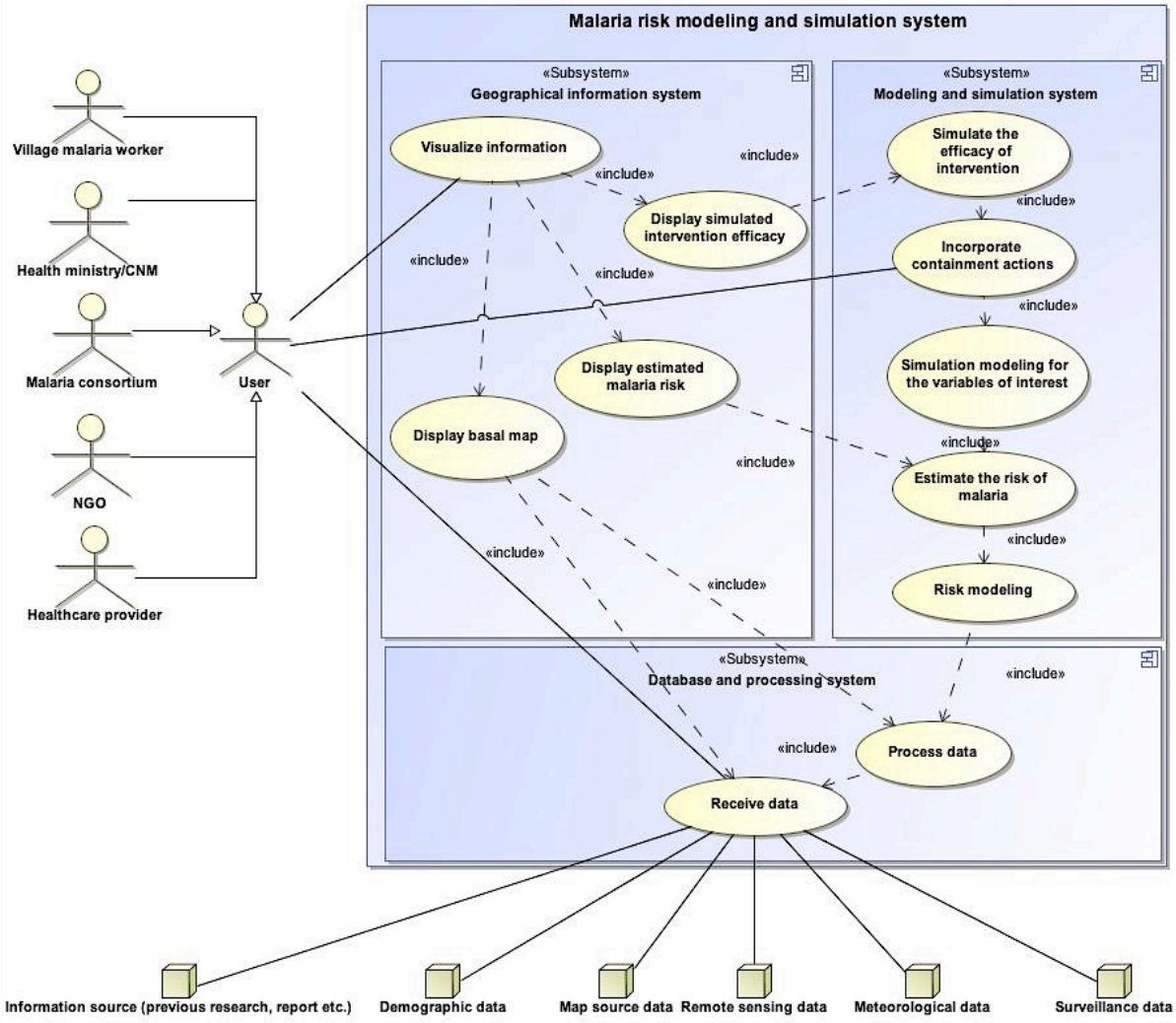


Figure 18 Use case diagram of the system

The use case diagram of the system is shown in Figure 18. The system collects various information from data sources and processes them to make malaria risk prediction model. The output the system is visualized on the map as GIS, which makes users easily identify the areas at high risk. Identified key stakeholders are as follows:

Village malaria worker (VMW)

Public officer such as staff in Ministry of Health

Organization engaged in malaria containment actions such as Malaria consortium

Non-Government Organization (NGO)

Health care provider at regional health center

Therefore, expected users are the professional healthcare provider and related party. Figure 19 shows the context diagram of the system. This diagram shows the primal function seen from each stakeholder, i.e. the interface of the system could be examined in this diagram. Displayed information plays as the key output of the system.

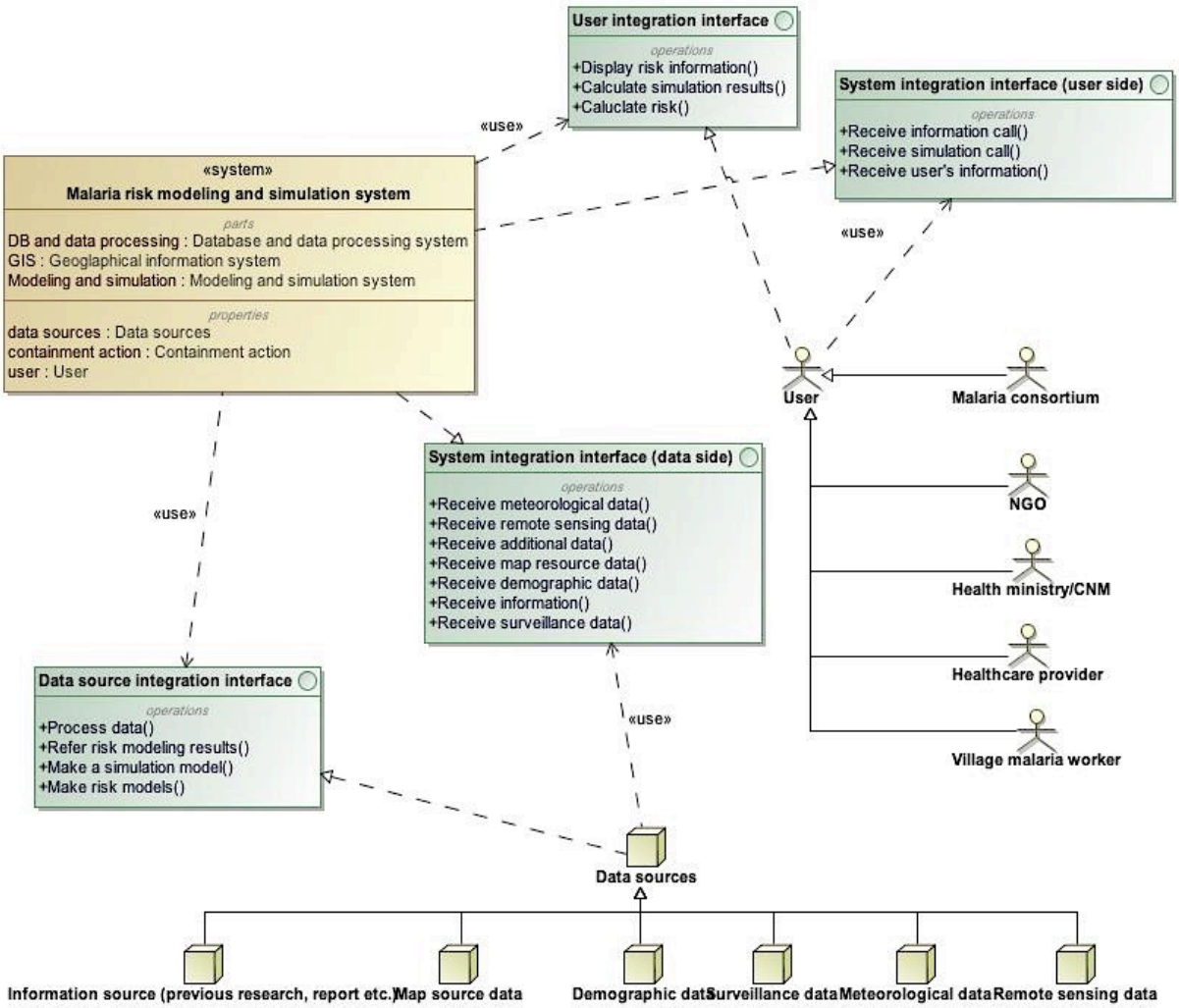


Figure 19 Context diagram of the system

3.2 System function

Based on the information from use case analysis and the context diagram, the sequence diagram for examining the interaction between the system and stakeholders was drawn.

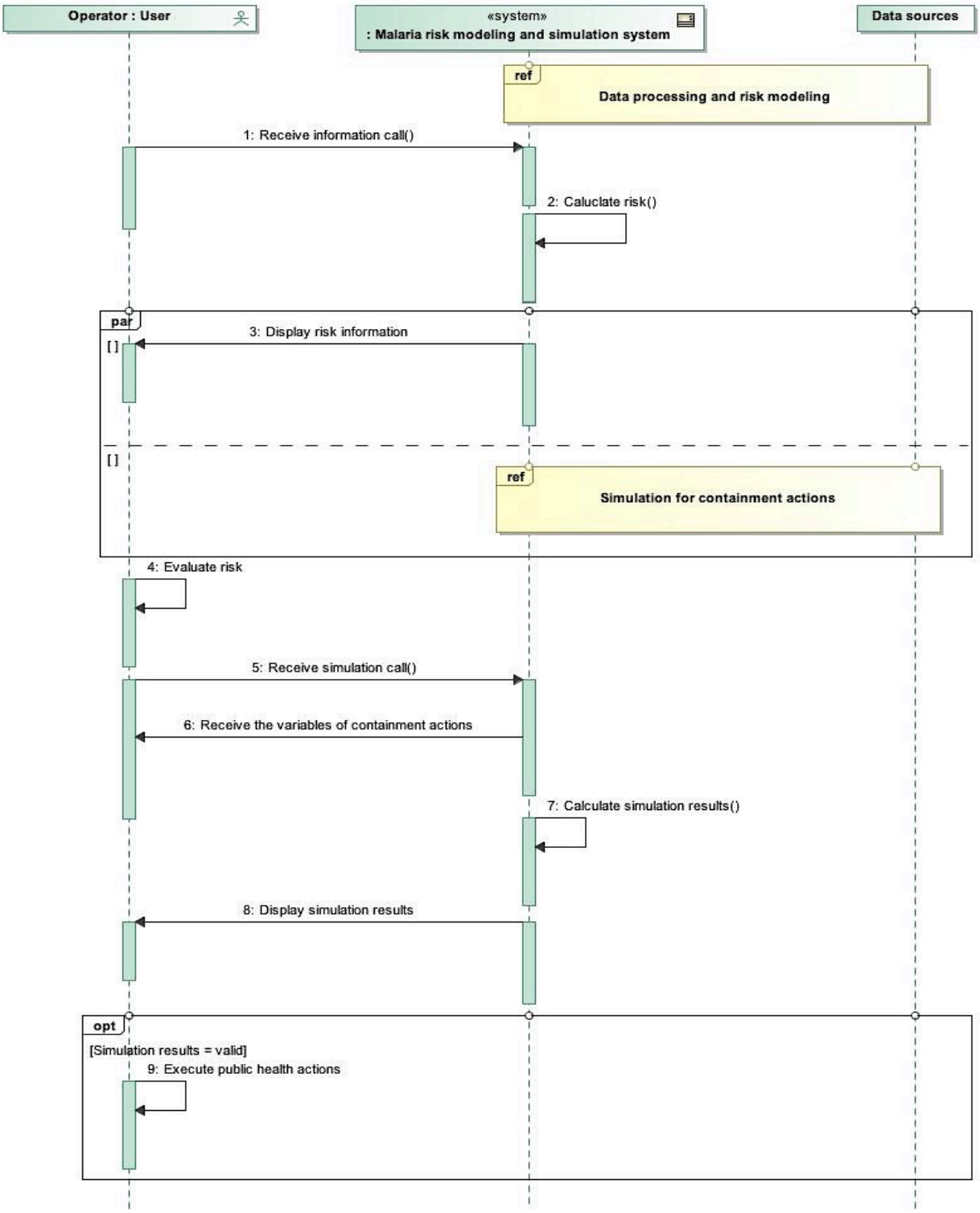


Figure 20 Sequence diagram of the system

Figure 20 shows the concept of operation describing the interface between users and the system. The major insight from this analysis was the system interacts with users multiple times to collect the data, by which the system shall display the output as GIS. As users need to take practical actions for malaria containment, the displayed information needs to be interlinked to these actions, also the extent of regional risk should be clearly identified not only by decision makers such as the government officer but also the regional practitioner. Thus, the scale of the system output needs to cover both wide and fine-scale. In that sense, the GIS platform can provide both wide and fine-scale view.

The functional flow block diagram (FFBD) shown in Figure 21 provides the required system function extracted from this operational sequence.

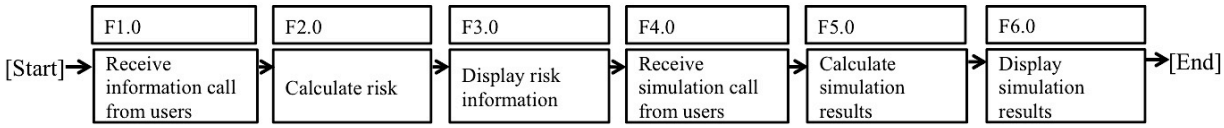


Figure 21 Functional flow block diagram

3.3 Requirement analysis

The requirement diagram provides a bridge between the typical requirements management tools and the system models [51].

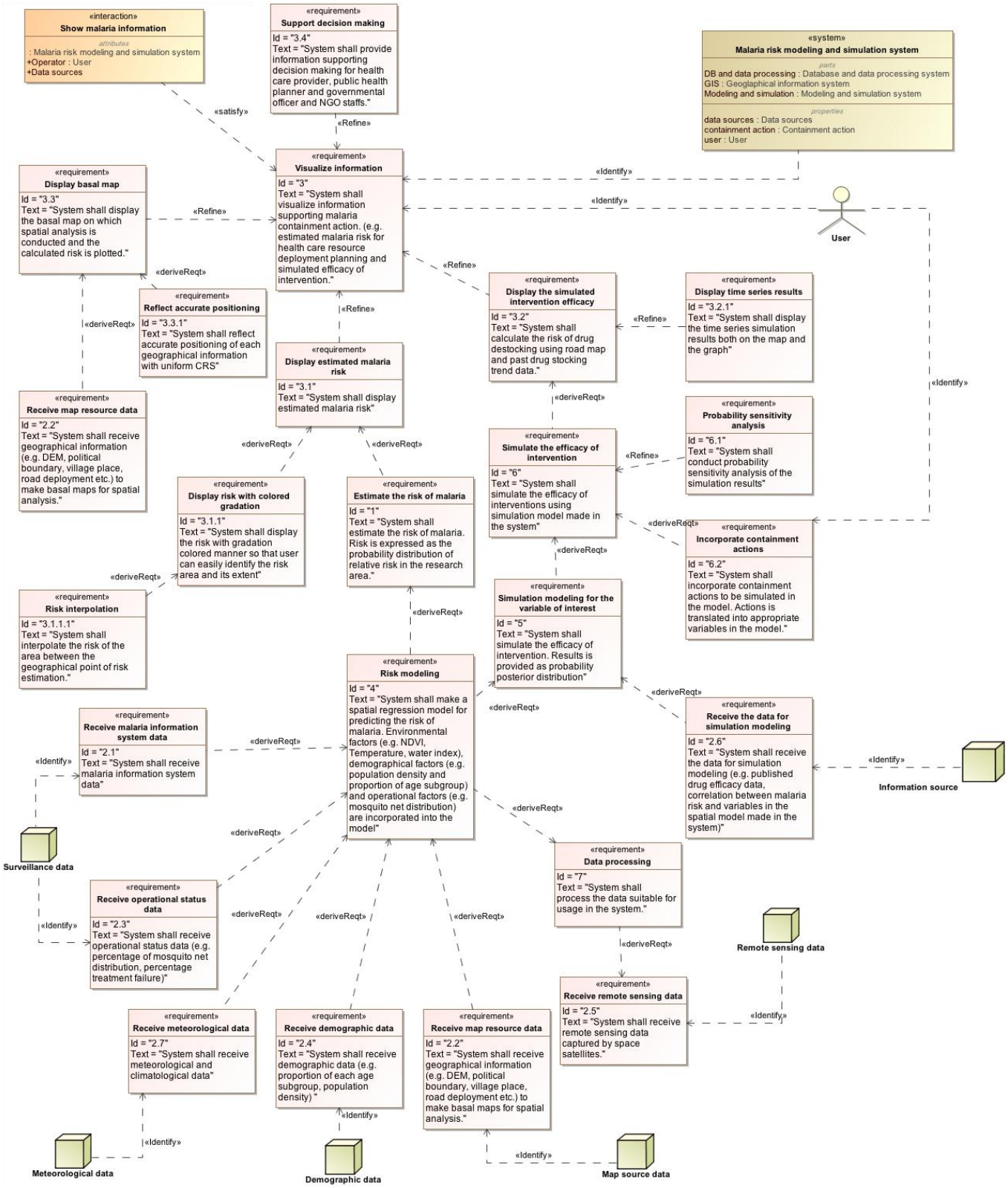


Figure 22 Requirement diagram of the system

Refined requirements indicate the needs for presenting risk with colored gradation to make it more visible interlinking to actions.

Table 2 Requirement table of the system

#	ID	Requirements	Description
1	1	Estimate the risk of malaria	System shall estimate the risk. Risk is expressed as probability distributions in the research area.
2	2	Receive data	System shall receive data provided by stakeholders and other information sources.
3	2.1	Receive malaria information system data	System shall receive malaria information system data
4	2.2	Receive map resource data	System shall receive geographical information (e.g. DEM, political boundary, village and road etc.) to make basal maps for spatial analysis.
5	2.3	Receive containment status data	System shall receive containment status data (e.g. percentage of mosquito net distribution and treatment failure).
6	2.4	Receive demographic data	System shall receive demographic data (e.g. proportion of each age subgroup and population density)
7	2.5	Receive remote sensing data	System shall receive remote sensing data captured by space satellites.
8	2.6	Receive the data for simulation modeling	System shall receive data for simulation modeling (e.g. published drug efficacy data, correlation between malaria risk and variables in the spatial risk model made in the system)
9	2.7	Receive meteorological data	System shall receive meteorological and climatological data.
10	3	Visualize (Provide) information	System shall visualize information supporting malaria containment actions (e.g. estimated malaria risk for healthcare resource deployment planning and simulated efficacy of intervention).
11	3.1	Display estimated malaria risk	System shall display estimated malaria risk.
12	3.1.1	Display risk with colored gradation	System shall display the risk with gradation colored manner so that user can identify the risk area and its extent.
13	3.1.1.1	Risk interpolation	System shall interpolate the risk of the area between the geographical points of estimation.
14	3.2	Display the simulation results	System shall display the result of computational simulations.
15	3.2.1	Display time series results	System shall display the time series of simulation results both on the maps and graphs.
16	3.3	Display basal map	System shall display the basal map on which spatial analysis is conducted and calculated risk is plotted.
17	3.3.1	Reflect accurate positioning	System shall reflect accurate positioning of each geographic information with uniform coordinates reference system (CRS).
18	3.4	Support decision making	System shall provide information supporting decision making for healthcare providers, public health planners, government official and NGO staffs.

19	4	Risk modeling	System shall make spatial regression modeling for predicting the risk of malaria. Environmental factors (e.g. NDVI, Water indices and Temperature), demographical factors (e.g. Population density and Proportions of each age subgroup) and containment status indicators (e.g. Mosquito net distributions) are incorporated into the model.
20	5	Simulation modeling for the variables of interest	System shall simulate the expected outcomes using combinations of containment status indicators. Results are provided as probability distribution.
21	6	Simulate the efficacy of intervention	System shall simulate the efficacy of interventions using simulation model made in the system.
22	6.1	Probability sensitivity analysis	System shall conduct the probability sensitivity analysis of the simulation results.
23	6.2	Incorporate containment actions	System shall incorporate containment actions simulated in the model. Actions are translated into appropriate variables in the model.
24	7	Data processing	System shall process the data suitable for usage in the system.

3.4 System architecture

Based on the information from the analysis, candidates of the system architecture were considered and selected as below (Figure 23). Then the system functions were allocated to each subsystem.

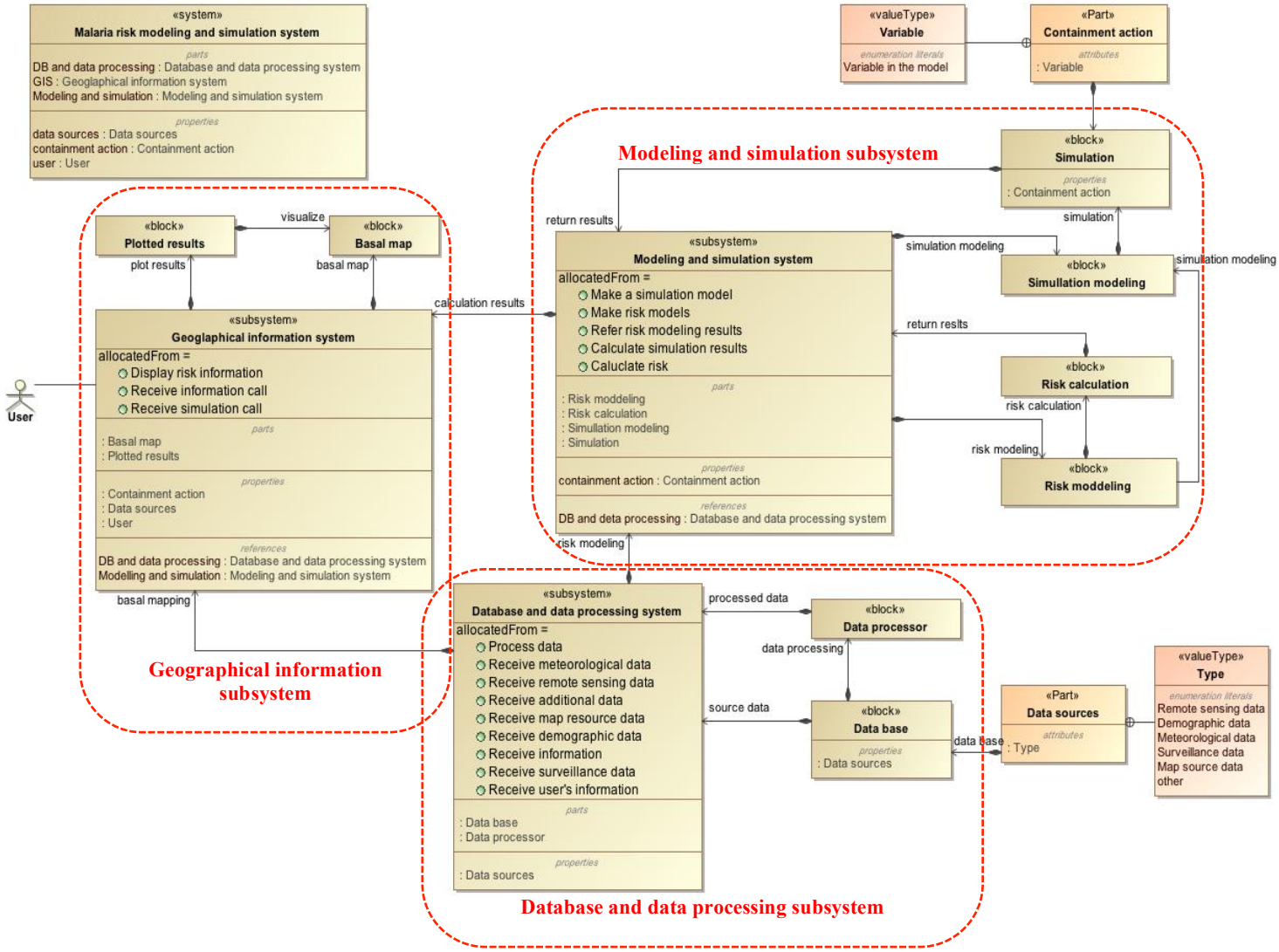


Figure 23 Block definition diagram of the system

The system consists of three subsystems, the database and processing subsystem, the modeling and simulation subsystem and the geographical information subsystem. These subsystems act in conjunction to support the malaria containment actions by each stakeholder.

Data are received at the database and processing subsystem and processed to the appropriate variables used in the model. This subsystem acts as database collecting data continuously and provides up to date information used for the spatial risk distribution modeling. Geospatial analysis (i.e. modeling and simulation) is conducted in the modeling and simulation subsystem. Parameters are combined to calculate the risk in accordance with risk calculation model. Then the results are transmitted to the geographical information subsystem that displays the risk easily identifiable for users. These subsystems work separately and in conjunction with each other, which let the system have more reliable sustainment of the traceability of each component and requirements. Furthermore, each component can be replaced in accordance with situational changes in the environment and advancement in the technologies.

3.5 Customer value chain analysis

An analysis of expected value flow was performed with Customer Value Chain Analysis (CVCA) as shown in Figure 24.

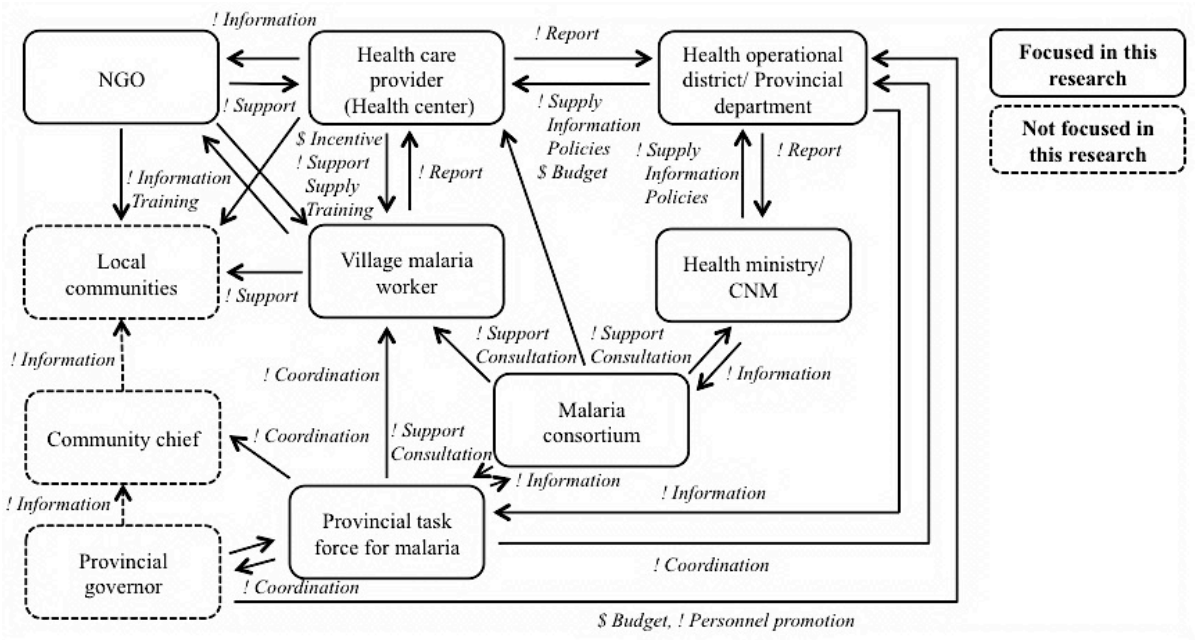


Figure 24 Customer value chain analysis

Based on the analysis, it was assumed that the budget and supply of diagnostic testing kit and medicines are strongly controlled by central government even for the field health care providers such as VMW and regional health center. Whereas the support for such field health care providers relies on the information based on the report

from field. The reporting pathway is the opposite direction of resource supply. As several stakeholders are involved in this reporting pathway, the reliability of data should be carefully monitored from the fact that there is a possibility that the falsification and misreporting occur. Whereas the importance of data reliability, the work load of field healthcare providers is sometimes intensive so that the timely and appropriate support is critical to attain the expected outcome from the current patient care system for malaria. For instance, intensive supervising during high malaria endemic seasons, trainings, educations and designing incentive system for offering the reward commensurate with their contribution can be considered. In any event, it is assumed that by prediction of regional emdemicity and expected outcomes attained from targeted containment status such resource optimization can be much more facilitated. Especially, it is important for the low-to-moderate transmission setting like in Cambodia, as the budget and healthcare resource needs to be managed within the reasonable extent while achieving the effect of as much as possible.

3.6 Summary of the system design

Through the system design process described here, following design products were identified. These products provided not only the insight for the current malaria care in Cambodia and the requirement for the system but also an opportunity to validate the concept design with key stakeholders at early development phase and the direction for the system verification and validation. Details of the system verification and validation are reviewed in chapter 7.

(1) *Stakeholders list*
(2) *Concept of operation*
(3) *Requirement table*
(4) *System architecture*
(5) *Customer value chain*

4. DEVELOPMENT AND EVALUATION OF THE SPATIAL MALARIA RISK MODEL

4.1 Malaria data collection

Malaria data were collected from Cambodia malaria bulletin report in 2010 to 2013 [52-53]. This comprehensive dataset was made from the case report collected through the effort of the malaria information system (MIS) and the national facility-based health information system (HIS) using common coding system [54] and contains API (per 1000 people) in each health operational district for two malaria species (*P.falciparum* and *P.vivax*) from two respective reporting pathways (by healthcare facilities or village malaria workers), which is reported periodically by National Center for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia. SMR, standardized mortality or morbidity ratio is a quantity, expressed as a ratio or percentage of quantifications compared with the general population of interest (equation 1, 2) [55].

$$SMR = \hat{\theta}_i = \frac{o_i}{e_i} \qquad (1)$$

$$e_i = \sum_k n_{ik} P_k \qquad (2)$$

Where, o_i : observed case number in *i* area, e_i : expected case number in *i* area, n_{ik} : population of *k* age group at *i* area and P_k : incidence in *k* age group in reference population. e_i was estimated by multiplying age population and reported incidence in each age group in western-Cambodian 10 provinces [56]. SMR, $\hat{\theta}_i$ in given *i* district, was then calculated dividing API by e_i per 1,000 people. Assuming small case number and relatively large dispersions under the low-to-moderate transmission settings, the case count data can be assumed to follow the negative binomial distribution, $o_i \mid \mu_i$, where μ_i is the corresponding distribution mean and ρ is the scale parameter (equation 3). With this, by transforming equation 1, μ_i can be derived by multiplying e_i and the relative malaria risk, $\hat{\theta}_i$ (equation 4) [57].

$$o_i \mid \mu_i \sim \text{NegBin}\,(\mu_i, \rho) \qquad (3)$$

$$log\ \mu_i = log\ e_i\ +\ log\ \hat{\theta}_{\ i} \qquad (4)$$

Considering the small number of observed cases compared with population size and the modifiable areal unit problem in geographical analysis, SMR for each health operational district was smoothed by empirical Bayesian method (EBSMR) [58] for adjusting the influence of different population size in area units. EBSMR was calculated by equation 5-7, given that $\hat{\theta}_{\ i}$ follows gamma distribution (equation 8) and observed o_i under θ_i follows Poisson distribution (equation 9).

$$EBSMR = \hat{\theta}_{\ i} =\ e[\theta_i | o_i, e_i] = \frac{o_i + \upsilon}{e_i + \alpha} \qquad (5)$$

$$\frac{\hat{\upsilon}}{\hat{\alpha}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{\ i} \qquad (6)$$

$$\frac{\hat{\upsilon}}{\hat{\alpha}^2} = \frac{1}{n+1}\sum_{i=1}^{n}\left(1 + \frac{\hat{\alpha}}{e_i}\right)\left(\hat{\theta}_{\ i} - \frac{\hat{\upsilon}}{\hat{\alpha}}\right)^2 \qquad (7)$$

$$\theta_i \sim Ga(\upsilon, \alpha) \qquad (8)$$

$$o_i | \theta_i \sim Po(\theta_i, e_i) \qquad (9)$$

4.2 Environmental and non-environmental anthropogenic predictor variables

Predictor variables incorporated into the modeling framework were described in Table 3. The normalized difference vegetation index (NDVI), the normalized difference water index (NDWI) and the land surface water index (LSWI) were calculated from Terra-MODIS 8-day composite data (http://LPDAAC.usgs.gov) from 2010 to 2013. As EBSMR was represented as yearly average, these environmental variables were averaged to the mean values for each year. NDVI, an index correlating with the extent of vegetation and used for forest monitoring was calculated using the reflectivity of red in visible range (R) and near infrared radiation range (IR) collected by satellite sensor (equation 10). For the MODIS satellite, IR corresponds to band 2 and R corresponds to band 1 (equation 11).

$$NDVI = \frac{IR - R}{IR + R} \qquad (10)$$

$$NDVI = \frac{(Band2) - (Band1)}{(Band2) + (Band1)} \qquad (11)$$

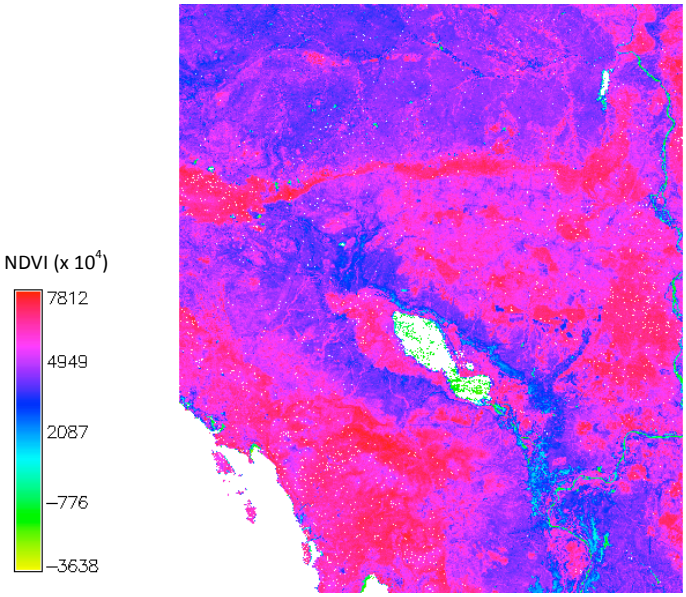


Figure 25 Example of NDVI calculation using MODIS satellite data

Note: NDVI values were multiplied by 10^4.

In the same manner, NDWI and LSWI, indices of water, can be calculated by combinations of the reflectivity of different wave lengths (equation 12-13).

$$NDWI = \frac{(Band2) - (Band5)}{(Band2) + (Band5)} \qquad (12)$$

$$LSWI = \frac{(Band2) - (Band6)}{(Band2) + (Band6)} \qquad (13)$$

The digital elevation model at 30 m resolution was extracted from ASTER GDEM database (http://gdem.ersdac.jspacesystems.or.jp) [59] and used to estimate the altitude. The topographic wetness index (TWI) was calculated using this altitude model and estimated by the method described in previous report [60]. Considering interactions between surrounding environment and people in the malaria transmission process, we extracted these data from multiple surrounding circular buffers with different radius distance (For each 1km from 1 to 5km) from villages, that could potentially indicate the human interactions with surrounding environment, and compared them by correlation efficient of the models. Distance and surrounding circular buffers were generated by the Quantum GIS software. As temperature influences the ecology of mosquito breeding habitat, i.e. malaria transmission [38], the *Plasmodium* temperature suitability index [61] was extracted from Malaria Atlas Project database [62]. The rapid urbanization is related to the change in the risk patterns of malaria transmission compared with rural sparsely populated areas [63-64] and the susceptibility of these populations is influenced by implementations of containment actions. Population density per km^2 was calculated using the record in Cambodia Malaria bulletin divided by areas of each health operational district as a variable reflecting the extent of urbanization. Besides, we used the reported proportion of sufficient ownership of long lasting insecticide-treated nets (LLIN) [56] and treatment failure rate of artemisinin (TF_{rate}) [65] as containment status indicators. Sufficient ownership of LLIN ($LLIN_{suf}$) is defined as proportion of household in which distributed mosquito net covers 2 persons or less per net. As no geographical localities could be obtained for these status indicators, they were aggregated to the provincial level and incorporated in the model development.

Table 3 Variables used to build the modeling framework to estimate EBSMR

Category	Variable	Data source	Data collection
Vegetation	NDVI	Terra-MODIS 8-day composite data 2010-2013	Extracted mean value from 1, 2, 3, 4, 5 km surrounding circular buffer from each populated village
Water	NDWI	Ditto	Ditto
	LSWI	Ditto	Ditto
Geography	TWI	Digital elevation model at 30 m resolution from ASTER GDEM database [59]	Ditto
Temperature	*P. falciparum* Temperature suitability index (*Pf*TSI)	Malaria Atlas Project database [62]	Averaged to mean value for each HOD
Population	Population density (/km^2)	Cambodia malaria bulletin report 2010-2013 [52-53]	Population record divided by total areas of each HOD
Vector control	Sufficient ownership of LLIN [a]	Cambodia malaria survey 2010 [56]	Used the values reported at each provincial level
Treatment	Treatment failure rate by artemisinin combination Therapy [b]	National Center for Parasitology, Entomology and Malaria Control [65]	Ditto

[a] Proportion of household in which distributed mosquito net covers 2 persons or less per net.

[b] Test positive for *P. falciparum* on day 28 or day 42

EBSMR, Standardized morbidity ratio estimated by empirical Bayesian method; NDVI, Normalized difference vegetation index; NDWI, Normalized difference water index; LSWI, Land surface water index; LLIN, Long lasting insecticide-treated net; Topographical wetness index; HOD, Health operational district

4.3 Spatial risk distribution modeling

The relationship between EBSMR ($\hat{\theta}$) and predictive variables was modeled using a generalized linear regression model as a function of the N predictive variables (X, Z), given that the logarithmic $\hat{\theta}$ follows the Gaussian distribution.

$$\hat{\theta} = e^{\lambda} \quad (14)$$

$$\lambda = \alpha + \sum_{N} \beta_N X_N + \sum_{N} \gamma_N Z_N + \varepsilon \quad (15)$$

Where α is the model intercept, β is the parameter associated with environmental covariates *X* and γ with non-environmental anthropogenic covariates *Z*. The likelihood of observed data given to the model and the input predictors were calculated based on this modeling frame (equation 14-15). The method used for model fitting can be the maximum likelihood method or Markov Chain Monte Carlo (MCMC). Firstly, we chose the maximum likelihood method for the examination of predictor variables and then, based on that information, MCMC using the Bayesian modeling frame was applied to estimate the uncertainty about the relationships represented by α and β (equation 15) and cross-scale predictions. Models were fitted using the R software (https://www.r-project.org). Predictor variables were entered into the initial models in a stepwise manner for the identification of the variables incorporated in the model, and then entered jointly into the model. This approach was repeated until all remaining variables in the final model were significant at α=0.05. An MCMC sampler in JAGS [66] was used for the Bayesian model fitting. 3 MCMC chains with 50,000 iterations as burn-in and 30,000 iterations thinned every 30 were stored for parameter estimates. Convergence of the model was examined by Gelman-Rubin diagnostics [67] and visual assessment of trace plots of chains.

4.4 Results

Of the 329,830 cases reported in 2011–2013, 124,888 cases in 18 operational health districts in western-Cambodian 10 provinces were included in the analysis. SMR in each health operational district were smoothed using an empirical Bayesian method. In contrast to the decreasing tendency of the API in each district, estimated EBSMR suggested remaining or even the increasing tendencies in endemic areas (Figure 26-28).

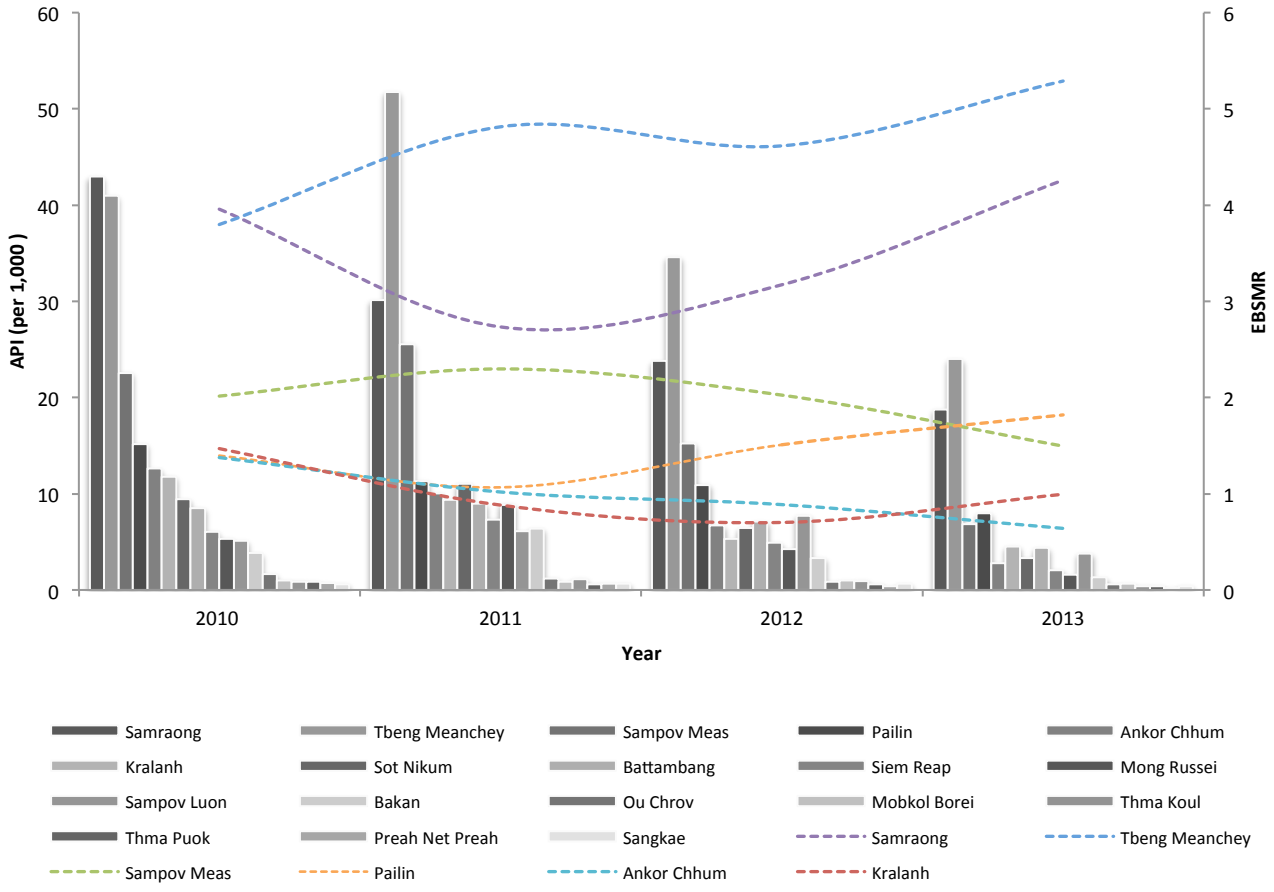


Figure 26 Yearly change of annual parasite incidence (API) for western-Cambodian health district and empirical Bayese estimated standardized morbidity ration (EBSMR) for six operational health districts at high EBSMR [a].

Bar graph represents API in each health operational district
and dotted line represents EBSMR of five provinces at high EBSMR.
[a] District at higher EBSMR than 1.0

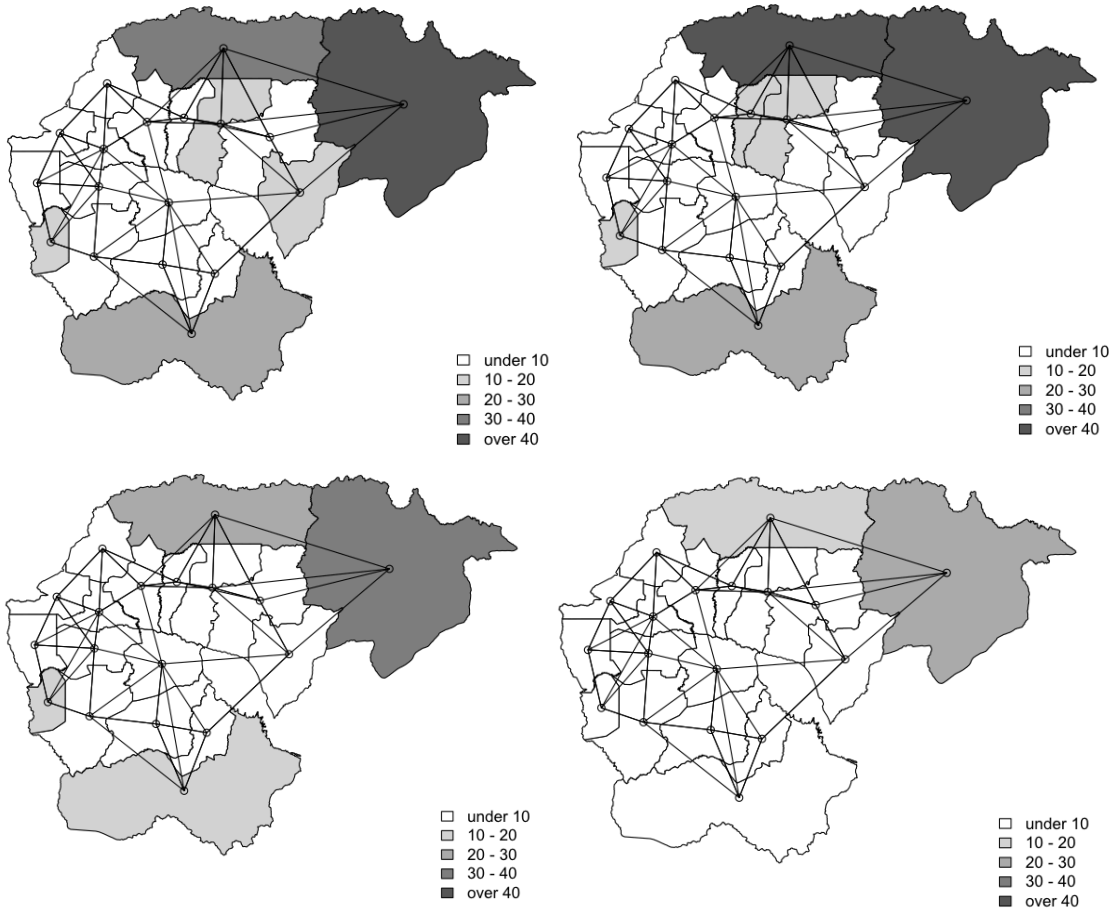


Figure 27 Maps of API for western-Cambodian health districts from 2010 to 2013

API was smoothed by empirical Bayese method with *k* near length method (nearest 3 districts were considered). The map on the top left corresponds to the map of 2010, top right for 2011, down left for 2012 and remaining down right for 2013.

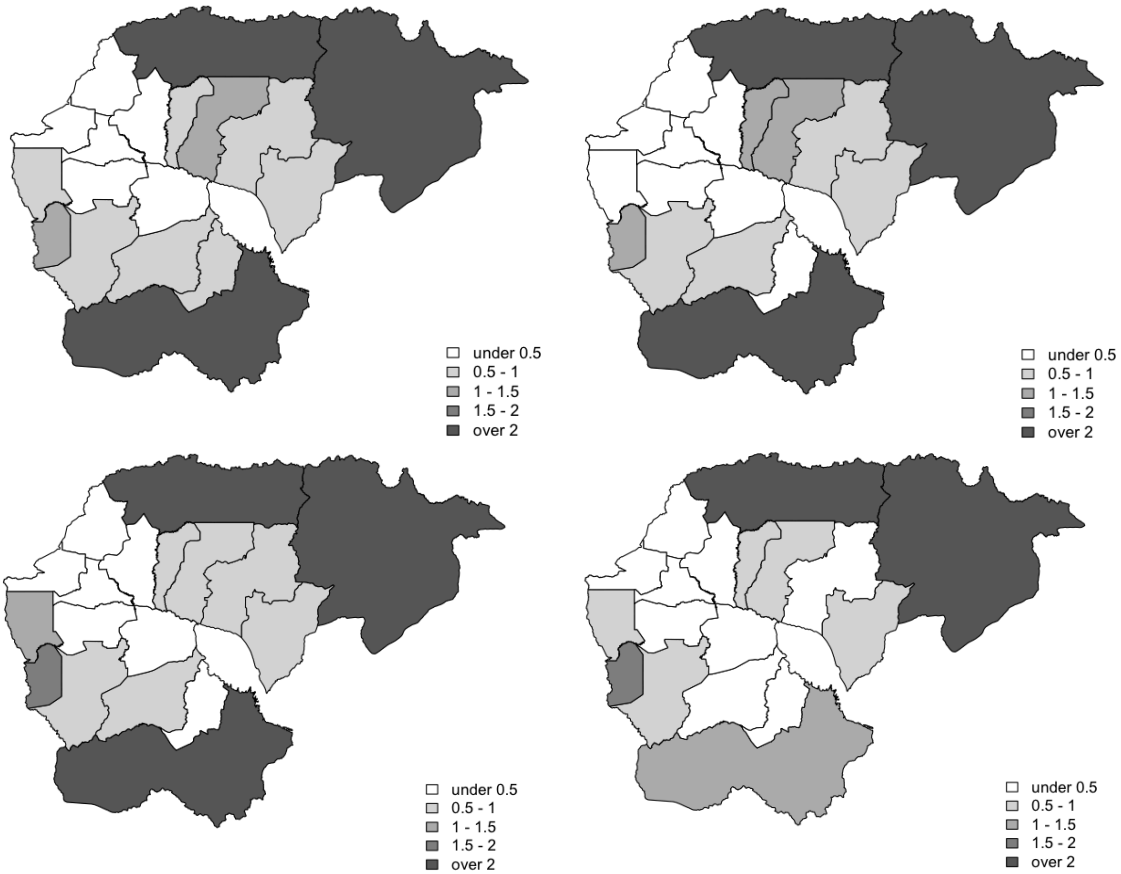


Figure 28 Maps of EBSMR for western-Cambodian health districts from 2010 to 2013

The map on the top left corresponds to the map of 2010, top right for 2011, down left for 2012 and remaining down right for 2013.

Within 5km distance from villages, absolute correlation values between environmental variables (NDVI, LSWI and TWI) extracted from surrounding circular buffers (from 1 to 5km) and EBSMR were highest at 5 km, whereas at 1 km for NDWI (Figure 29). Correspondingly, the correlation efficient of the model differs at each distance.

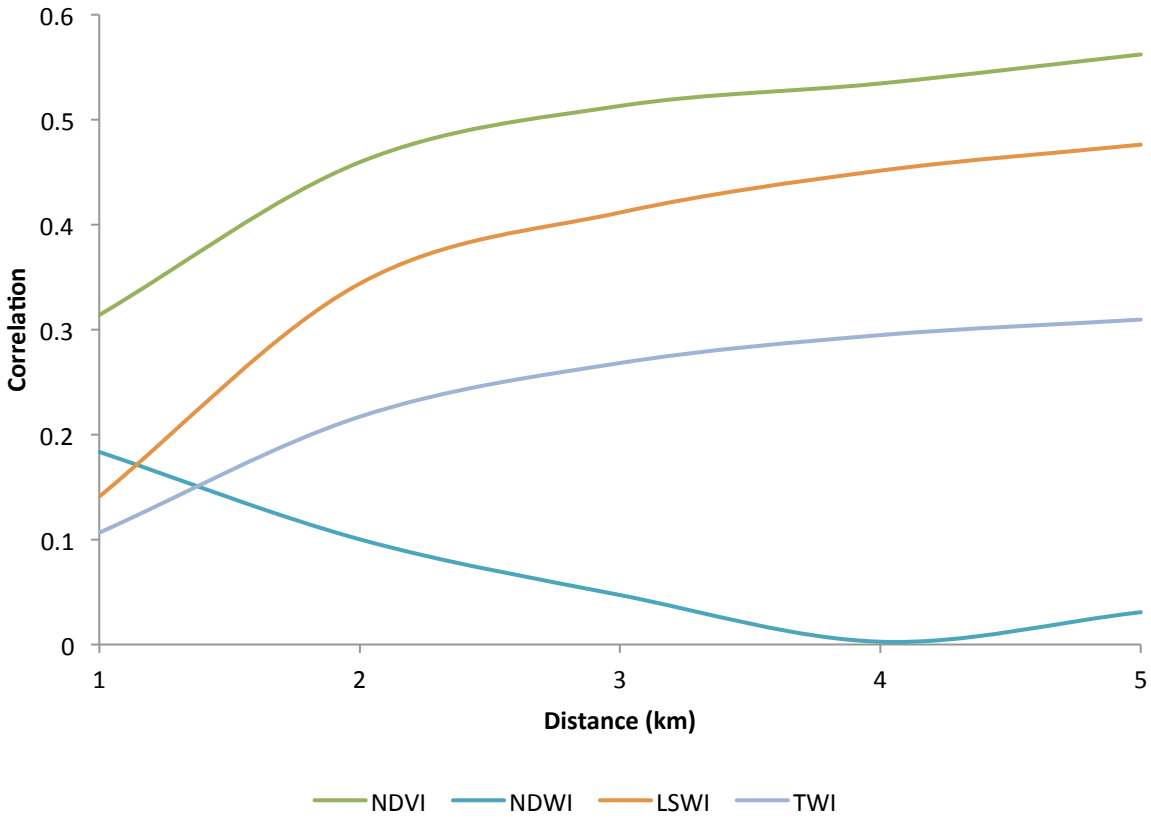


Figure 29 Absolute value between environmental variables (NDVI, NDWI, LSWI and TWI) extracted from surrounding circular buffer from populated villages (from 1 to 5km) and EBSMR

Values were extracted from each 1 km distance circular buffer (1, 2, 3, 4, 5 km) from populated villages and then averaged to mean values.
NDVI, Normalized difference vegetation Index; NDWI, Normalized difference water index; LSWI, Land surface difference index; TWI, Topographical wetness index

Thus, the data collection range chosen for the model was 5 km for NDVI, LSWI, TWI and 1 km for NDWI, respectively. After the variable selection, the final model well estimated the SMR of each area (adjusted $R^2 = 0.774$, AIC = 149.423). This model included NDVI, NDWI, Topographical wetness index, *P. falciparum* temperature suitability index, $LLIN_{suf}$ and TF_{rate}. Parameter estimates for each variable are shown in table 4.

Table 4 Parameter estimates selected for the final generalized linear regression model

Category	Variable	Parameter estimate	Standard error	P-value
Vegetation	NDVI (5 km)	7.446	1.947	<0.001
Water	NDWI (1 km)	-24.330	5.009	<0.001
Geography	TWI (5km)	-1.707	0.6346	0.009
Temperature	*P. falciparum* Temperature suitability index (*Pf*TSI)	0.0002681	0.0000403	<0.001
Vector control	Sufficient ownership of LLIN [a]	-0.06387	0.007157	<0.001
Treatment	Treatment failure rate by artemisinin combination Therapy [b]	0.03611	0.008309	<0.001

[a] Proportion of household in which distributed mosquito net covers 2 persons or less per net.

[b] Test positive for *P. falciparum* on day 28 or day 42

NDWI, Normalized difference water index; NDVI, Normalized difference vegetation index; TWI, Topographical wetness index; LLIN, Long lasting insecticide-treated net

The calibration plot of final model represented the good fitting of the predicted and actual values (Figure 30A). Mean absolute error (MAE) of this final model was 0.499. Figure 30B shows the proportions of predicted values within the range of absolute error from 0.1 to 2. 55.56% of predicted values were covered within the range of ± 0.2, 75% were in ± 0.5 and 87.5% were in ± 1, respectively.

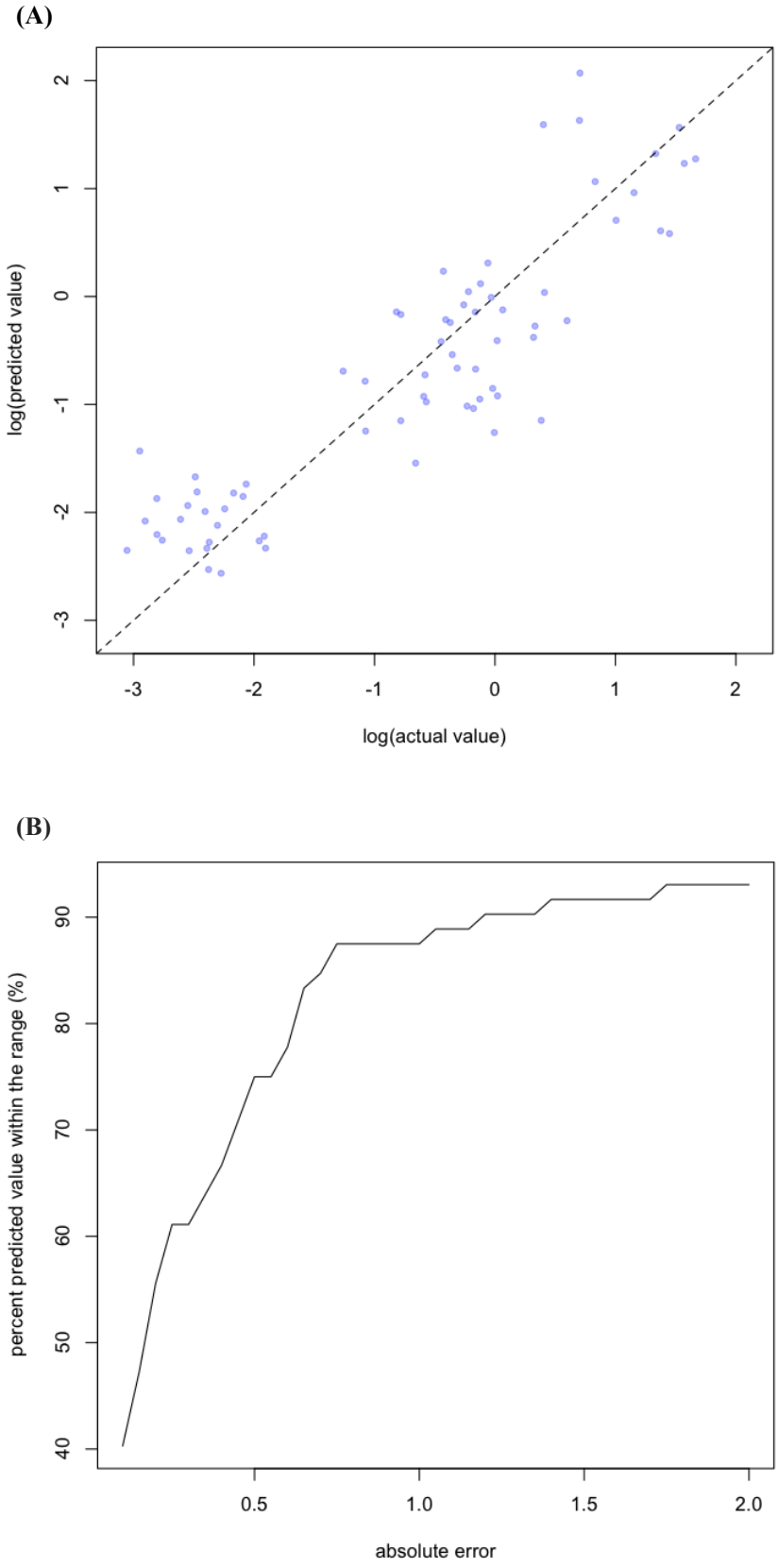
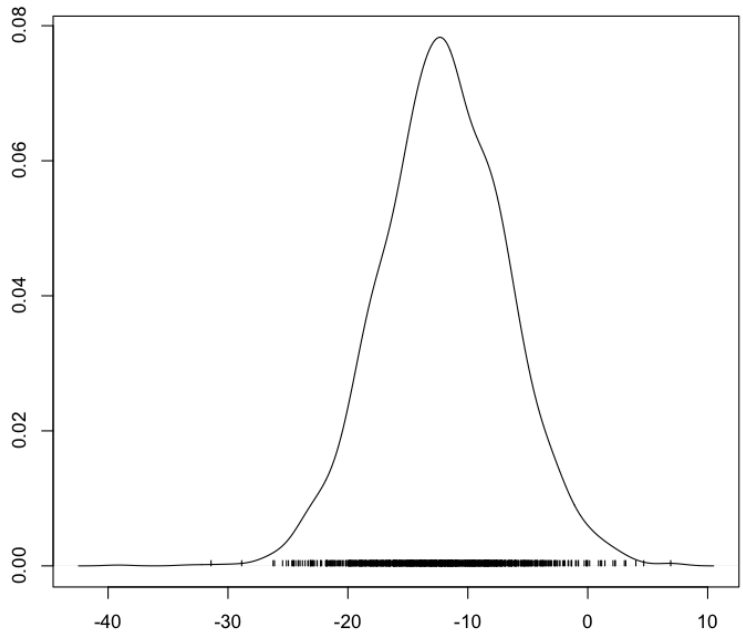


Figure 30 The calibration plot (A) and the proportion of predicted values within the range of absolute error (B) of the final model

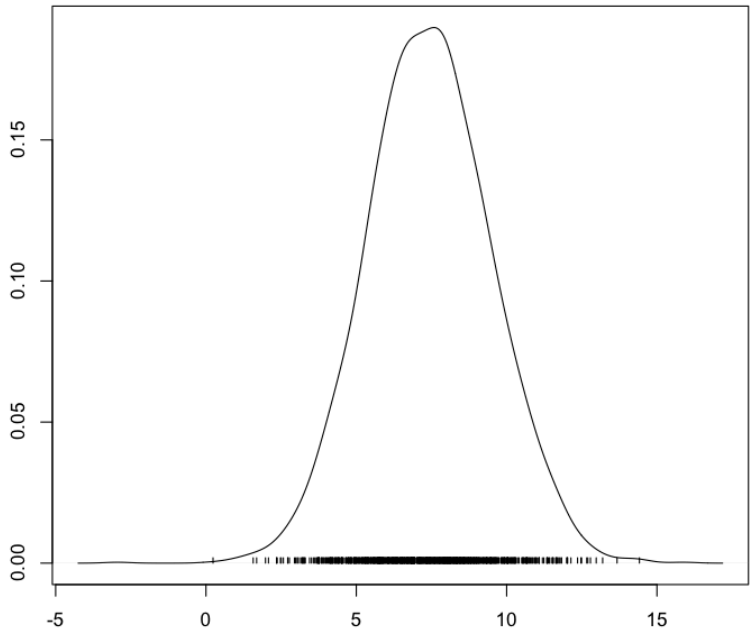
The dashed line in figure (A) represents 1:1 relationship of actual and predicted value.

Based on the information from generalized linear regression modeling, the Bayesian modeling frame was applied to estimate the uncertainty about the relationships represented by α and β (equation 15). The trace plots of the Bayesian modeling frame were monitored to examine the convergence of cross scale prediction (Appendix). The model was settled with given condition for MCMC and provided the range of posterior distribution of parameters (Figure 31).

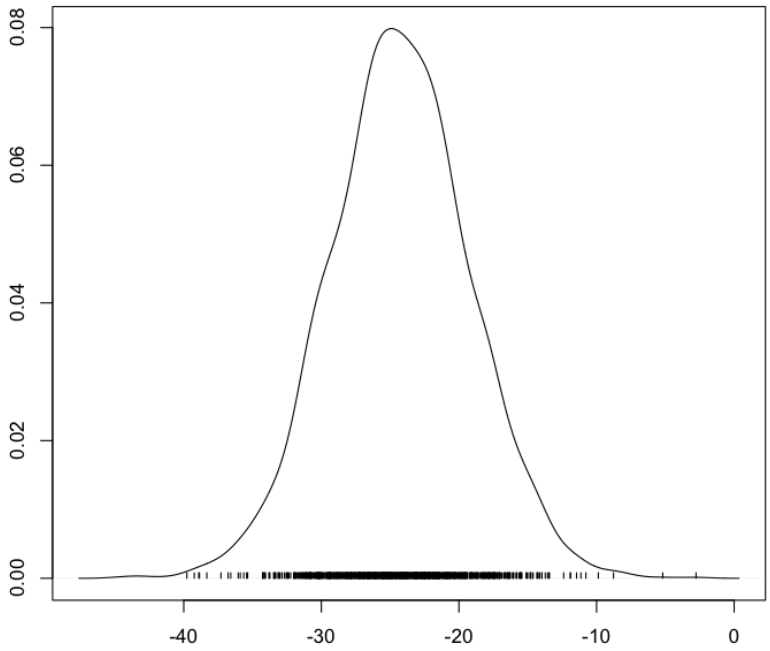
(A)



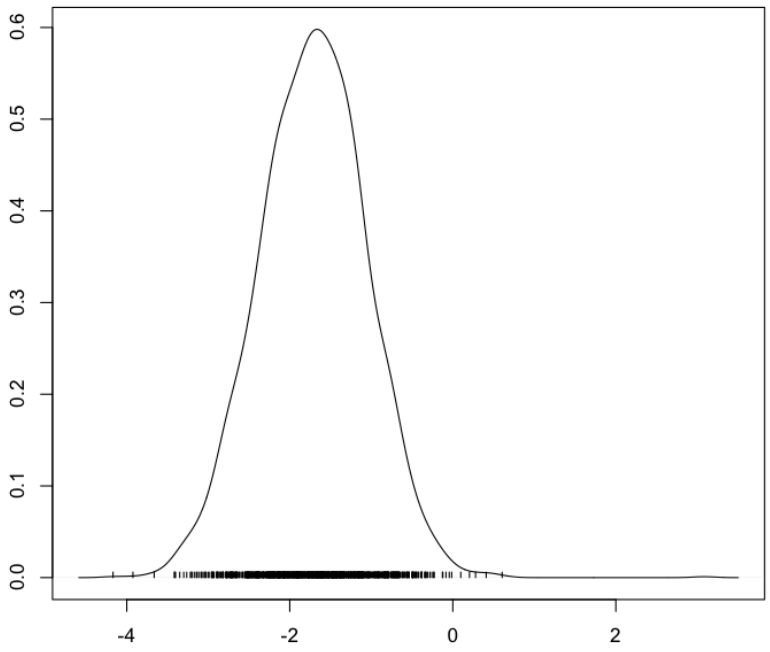
(B)



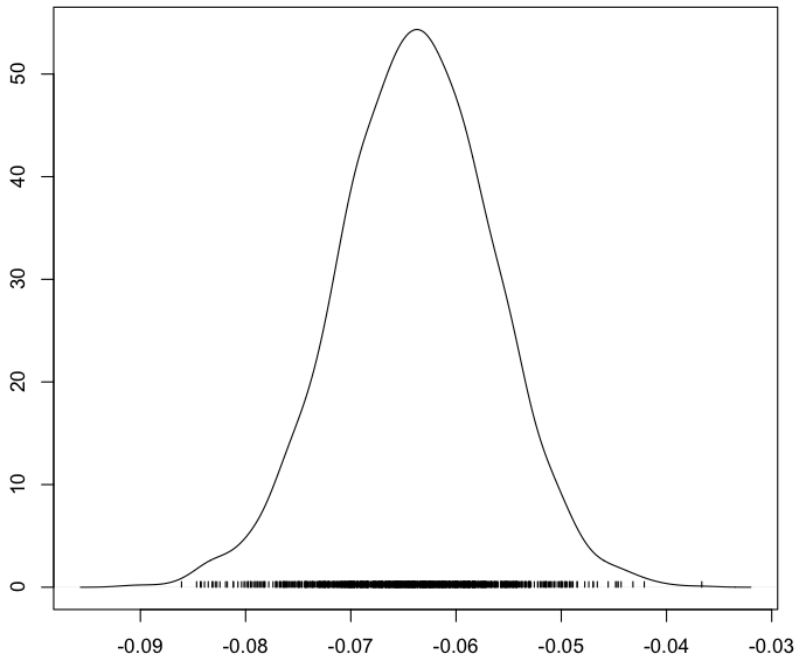
(C)



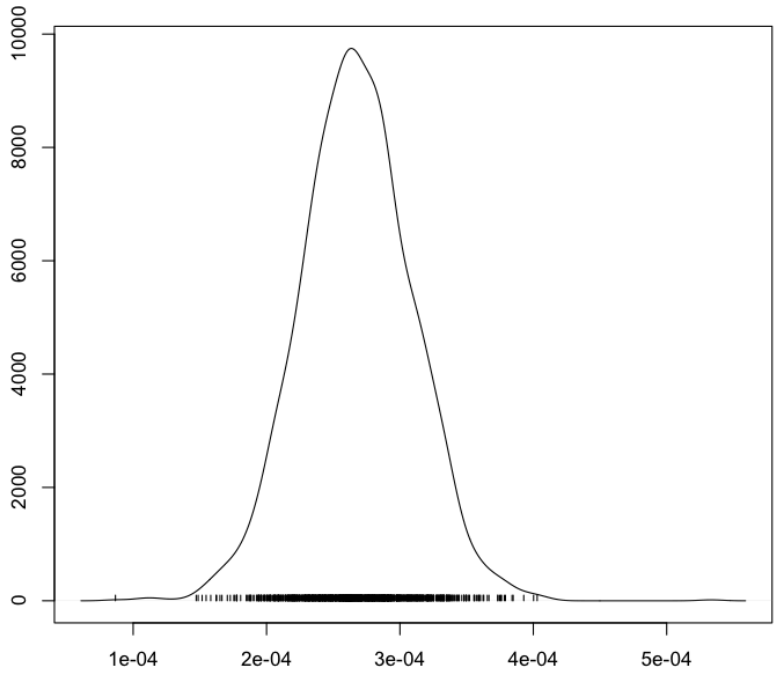
(D)



(E)


0
10
20
30
40
50
-0.09
-0.08
-0.07
-0.06
-0.05
-0.04
-0.03

(F)


0
2000
4000
6000
8000
10000
1e-04
2e-04
3e-04
4e-04
5e-04

(G)

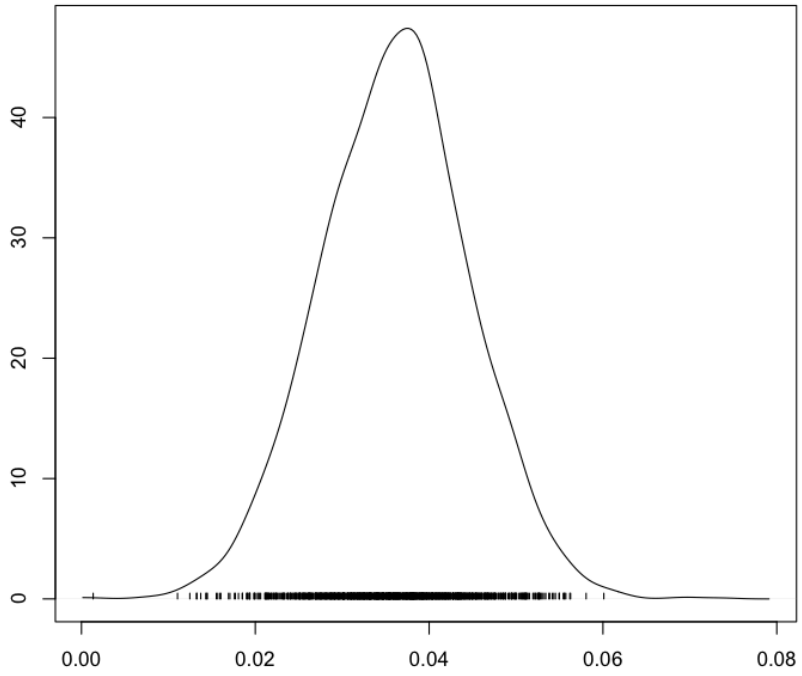


Figure 31 Density plot of posterior distributions of each parameter

Parameter distributions for (A): Intercept, (B): NDVI, (C): NDWI,
(D): TWI, (E): LLIN, (F): Temperature and (G): TF
Horizontal axis of each graph indicates estimated kernel density of each parameter
NDWI, Normalized difference water index; NDVI, Normalized difference vegetation index;
TWI, Topographical wetness index; LLIN, Long lasting insecticide-treated net

5. FINE-SCALE RISK MAPPING USING SPATIAL MALARIA RISK MODEL

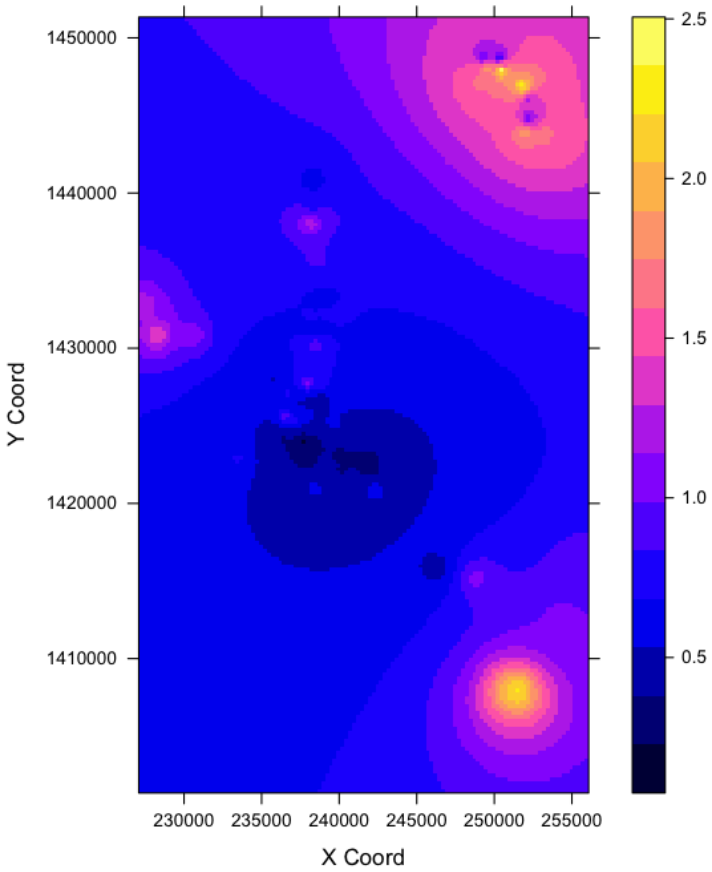
5.1 Fine-scale mapping and its evaluation

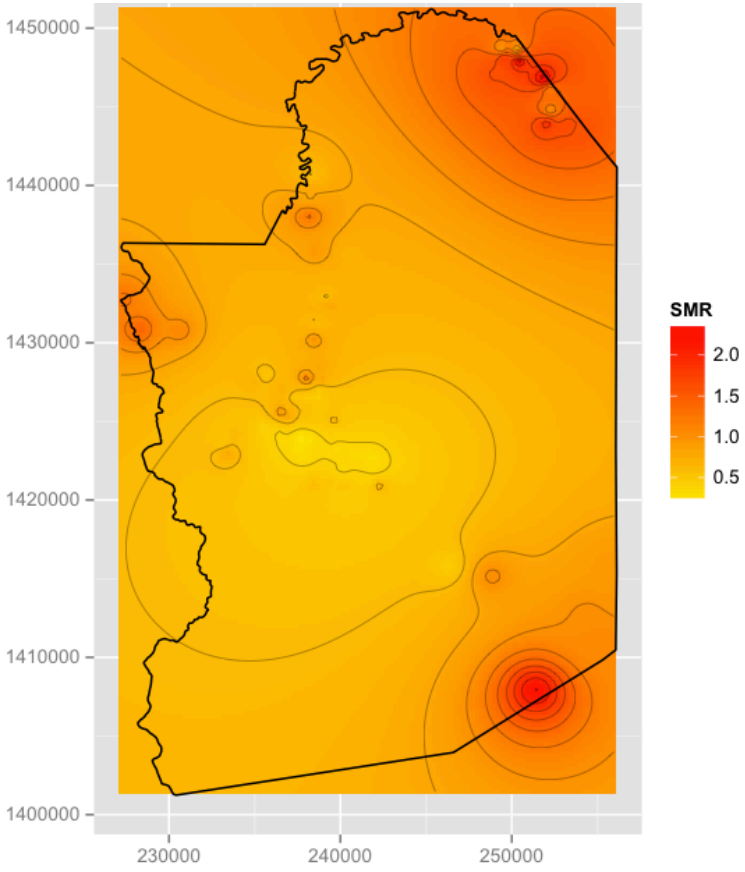
Using Bayesian modeling frame, estimated SMR for each village was calculated. The fitted model was then applied in conjunction with spatial covariates aggregated from the location of each village to estimate the village level SMR. We created maps visualizing the risks of two provinces in western Cambodia, Pailin and Preah Vihear, by the inverse distance weighed method (IDW) and ordinal kriging interpolation of estimated SMR at each village. For accuracy evaluations of cross-scale prediction from the model, predicted SMR was compared with geocoded case data in Pailin [68] and Preah Vihear [69] extracted from Malaria Atlas Project database [62] using Spearman's rank correlation [70] and Welch's t-tests for unequal variances [71]. As the source data of our map were mostly depending on the report from VMW and HIS based on the rapid diagnostic kit (RDT) and Microscopy detection, these data were chosen for the reason of the detection method (RDT / Microscopy) and closer report period among the available data. To exclude the incidental nature for spearman's correlation with this sample data, we conducted 2,000 times of resampling of this dataset with replacement to create confidence interval by non-parametric bias corrected and accelerated percentile method [72] and assess the distribution of correlation values. Visual representations of risk distributions in the maps were also validated through interviews with healthcare providers in regional health center and professionals of GIS.

5.2 Results

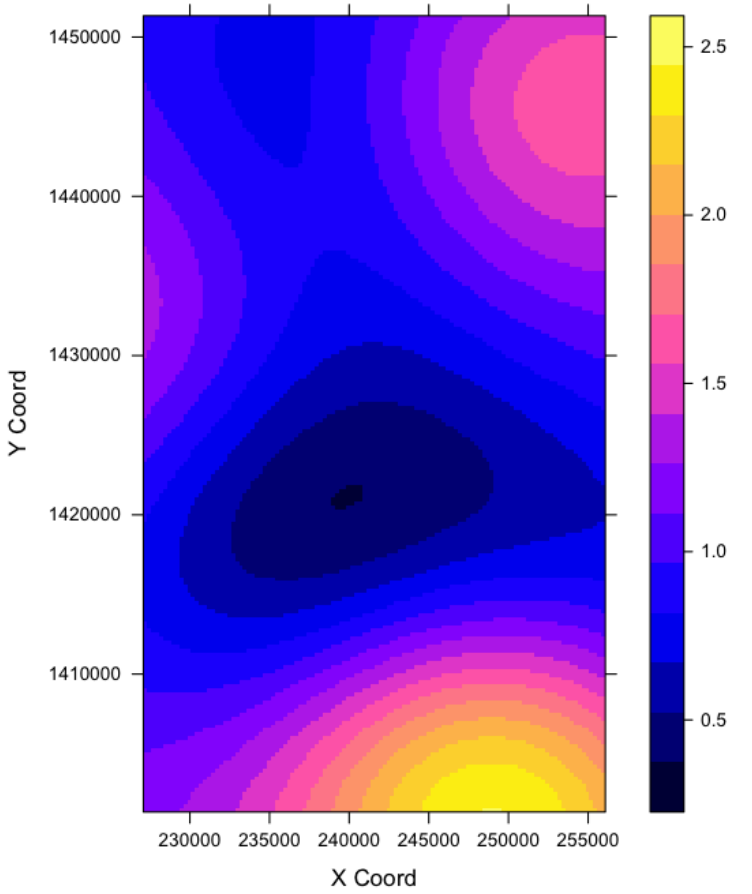
Fine-scale maps were created by the inverse distance weighed method (IDW) and ordinal kriging interpolation. Figure 32 shows the maps created from the predictive models for Pailin and Preah Vihear province. Each map represents different risk representation pattern in accordance with respective interpolation method. The map interpolated by IDW showed more spotted risk, which helps identification of localized risky hotspots, whereas the map interpolated by ordinal kriging showed broader patterns providing bigger perspective for optimizing healthcare resource distributions.

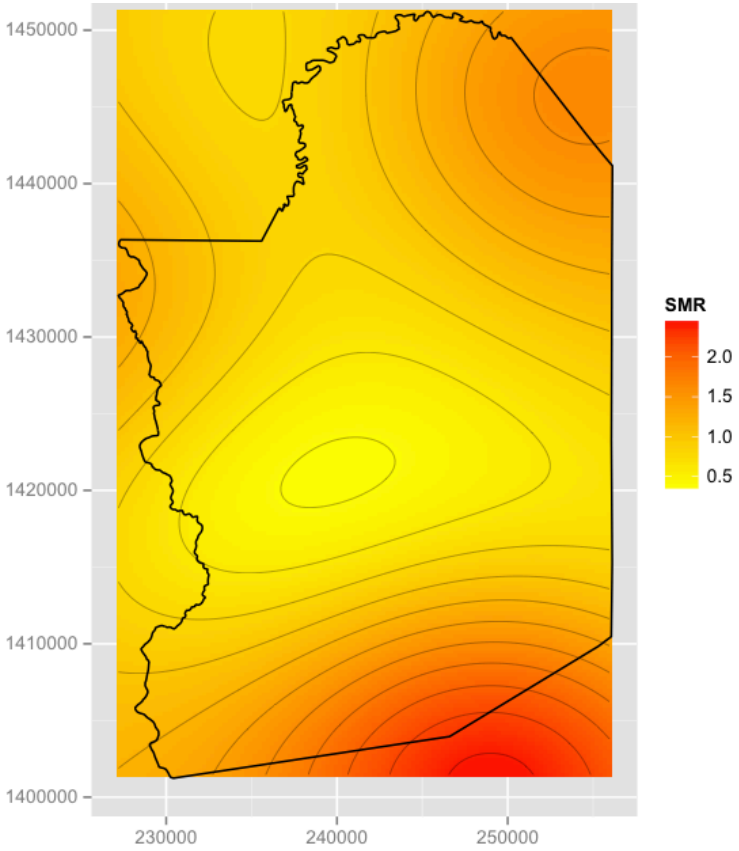
(A) Pailin, 2010 (IDW)



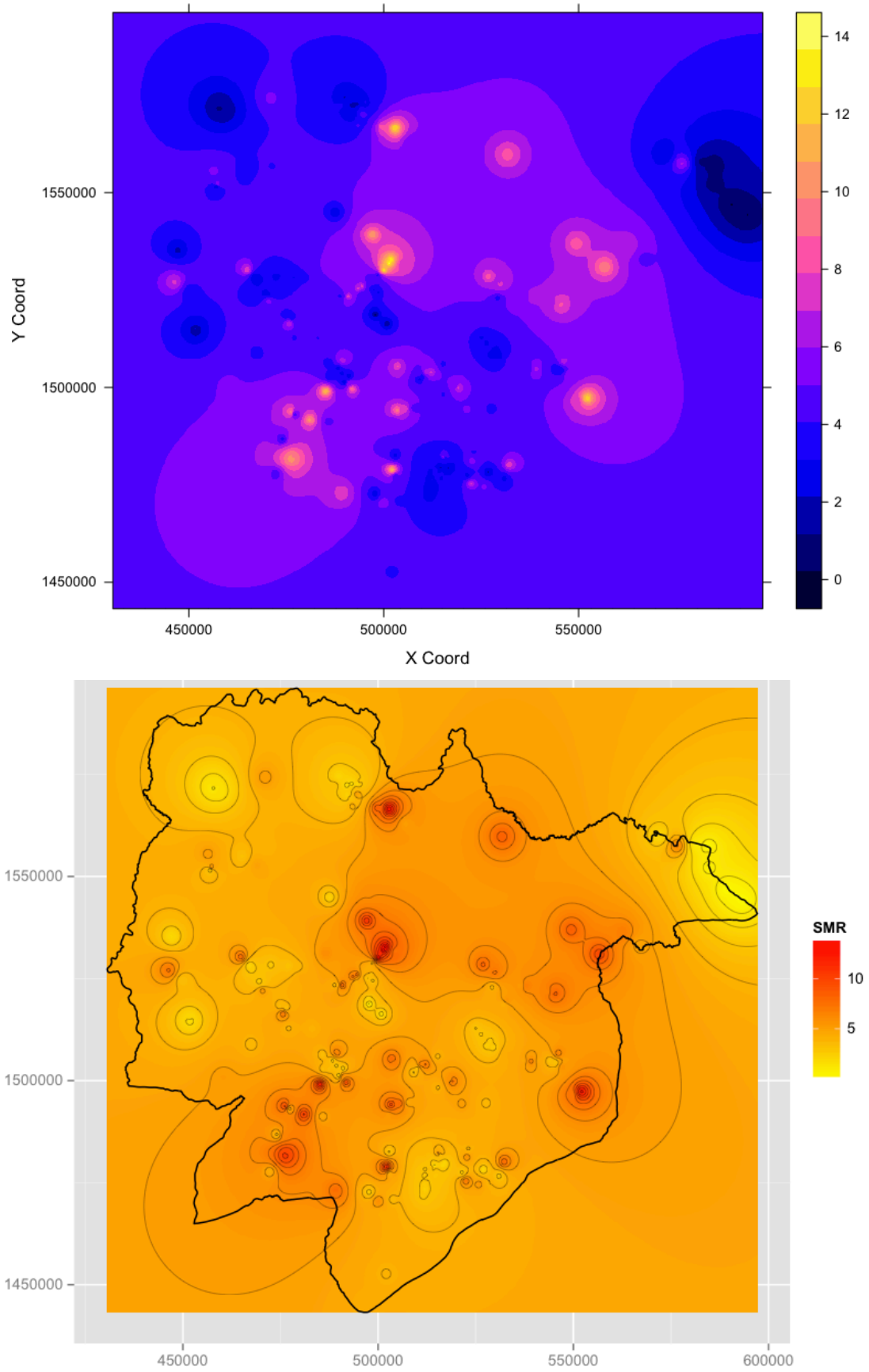


(B) Pailin, 2010 (Ordinal kriging)

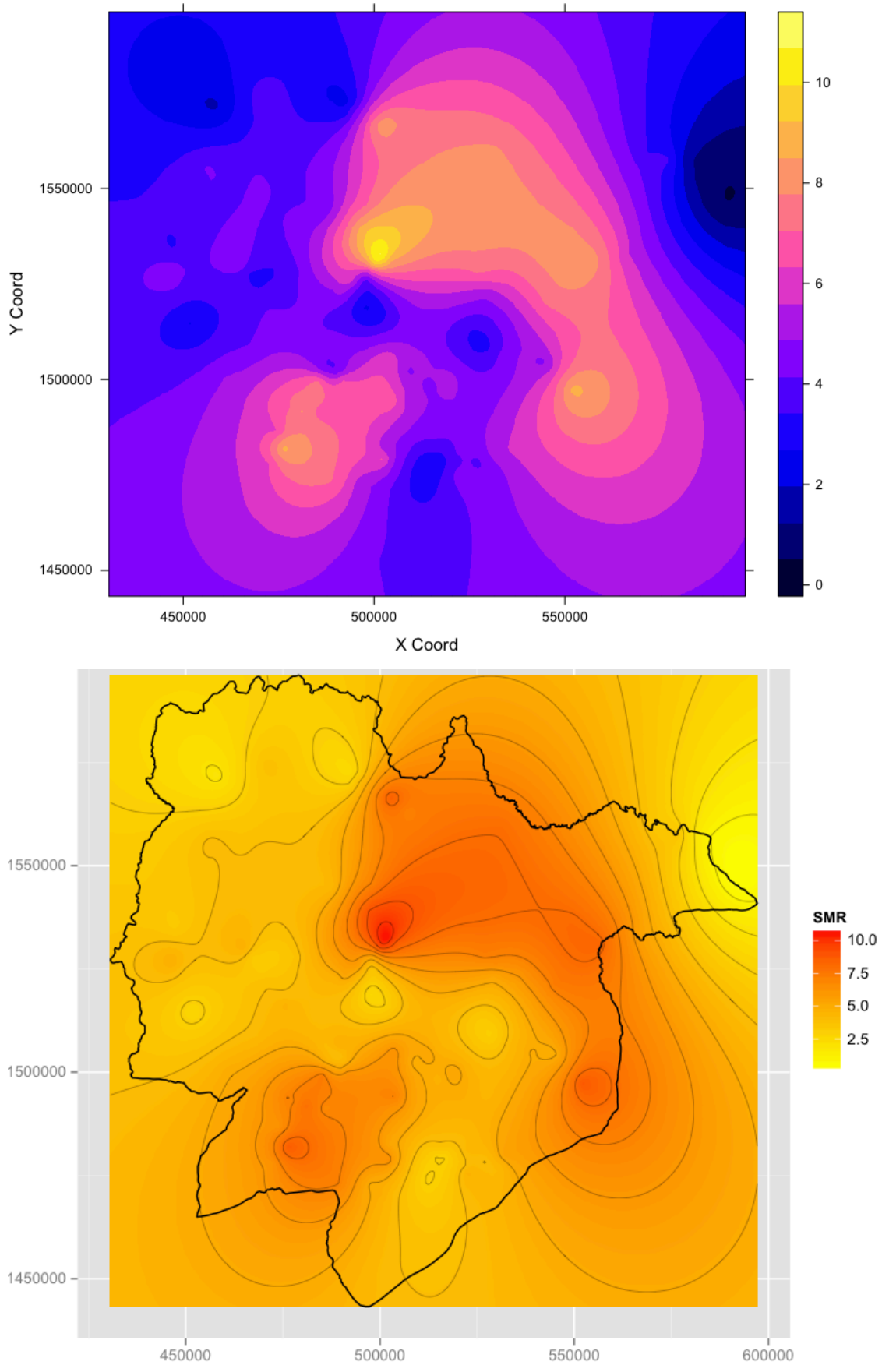




(C) Preah Vihear, 2010 (IDW)

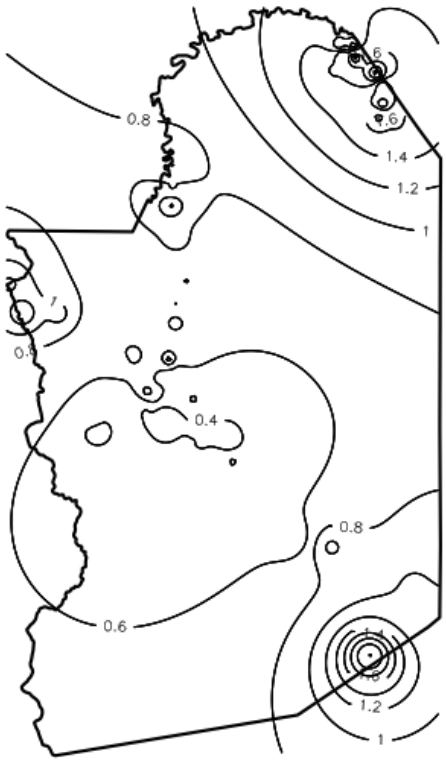


(D) Preah Vihear, 2010 (Ordinal kriging)

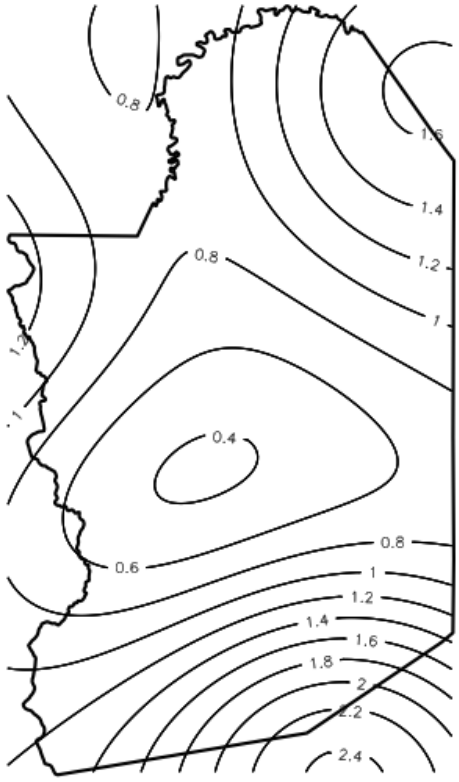


Risk contour

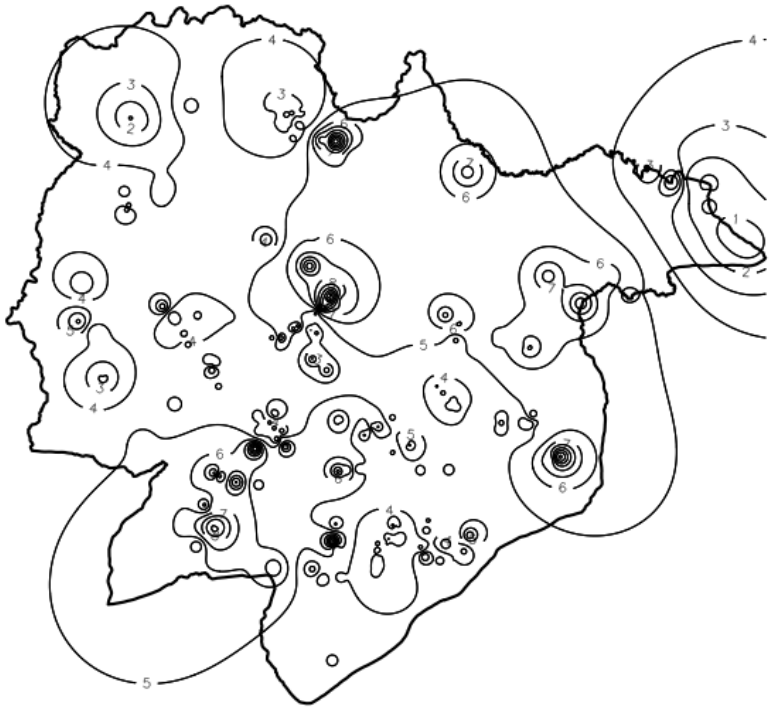
A



B



C



D

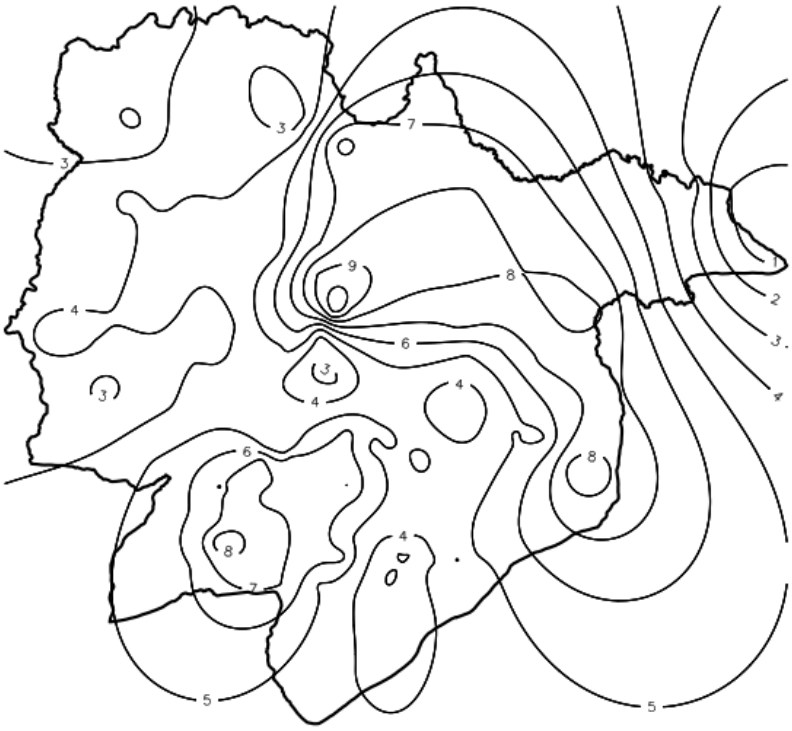
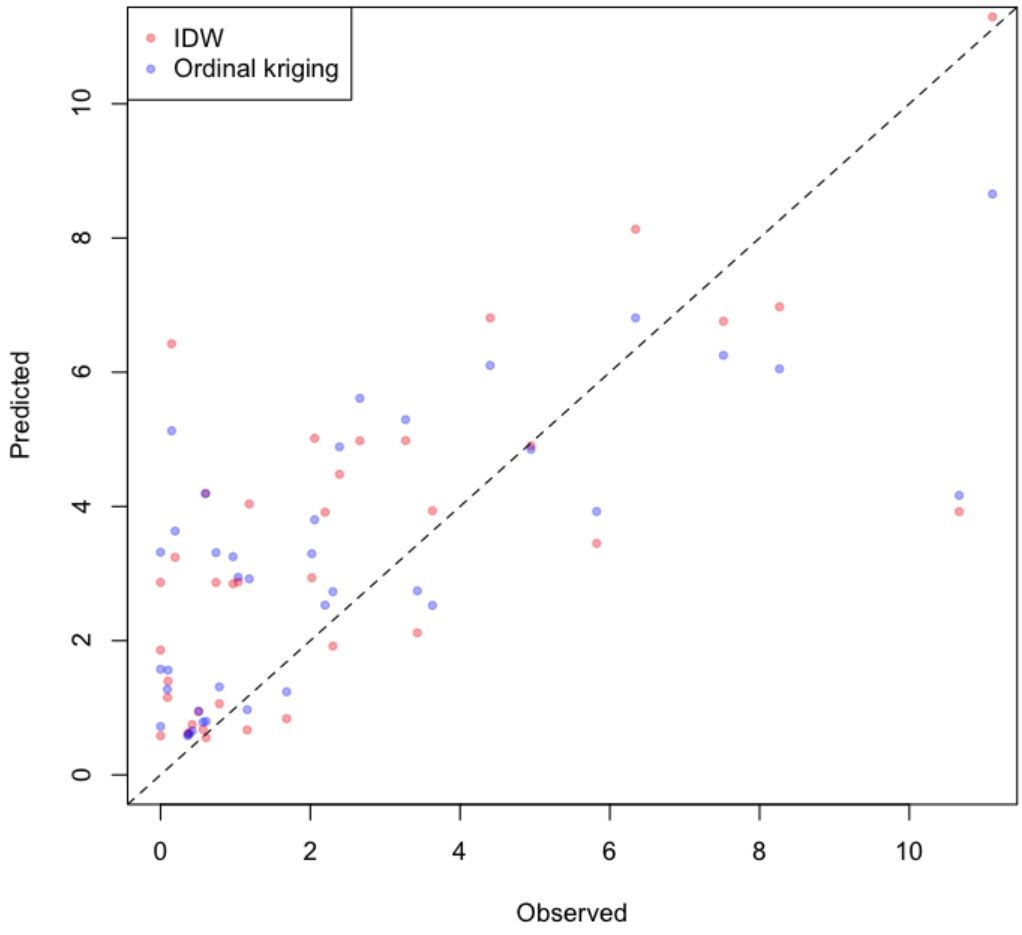


Figure 32 Representative maps created from the predictive models for Pailin (A, B) and Preah Vihear (C, D) province in 2010

Map (A, B) are the representative maps of Pailin province and (C, D) for Preah Vihear province in 2010. Map (A) and (C) were the risk maps created by the inverse distance weighed interpolation method (IDW) and map (B) (C) correspond to the maps created by the ordinary kriging.

In comparison with geocoded case data, corresponding predicted values in this map showed conformity (Spearman's rank correlation; r = 0.662 in IDW and 0.645 in ordinal kriging, Welch's t-test; N.S.) showing that the cross-scale predictions were corresponding to the actual case reports (Figure 33A). 95% confidential intervals for both IDW and ordinal kriging were 0.414 – 0.827 and 0.368 – 0.813, respectively, showing steep peak in the kernel density plot at around 0.65 – 0.7 (Figure 33B). The visual representations of hotspot shown in this map were confirmed that they were aligned with the actual areas at high risk, which could be identified by other sources [48, 62, 69], through the visual assessment by a number of healthcare providers and experts of GIS.

A



B

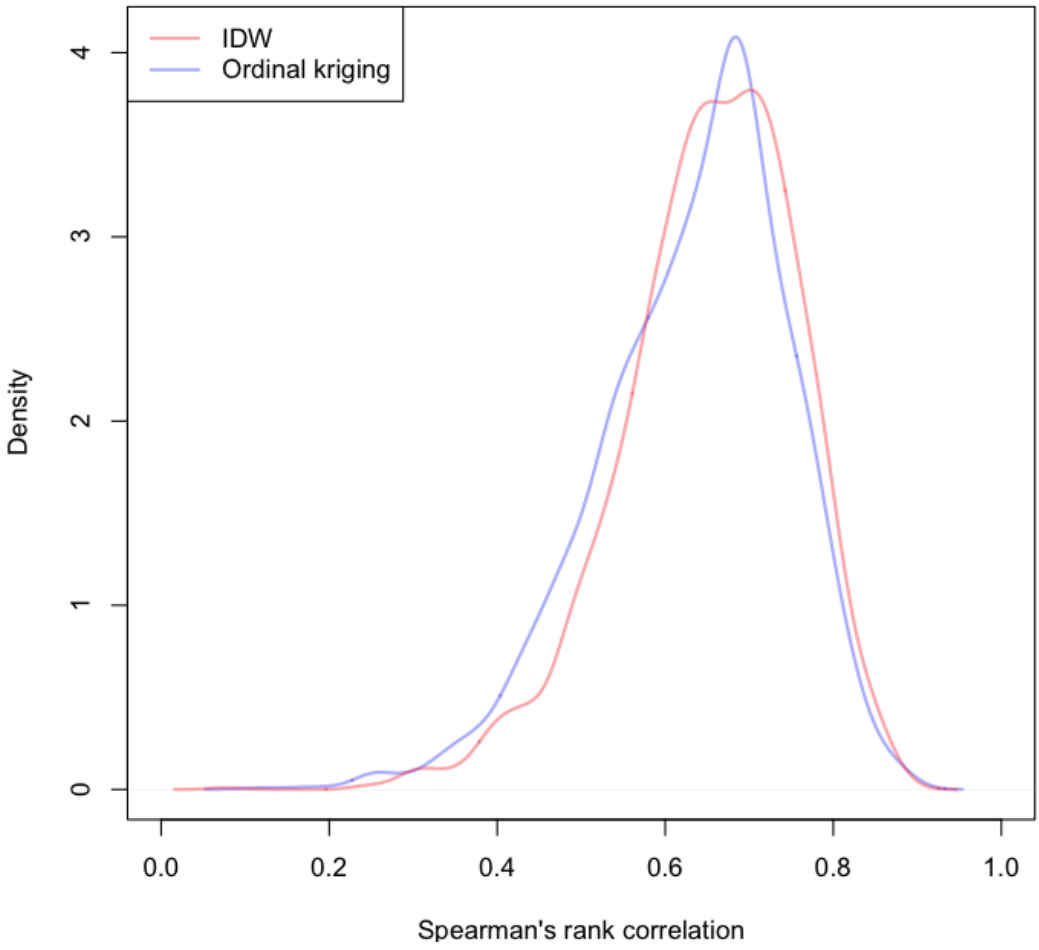


Figure 33 Comparison of the standardized morbidity ratio calculated from geocoded case data with corresponding predicted values (A) and the kernel density plot of the resampled spearman's rank correlation (B) in the risk map created by the risk-modeling frame work.

The dashed line represents 1:1 relationship of observed and predicted values
IDW, Inverse distance weighed method

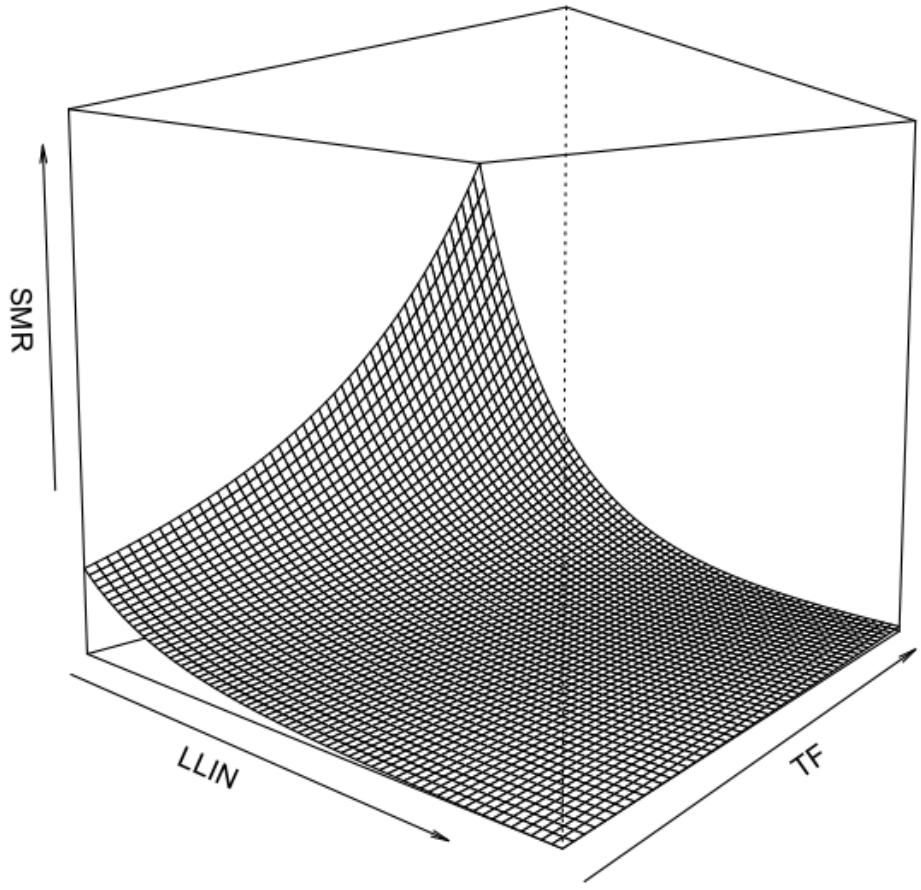
6. COMPUTATIONAL SIMULATIONS UNDER GIVEN CONTAINMENT STATUS INDICATORS

We conducted computational simulations to demonstrate the relationship of expected outcomes and containment status indicators using the model developed here. The followings are the results and implications.

6.1 Simulation results of expected outcomes under given containment status indicators

Using the model, computational simulations of expected outcomes under given conditions of $LLIN_{suf}$ and TF_{rate} were conducted. The visual representations of simulation results demonstrated not only multiplicative contribution of these indicators but also the different patterns of expected outcomes from the combination of these two containment status indicators in respective areas (Figure 34).

A



B

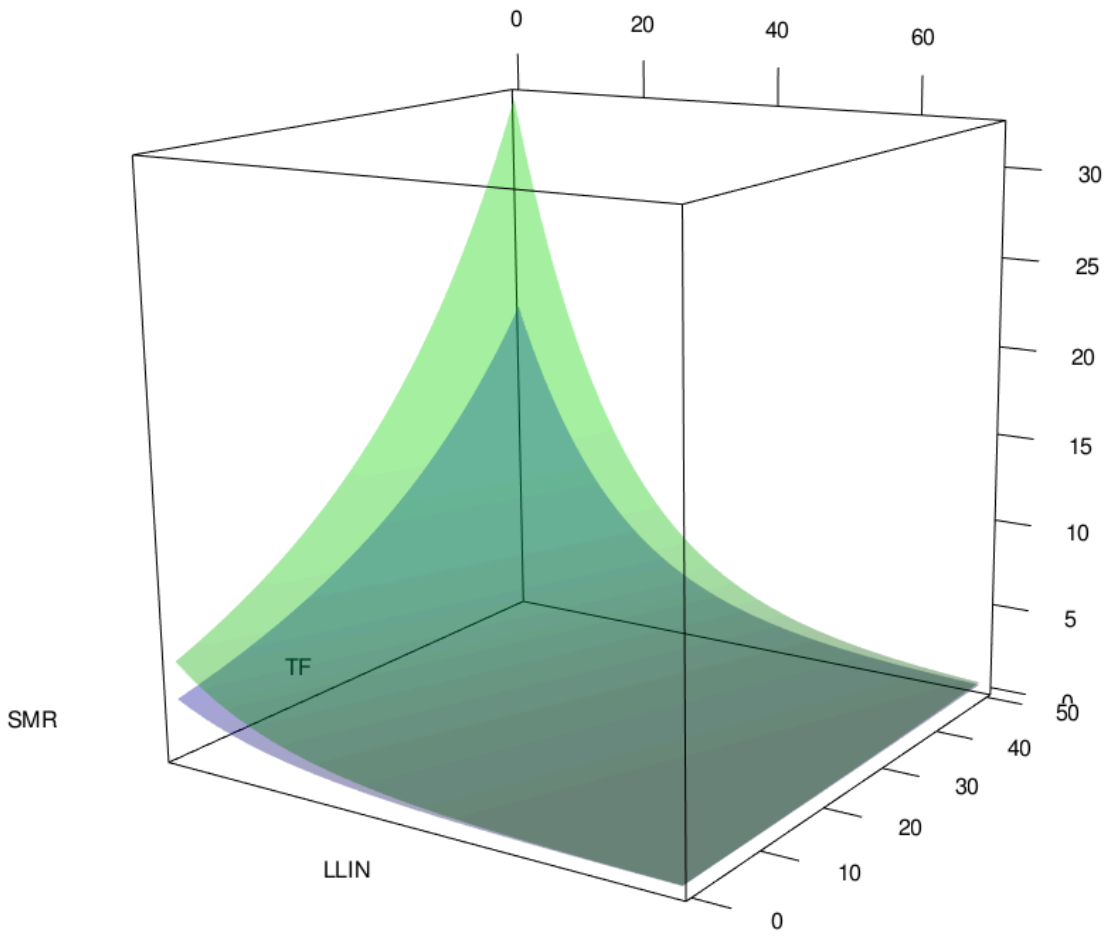


Figure 34 Computational simulations of expected standardized morbidity ratio (SMR) under various conditions of LLIN coverage and Treatment failure rate of artemisinin

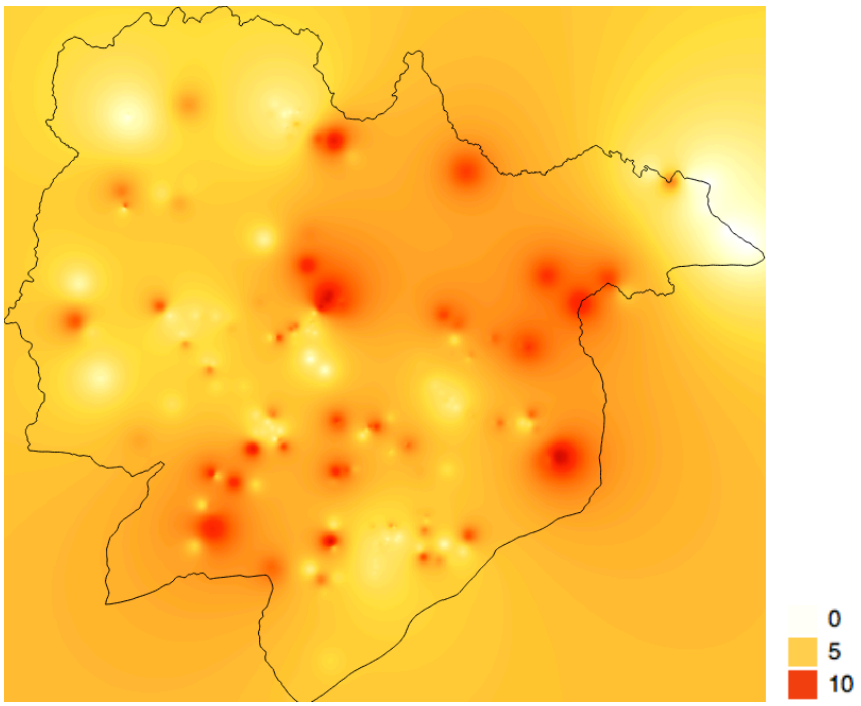
Figure (A) represents the relationship of two containment status indicators with expected SMR in Pailin province. Figure (B) represents the different pattern of expected outcomes from the combination of these two containment status indicators in respective areas. The green surface corresponds to that in Pailiin, and the blue surface corresponds to Preah Vihear province. LLIN, Long lasting insecticide-treated net; TF, Treatment failure rate of artemisinin defined as test positive percentage on day 28 or day 42.

6.2 Geographical analysis for expected outcomes from the targeted containment status

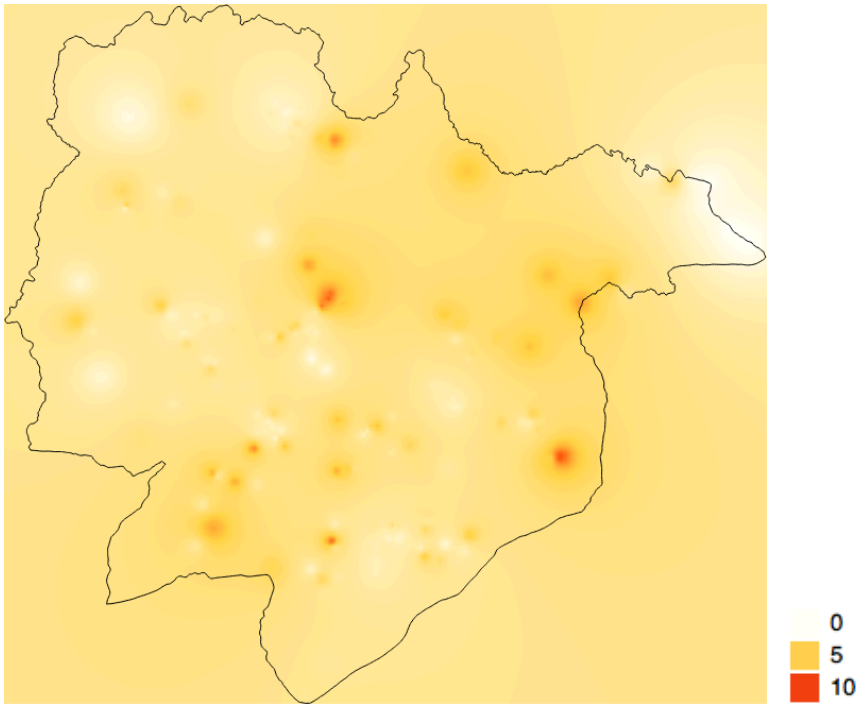
The simulation results could be mapped to examine the geographical effect expected from targeted containment status. Figure 35 shows an example of geographical analysis. The geographic view of the effect from each or combined containment status

could be obtained from this analysis, providing geographical perspective for decision makers.

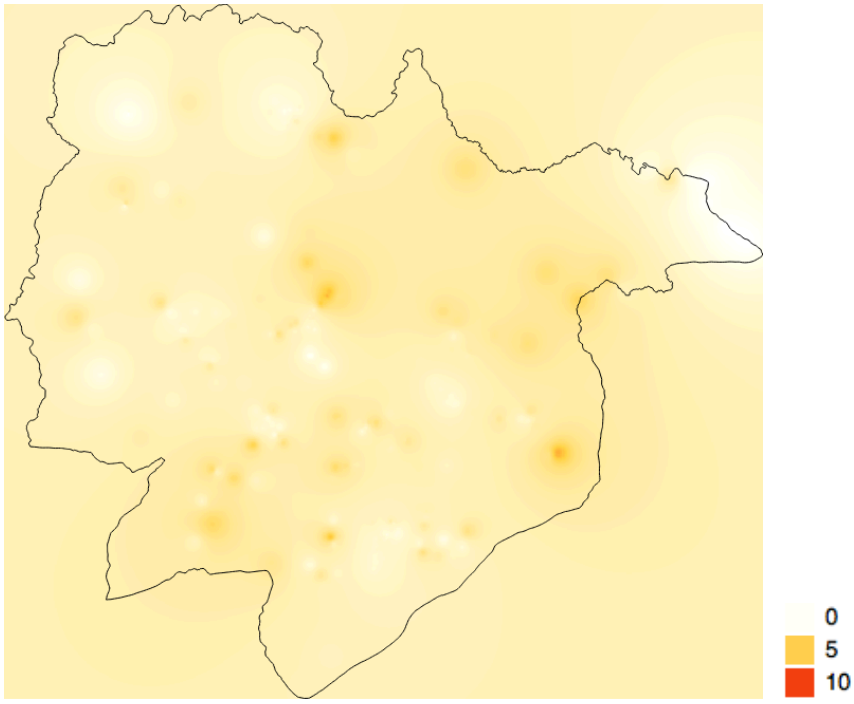
A: Current state



B: TF decreased to 0% (from current 14.7%)



C: LLIN increased to 20% (from current 5.5%)



D

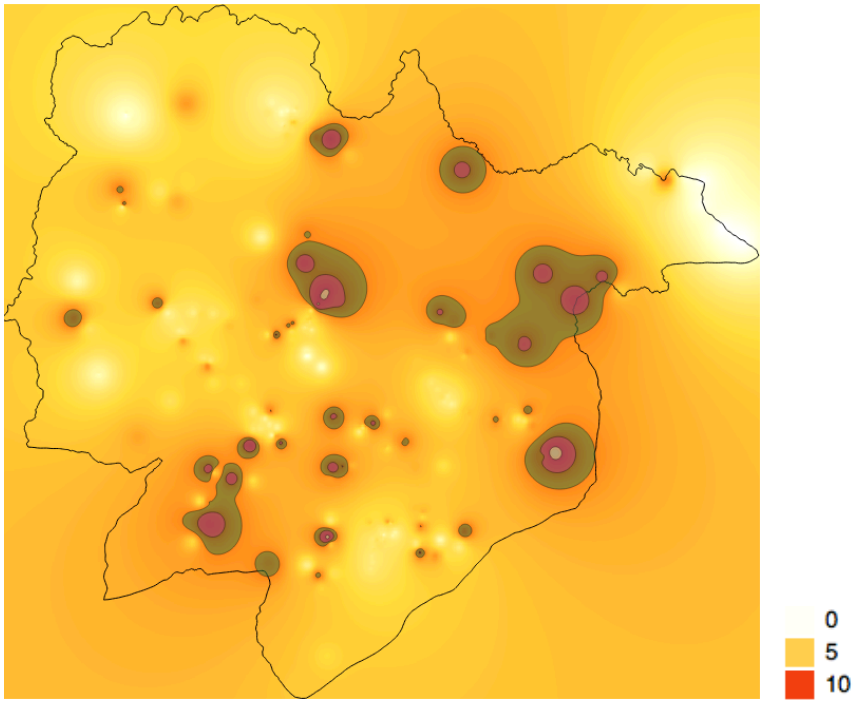


Figure 35 An example of geographical analysis of expected outcomes from targeted containment status in Preah Vihear

Figure (A) represents the current predicted state. Figure (B, C) mapped geographical view of expected outcomes from targeted containment status. Simulation outcomes can be visualized on the same map (D), by which the effect can be examined. The green (for TF decrease), purple (for LLIN coverage increase) and gray (combined) colored area corresponds to areas at more than 5 SMR. LLIN, Long lasting insecticide-treated net; TF, Treatment failure rate of artemisinin.

7. VERIFICATION AND VALIDATION OF THE SYSTEM

7.1 The Requirements Verification Traceability Matrix of the system

The Requirement Verification Traceability Matrix for the requirements need evaluations is shown in Table 5. Almost all the requirements were verified except the requirement 3.2.1. As available data were not sufficient to conduct spatio-temporal analysis, we did not consider time series transition in the regional malaria risk during the data collection period in the modeling frame developed here on condition that the relative regional risk does not change drastically in a short period. This issue needs to be covered by future work to demonstrate the spatio-temporal transition of malaria risk in fine-scale. This issue is discussed in chapter 8.

Table 5 Requirements Verification Traceability Matrix

ID	Sources	Requirements	Verification method	Results
1	Sequence diagram, FFBD	Estimate the risk of malaria	Section 4.4: **Analysis** Calibration plot and the model examination Section 5.2: **Analysis** Comparison of geocoded data and predicted data on the map	Met
2	Sequence diagram, FFBD	Receive data	Section 4.4: **Test** Variables used to build modeling framework	Met
2.1	Use case, Context diagram	Receive malaria information system data	Section 4.4: **Test** Variables used to build modeling framework	Met
2.2	Use case, Context diagram	Receive map resource data	Section 4.4: **Test** Variables used to build modeling framework	Met
2.3	Use case, Context diagram	Receive containment status data	Section 4.4: **Test** Variables used to build modeling framework	Met
2.4	Use case, Context diagram,	Receive demographic data	Section 4.4: **Test** Variables used to build modeling framework	Met
2.5	Use case, Context diagram	Receive remote sensing data	Section 4.4: **Test** Variables used to build modeling framework	Met

2.6	Use case, Context diagram	Receive the data for simulation modeling	Section 4.4: **Test** Variables used to build modeling framework	Met
2.7	Use case, Context diagram	Receive meteorological data	Section 4.4: **Test** Variables used to build modeling framework	Met
3	Use case, Sequence diagram	Visualize (Provide) information	Section 7.2: **Demonstration/Inspection** Stakeholders interview	Met
3.1	Use case	Display estimated malaria risk	Section 5.2: **Inspection** Inspection of visual representation of the map	Met
3.1.1	Requirement diagram	Display risk with colored gradation	Section 7.2: **Demonstration/Inspection** Stakeholders interview	Met
3.1.1.1	Requirement diagram	Risk interpolation	Section 5.2: **Inspection** Inspection of visual representation of the map	Met
3.2	Use case	Display the simulation results	Section 6.1/6.2: **Demonstration** Simulation and visual inspection of the results	Met
3.2.1	Requirement diagram	Display time series results	N/A	Not met
3.3	Requirement diagram	Display basal map	Section 5.2: **Demonstration** Visual inspection of the map	Met
3.3.1	Requirement diagram	Reflect accurate positioning	Section 5.2: **Analysis/Inspection** Visual inspection and comparison of geocoded data and predicted data on the map	Met
3.4	Sequence diagram	Support decision making	Section 7.2: **Inspection** Stakeholders interview	Met
4	Requirement diagram	Risk modeling	Section 4.4: **Analysis** Calibration plot and the model examination	Met
5	Requirement diagram	Simulation modeling for the variables of interest	Section 4.4: **Analysis** Calibration plot and the model examination	Met
6	Use case, Sequence diagram	Simulate the efficacy of intervention	Section 6.1/6.2: **Demonstration/Inspection** Simulation and visual inspection of the results	Met
6.1	Requirement diagram	Probability sensitivity analysis	Section 5.2: **Analysis** Comparison of geocoded data and predicted data on the map	Met

7.2 Stakeholders interview

7.2.1 Method

To validate the system, we conducted series of interviews with key stakeholders. At first, following 3 key questions were asked to the interviewee while showing the process and products of the fine-scale map created here and the map of Preah Vihear province from Malaria Atlas Project database as a reference comparator. Additional comments from stakeholders made in the interview were also recorded.

Table 6 Key questions for the stakeholders interview

No.	Questions
1	Is this approach useful for supporting the planning of healthcare resource distributions and malaria containment actions?
2	From visual feature perspective, are the hotspots easily identified?
3	What kinds of improvements are needed to make this be more valid? Possibilities of other applications?

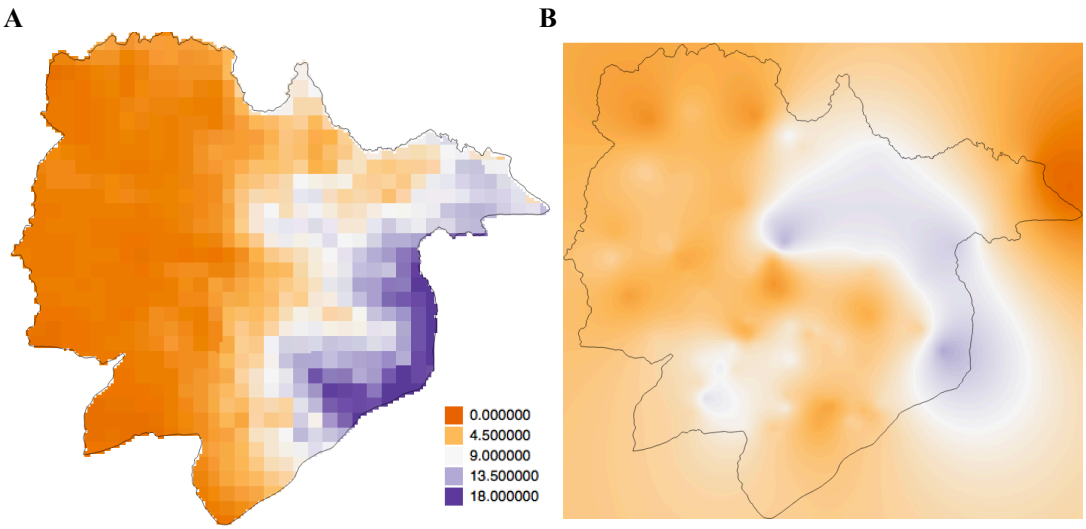


Figure 36 Visual representations of the map from Malaria Atlas Project database [62] and the fine-scale map created by the risk prediction model developed

Figure (A) corresponds to the map from Malaria Atlas Project database. Figure (B) is the fine-scale map created by the method developed in this research.

7.2.2 Interview with NGO staff

Interviewee: Ms. Akemi Takahashi (Chief of Cambodia office, Foundation for International Development / Relief)

The visual representation of the fine-scale map is very similar to that of Malaria Atlas Project. It makes it easier for the initial selection of interventional target area and its control, which facilitates the impact evaluation by visualize the effectiveness of planned interventions. However, this does not reflect the selective effect on the targeted hotspot. Introducing this map to the staff in the health ministry and the health operational district can be recommended. There are too many reporting steps in the current health information system in Cambodia. The educational level is also related and need consideration. In terms of ensuring appropriate drug use, this approach can be applied to the case of tuberculosis.

7.2.3 Interview with GIS professional and engineer

Interviewee: Professor Sophiap Seng (President of National Institute of Posts, Telecommunications and ICT) and Professor Sarann Ly (Professor of rural engineering, Institute of Technology of Cambodia)

This approach is very interesting and the visual representation of the fine-scale map is similar to that of Malaria Atlas Project. Co-kriging can be considered for interpolation method to reflect the regional context appropriately. The most attentive thing is that the map was created mainly from open data sources. However, I think this kind of approach has already done a lot in this research field. Therefore, the uniqueness of this approach needs to be clearly conveyed. The output needs to be simpler, especially for the part of practical application of risk model. Sequence diagram is one option. For country level, this map could give good view. There are many stakeholder acts locally but their action is too detailed. Good customer is the staff of the health ministry. Bill and Melinda Gates foundation could be interested in this product. As for the color presentations, hotspot should be presented as red like colors reminding risk.

7.2.4 Interview with field healthcare provider

Interviewee: Mr. Phin Canmonin (Director of provincial health center in Kampon Chhnang province)

He thought that the map was useful to see the hotspot and conduct intervention. By

2025, Cambodian government is aiming for eliminating malaria. He thought that this map could support these activities for achieving this target. Most patients come to health center (HC) in January. In August, in the middle of rainy season, people get infected especially in mountain and forest. They regularly go to the mountainside for their work. 24 cases came to the hospital in August. Once patients get treated at HC, they never come back to the HC, unless they get infected again. There are 4 species of malaria, 2 common species are *falciparum* and *vivax*. *P. falciparum* is the most common case in this HC. He reports all the malaria cases come to the HC to the health operational district.

7.2.5 Interview with business owner working with farmers in Pailin province

Interviewee: Mr. Kengo Kitaura (Founder of Agribuddy ltd.)

Basically, the hotspots shown in this map seems to align with the actual risky areas. However, care needs to be taken for the reliability of source data. As is often the case in other areas, falsifications and misreporting can be observed in the data reporting system in Cambodia. Another issue is, in the current situation in Cambodia, it is not easy to make community residents understood and take actions based on the information from risk prediction. They cannot imagine the merit of intervention with only the future prediction. Instead of that, it may be better to design the incentive system for them to take any actions. In other words, it is important to consider how the intervention can be realized by sustaining the action.

Figure 37 Photography with Mr. Phin Chanmonin and health center staff (October 2015)

8. DISCUSSION

8.1 Results overview

As malaria elimination effort progresses, it became increasingly important to identify the residual foci of malaria transmission for addressing the remaining challenges such as preventing the residual transmission to protect immunologically susceptible populations from developing serious infectious symptoms and emerging artemisinin-resistant malaria. The maps created here can be utilized for the facilitation of efficient health resource distributions to make use of limited resources for areas where the effort for the malaria elimination has been made. This approach enables the targeted surveillance, preventive measures and monitoring for the treatment failure requiring intensive support for local health practitioners. Previous report in this area suggested that remarkable proportions of patients still had parasitaemia on day 3 after starting treatment of artemisinin combination therapy whereas symptom resolutions were seen in this period [17]. Thus, the treatment monitoring is important for preventing patient from discontinuing treatment and developing drug resistance, which requires the intensive support and supervision of local health practitioners. Interestingly, visual representations of the maps created here were similar to those of Malaria Atlas Project and displayed the finer level of risk distributions. Several differences were also seen and can be partially explained by spatial and temporal variations of source data. However, clearly highlighted areas at high-predicted risk can be distinguished from the other regions at low risk in finer scale, while providing information to quantify the expected outcomes from the combination of containment status, suggested that these fine-scale maps play important roles under current situations in Cambodia. Validation by comparing predicted risk with geocoded case data confirmed that the predicted areas at high risk of transmission are likely the areas where the attention and appropriate support are needed.

8.2 Application of SMR for spatial risk distribution modeling

This research also describes an application of SMR using API reported in routine aggregated surveillance data to quantify the spatial distributions of risk by capturing the environmental context and containment status indicators in the model under the

low-to-moderate transmission settings. Although there is room for investigations for its interpretation, API can be a reliable measure under the low malaria transmission setting [46]. We saw the remaining or even increasing tendency of SMR reflecting the relative risk of respective areas during the research period, which can be a useful measure for deciding the allocation of limited healthcare resources at least for a few years. Sturrock et al. built a prediction model using routine aggregated case data and created a fine-scale risk map in Swaziland [49]. In their model, mean temperature and travel time to health facilities are the predictors of both the pixel scale and the coarser district scale of risk. Whereas Lowe et al. reported various kinds of predictors such as altitude, living conditions, urbanizations, precipitations and temperature [57]. Variables chosen for our model were in line with these reports, in terms of using environmental and human behavior-related variables for malaria risk predictions. Although the altitude was thought to relate to malaria ecology, it was not incorporated into the model. Nevertheless, the risk was well explained. This could be explained by relatively flat terrain consisting most of this area, which did not affect the transmission pattern.

8.3 Environmental context disparities

Of note, the data collection distance from each village for environmental variables affected risk predictions of the model. Besides, the selected distance for model development for vegetation (NDVI) and water related variables (NDWI) were different each other. This partially describes human interactions with what exists around living communities. The relationships of malaria incidence and *Anopheles* mosquito numbers with distance from mosquito bleeding sites were reported in previous studies [73-75]. According to the surveillance report [56, 76-77], malaria prevalence was decreased by distance from forests. Also a number of studies have addressed the relationship with distance from environmental features for malaria spatial risk distribution modeling such as the proximity of water puddle [78], health facilities [49]. This distance effect for vegetation and water indices indicates that such environmental features are interrelating to human living communities by different way. Forest workers are usually going and working in the forests several kilometers away from their living communities whereas the activity range of vectors are limited to few distances from their breeding habitat. The maps created here suggested that the spatial heterogeneity of disease risk could be

explained by such environmental context disparities. The results suggest that the distance from living communities can serve as a useful reference when considering the environmental context. Through this approach, the relative risk specified from surrounding environmental context can be described over a wide area, while maintaining the uniformity of unknown conditions, using remote sensing data by earth observations of space satellites.

8.4 Implications from computational simulations

In Swaziland, the fine scale map was created using household level data. However, this kind of micro data is often inaccessible, hence they cannot be used for mapping. The encouraging results for fine-scale risk prediction here in the modeling framework enabled the visualization of the effect size from combinations of containment status indicators. The simulation results suggested that the predicted outcomes of containment status were different among each environmental context. Therefore, the results provided an opportunity for evaluating the interventions considering these environmental situations in target areas. Moreover, the simulation results of expected interventional outcomes can be mapped, by which the decision-maker can assess combinations of the interventional approach considering several constraints such as detailed characteristics, specific issues and resource constraints in the target area. Generally, under the low-to-moderate transmission settings, the situation surrounding malaria containment actions differs by area. For instance, the coverage of LLIN is higher in some area where the high malaria prevalence was reported and interventional effort has been made for that. Therefore the incremental cost for improving the conditions of each activity may also be different. Figure 38 shows an example of probability sensitivity analysis using the model developed here. The expected outcomes attained as containment status indicators differ by actions. The decision needs to be made based on the predictions. This example demonstrates the possibility for providing the tailored information for the targeted hotspot, by which the decision make can examine the action alternatives based on the required balance of cost and effectiveness.

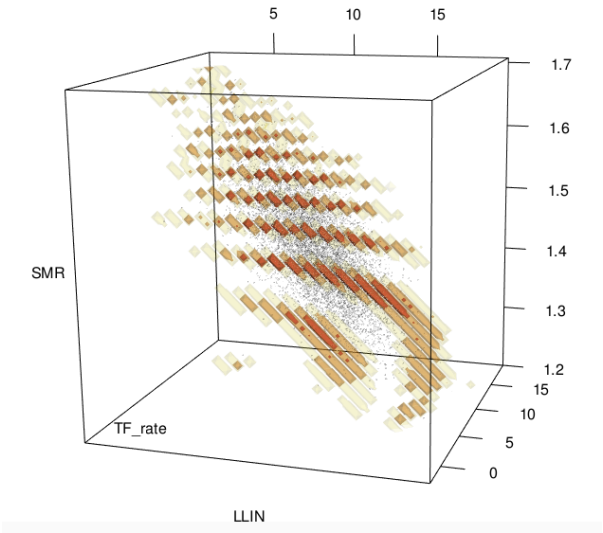


Figure 38 An example of provability sensitivity analysis under given containment status indicators in Preah Vihear province

Red-colored grid represents likelihood of expected outcomes under given containment status. Darker color indicates the higher likelihood expected from this simulation.

8.5 Reliability of data

Generally, the reliability of data is one critical factor for creating relevant models to be used in the real world practice. The current malaria reporting system in Cambodia relies on the aggregation of field report from village malaria workers and healthcare practitioners to the district, then the province and eventually the national level. Under the low transmission setting, it gets harder for the passive surveillance to capture the reliable case number reflecting the actual situations [79]. Although we cannot deny the possibilities of variations in the reliability of data, this mapping approach can add more reciprocity among stakeholders than simply recording the aggregated case numbers, which will encourage more effective report-and-utilization cycles improving the data quality and reliability. Therefore, this will complement the recent mobile phone based real time case reporting system [54], providing an opportunity for effective data utilization. Considering the structure of Cambodian malaria health information system, the system can be considered as a system of systems in its nature and output. In the last few years, SoS has been gaining increased attention as a means to understand the high

complexity of metasystems [80]. Boardman and Sauser proposed the five main properties of SoS: autonomy, belonging, connectivity, diversity, and emergence [81], [82]. Table 7 summarizes the properties thought to be applicable to be a SoS that the Cambodian information system has. As shown in Figure 39, the system has five or more independent systems: Health ministry, NGO such as Malaria consortium, Health operational district, Health center and the network of village malaria workers. All component system acts independently and cooperatively interacting each other for the purpose to collect the precise regional malaria risk to utilize malaria containment. Therefore, to this fact, one needs to understand the effectiveness of the system developed here in the system of systems.

Table 7 Summary of properties applicable for those of the system of systems

Properties	Descriptions
Autonomy	Cambodian malaria health information as a whole aims to provide updated precise information of disease prevalence. Currently this system succeeds in providing these information to public periodically. The reported information are managed or consolidated in the whole body of the system.
Belonging	The relationship among the national agency, health operational districts, provincial health centers and village malaria workers are connected as resource provision and the pathway of reported information flow. Some stakeholders such as village malaria workers are, in some occasion, independent from ownership of higher entity. Furthermore the supporters from outside of Cambodia such as NGO also play and provide important contributions. Malaria information system provided by Malaria Consortium is an example of unique contribution of the component system that is independently managed from the whole system. For the case of malaria, they interoperate together for the same purpose while sharing the mission of malaria elimination by 2025.
Connectivity	Despite component systems are geographically isolated, they are connected by reporting and resource provisional pathway. Furthermore, all the component systems have intimate relationship to communicate regularly each other to share the information, support and expertise.
Diversity	The system consists of various types of component systems such as public agencies, village health volunteers, NGO and Malaria Consortium, which allows the diversity and flexibility of the system capability.
Emergence	The system of village malaria workers sometimes, and also basically, the support from foreign agencies, foundations and NGOs are based on the voluntary aids of people.

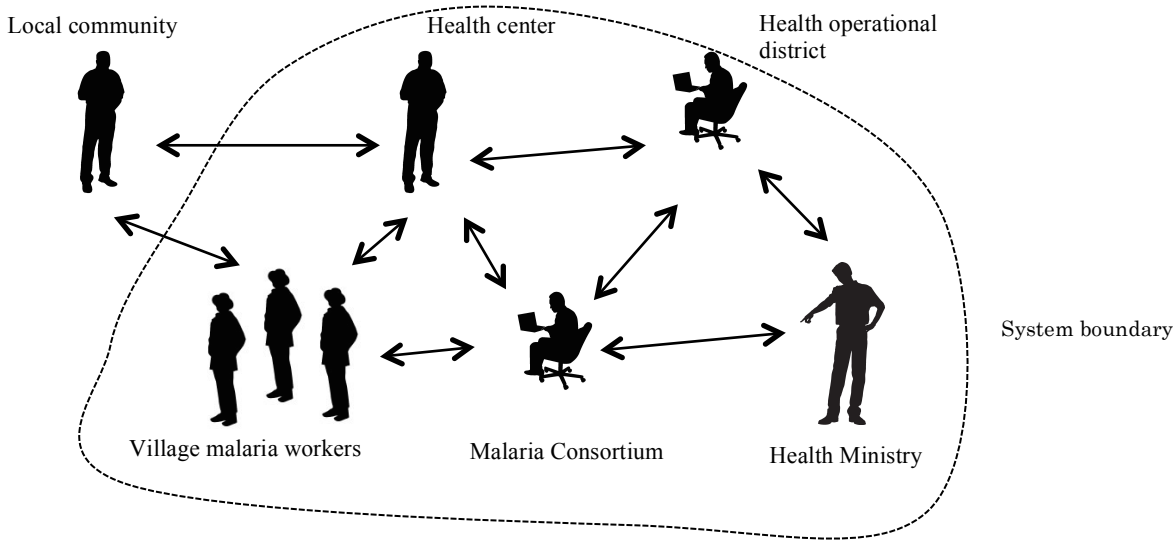


Figure 39 An example of the concept graphic of the Cambodian malaria health information system

8.6 Prospects of the value of the system

In spite of these considerations, like any programs, the malaria elimination action needs specific plans with realistic time limits and well-defined parasitological and entomological goals [46]. Maps created by the modeling framework here can provide the insight for establishing realistic goals based on the current tools. Furthermore, they can provide the useful information both in quantitatively and qualitatively for monitoring and evaluations of interventional activities while the providing decision-makers with a platform for cross-scale wandering to make a decision for the efficient healthcare resource use. Our approach outlined here is no more than a quantitative prediction technique for further utilization of existing dataset, thus may play only a part in the whole healthcare information system for malaria elimination. For sure, we might have to consider the possibilities of the divergence of prediction from real world. Still, these adjustments in malaria quantification provide us with further steps, working together in the system, toward malaria elimination.

9. FUTURE WORK

9.1 Limitations of this research

While our approach outlined here has generated several supportive results in terms for fine-scale risk prediction under low-to-moderate transmission settings, it has several important limitations and consideration for the future work. Firstly, we didn't consider other containment status indicators than $LLIN_{suf}$ and TF_{rate} for the development of modeling framework. The expected outcomes of interventional efforts may be provided from the results of various activities. This could not be explained by a simple additive effect but by a synergetic effect through the interaction of these activities. We considered this interaction between $LLIN_{rate}$ and TF_{rate}, however the result was not improved. Therefore, this issue has not been fully elucidated and should be considered for practical applications to describe the complex realty for assessing the effectiveness of interventions. Secondly, we didn't consider the influence of migrant population and time series variation of the risk in the modeling framework. The influence of these two factors may be important for considering practical situations. The dynamics of human carriers that drive parasite transportation between regions can be quantified using spatially explicit mobile phone data and malaria prevalence information [2]. More useful models can be developed incorporating such factors as people migrations between the regional boundaries and time series element into the modeling framework. As we used API for the calculation of SMR, the predictor variables related to environment were boiled down to yearly average. As is the same case in the spatial granularity of data, deciding appropriateness of the time granularity is a perplexing issue, because of the difficulties for catching adequate case numbers for the reliable spatial risk distribution modeling from detail level data. Furthermore, the appropriateness for deciding the region of interest for data collection is also unknown factor to be elucidated, as the calculation of the denominator for SMR is influenced by this factor.

9.2 Understanding the effectiveness of the system

One issue found in the research is reliability of reported data in Cambodia. There is possibility that our mapping approach can add more reciprocity among stakeholders

than simply recording the aggregated case numbers. Throughout the facilitation of more effective report-and-utilization cycles, data quality and reliability can be improved. However, this is not elucidated qualitatively and quantitatively. To address this hypothesis, we need to understand the structure of the health information system, by which the causal relationship of suspected data quality could be more clarified. The Cambodian health information system is expected to be a system of systems. Understanding the behavior and characteristics of SoS by systemic approach is one important topic of interest [80]. One of the metrics being interested for the quantification is the efficiency of information understanding among the stakeholders. It is still a challenging question but meaningful to understand the usefulness of this system.

9.3 Applications to other epidemiological issues

Strength of our approach is utilization of existing dataset. Generally, most of such data sources are open to public. In that sense, several more useful information can be added to the aggregated surveillance report in a real time manner. One area that we can consider this approach make contribution is the real time monitoring of emerging infectious disease such as Ebora [83] and recent severe acute respiratory syndrome (SARS) corona virus out break [84]. Also the application to non-communicable disease such as life-style disease needs to be elucidated considering future transition of disease structure globally. Recent advancement in the information technology enables the real-time collection of various kinds of data surrounding people. The recognition of Internet of Things (IoT) [85] is expanding and was actively introduced into the policy development in several countries [86]. Thus, we expect that the environmental context related to non-communicable disease could be understood using these advancing technologies such as atypical information surrounding people.

10. CONCLUSION

We conclude the validity of the system. Using routine aggregated surveillance reports combined with the environmental data and the non-environmental anthropogenic data, regional malaria risks can be well explained. Using this modeling framework, the fine-scale risk maps can be obtained under the low-to-moderate transmission setting where reinvestigations of existing risk modeling approaches are needed. Our contributions are; demonstration of the mathematical modeling approach for SMR using API from routine aggregated surveillance report, cross-scale prediction by modeling framework corresponding to environmental context disparities to create malaria risk maps in fine-scale. Different representations of simulated outcomes from containment status indicators provide us with useful insight for tailored planning of action alternatives considering regional malaria endemicity.

Further studies are needed to demonstrate how this system will provide the effect in the Cambodian health information system. The implications from this study suggested the system would provide increased reciprocity of the information, by which the improved quality of reported data could be expected.

11. PUBLICATON

This thesis has allowed the following scientific productions.

Conference

[1] Suguru Okami, Naohiko Kohtake, “Designing the GIS predicting regional malaria emdemicity in Cambodia”, Esri Health and Human Services GIS Conference, Grand Hyatt Atlanta in Buckhead, Atlanta, Georgia, 14 - 16 September 2015.

Journal

[1] Suguru Okami, Naohiko Kohtake, “Fine-scale mapping by spatial risk distribution modeling for regional malaria endemicity and its implications under the low-to-moderate transmission setting in western Cambodia” (submitted).

12. BIBLIOGRAPHY

[1]. Nakaya T (2008) Spatial epidemiology in Geographic Information System environments. *J. Natl. Inst. Public Health* **57** (2): 99-116.

[2]. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, et al. (2012) Quantifying the impact of human mobility on malaria. *Science* **338**: 267-270.

[3]. Hay SI, Snow RI, Rogers DJ. (1998) From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives. *Parasitology Today* **14** (8): 306-313.

[4]. Shimodaira H, Nozaki M, Kwon Y, Kamimura N, Kaiho F (2014) Analysis of adverse reaction in Kampo-medicines using JADER database of PMDA. *Jpn. J. Drug Inform.* **16** (1): 16-22.

[5]. World Health Organization (2015) Fact sheet N^{o} 94. Geneva: World Health Organization.

[6]. WHO-UNICEF Child Health Epidemiology Reference Group (2013) Causes of deaths among children under 5 years, 2011. Geneva: World Health Organization.

[7]. World Health Organization (2014) WHO report Trends in reported malaria incidence, 2000-2012. Geneva: World Health Organization.

[8]. World Health Organization (2007) Malaria elimination. A field manual for low and endemic countries. Geneva: World Health Organization.

[9]. Tatem AJ, Smith DL, Gething PW, Kabaria CW, Snow RW, Hay SI (2010) Ranking of elimination feasibility between malaria-endemic countries. *Lancet* **376**: 1579-1591.

[10]. National Center for Parasitology, Entomology and Malaria Control (2011) Strategic plan for elimination of malaria in Cambodia 2011-2025. Phnom Penh, Cambodia: Ministry of Health.

[11]. Maude RJ, Nguon C, Ly P, Bunkea T, Ngor P, et al. (2014) Spatial and temporal epidemiology of clinical malaria in Cambodia 2004-2013 *Malaria Journal* **13**: 385.

[12]. World Health Organization (2014) World malaria report 2014 Geneva: World Health Organization.

[13]. World Health Organization (2013) Emergency response to artemisinin resistance in the Greater Mekong subregion. Regional framework for action 2013-2015. Geneva: World Health Organization.

[14]. World Health Organization (2015) Guideline for the treatment of malaria. Third edition. Geneva: World Health Organization.

[15]. World Health Organization (2015) Q&A on artemisinin resistance. Geneva: World Health Organization.

[16]. World Health Organization (2015) Update on artemisinin and ACT resisitance – September 2015. Geneva: World Health Organization.

[17]. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, et al. (2014) Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *New Engl J Med* **371**: 411-423.

[18]. Bosman P, Stassijns J, Nackers F, Canier L, Kim N, et al. (2014) *Plasmodium* prevalence and artemisinin-resistant falciparum malaria in Preah Vihear province, Cambodia: a cross-sectional population-based study. *Malaria Journal* **13**: 394.

[19]. Amaratunga C, Sreng S, Suon, S, Phelps ES, Stepniewska K, et al. (2012) Artemisinin-resistant *Plasmodium falciparum* in Pursat province, western Cambodia: a parasite clearance rate study. *Lancet Infect Dis.* **12**(11): 851-858.

[20]. Laurent BS, Miller B, Burton TA, Amaratunga C, Men S, et al. (2015) Artemisinin-resistant *Plasmodium falciparum* clinical isolates can infect diverse mosquito vectors of Southeast Asia and Africa. *Nature Communications* **6**: 8614 doi: 10.1038/ncomms9614.

[21]. Tabernero P, Fernández FM, Green M, Guerin PJ, Newton PN. (2014) Mind the gaps – the epidemiology of poor-quality antimalarials in the malarious world – analysis of the WorldWide Antimalarial Resistance Network database. *Malaria Journal* **13**: 139.

[22]. Leang R, Barrette A, Bouth DM, Menard D, Abdur R, et al. (2013) Efficacy of Dihydroartemisinin-piperaquine for treatment of uncomplicated *Plasmodium falciparum* and *Plasmodium vivax* in Cambodia 2008 to 2010. *Antimicrobial Agents and Chemotherapy* **57**: 818-826.

[23]. Leang R, Taylor WRJ, Bouth DM, Song L, Tarning, J, et al. (2015) Evidence of Plasmodium falciparum malaria multidrug resistance to artemisinin and piperaquine in western Cambodia: Dihydroartemisinin-piperaquine open-label

multicenter clinical assessment. *Antimicrobial Agents and Chemotherapy* **59**(8): 4719-4726.

[24]. National Center for Parasitology, Entomology and Malaria Control (2014) Benefits to national malaria programs from regional support: the Cambodia case Phnom Penh, Cambodia: Ministry of Health

[25]. Hoyer S, Nguon S, Kim S, Habib N, Khim N, Sum N, et al. (2012) Focused screening and treatment (FSAT): A PCR-based strategy to detect malaria parasite carriers and contain drug resistant *P. falciparum*, Pailin, Cambodia. *PLOS ONE* **7**(10): e45797.

[26]. Cox J, Soley LD, Bunkea T, Sovannaroth S, Ty KS, et al. (2014) Evaluation of community-based systems for the surveillance of day three-positive *Plasmodium falciparum* cases in western Cambodia. *Malaria Journal* **13**: 282.

[27]. Lwin KM, Imwong M, Suangkanarat P, Jeeyapant A, Vihokhern B, et al. (2015) Elimination of *Plasmodium falciparum* in an area of multi-drug resistance. *Malaria Journal* **14**: 319.

[28]. Sabot O, Tulloch J, Basu S, Dyckman W, Moonasar D, Moonen B. (2009) Getting to zero. *A Plospectus on Malaria Elimination* 19-39.

[29]. Martin C, Curtis B, Fraser C, Sharp B. (2002) The use of GIS-based malaria information system for malaria research and control in South Africa. *Health & Place* **8**: 227-236.

[30]. Clements AC, Reid HL, Kelly GC, Hay SI. (2013) Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? *Lancet Infect Dis.* **13**(8): 709-718.

[31]. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, et al. (2009) A world malaria map: Plasmodium falciparum endemicity in 2007. *PLOS Medicine* **6**(3): e1000048.

[32]. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, et al. (2011) A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malaria Journal* **10**: 378.

[33]. Patil AP, Gething PW, Piel FB, Hay SI. (2011) Bayesian geostaticstics in health cartography: the perspective of malaria. *Trends in Parasitology* **27**(6): 245-252.

[34]. Nihei N, Hashida Y, Kobayashi M, Ishii A. (2002) Analysis of malaria endemic areas on the Indochina Peninsula using remote sensing. *Jpn J Infect Dis.* **55**(5): 160-166.

[35]. Cohen JM, Ernst KC, Lindblade KA, Vulule JM, John CC, et al. (2010) Local topographic wetness indices predict household malaria risk better than land-use and land-cover in the western Kenya highlands. *Malaria Journal* **9**: 328.

[36]. MacCann RS, Messina JP, MacFarlane DW, Bayoh MN, Vulule JM, et al. (2014) Modeling larval malaria vector habitat locations using landscape features and cumulative precipitation measures. *International Journal of Health Geographics* **13**: 17.

[37]. Cianci D, Hartemink N, Justicia I. (2015) Modeling the potential spatial distribution of mosquito species using three different techniques. *International Journal of Health Geographics* **14**: 10.

[38]. Craig MH, Snow RW, Sueur DI. (1999) A climate-based distribution model of malaria transmission in Sub-Saharan Africa. *Parasitology Today* **15**(3): 105-111.

[39]. Oloukoi G, Bob U, Jaggernath J. Perception and trends of associate health risks with seasonal climate variation in Oke-Ogun region, Nigeria. *Health & Place* **25**: 47-55.

[40]. Cohen JM, Dlamini S, Novotny JM, Kandula D, Kunene S, et al. (2013) Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malaria Journal* **12**: 61.

[41]. Macdonald, G, (1957) The epidemiology and control of malaria. London: Oxford University press.

[42]. Dietz K., Molineaux L, Thomas A. (1974) A malaria model tested in the African savannah. *Bulletin of the World Health Organization* **50**: 347-357.

[43]. Chikodzi D. (2013) Spatial modeling of malaria risk zones using environmental, anthropogenic variables and geographical information systems techniques. *Journal of Geoscience and Geomatics* **1**(1): 8-14.

[44]. Ellis EC, Ramankutty N. (2008) Putting people in the map: anthropogenic biomes of the world. *Front Ecol. Environ.* 6(8): 439–447.

[45]. Hay SI, Smith DL, Snow RW. (2008) Measuring malaria endemicity from intense to interrupted transmission. *Lancet Infect Dis* **8**: 369-378.

[46]. Smith DL, Smith TA, Hay SI. (2009) Measuring malaria for elimination. *A Plospectus on Malaria Elimination* 108-126.

[47]. Okell LC, Bousema T, Griffin JT, Ouédraogo AL, Ghani AC. (2012) Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. *Nature Communications* **3**: 1237.

[48]. Hoyer S, Nguon S, Kim S, Habib N, Khim N, et al. Focused screening and treatment (FSAT): A PCR-based strategy to detect malaria parasite carriers and contain drug resistant *P. falciparum*, Pailin, Cambodia. *PLOS ONE* **7**(10): e45797.

[49]. Sturrock HJW, Cohen JM, Keil P, Tatem AJ, Menach AL, et al. (2014) Fine-scale risk mapping from routine aggregated case data. *Malaria Journal* **13**: 421.

[50]. Center for Disease Control and Prevention. (2006) Principles of epidemiology in public health practice, third edition. An introduction to applied epidemiology and biostatistics. Atlanta, Georgia: CDC.

[51]. OMG. (2016) OMG systems modeling language [Online]. Available: http://www.omgsysml.org.

[52]. National Center for Parasitology, Entomology and Malaria Control. (2011) Cambodia malaria bulletin, December 2011. Phnom Penh, Cambodia: Ministry of Health.

[53]. National Center for Parasitology, Entomology and Malaria Control. (2013) Cambodia malaria bulletin, December 2013. Phnom Penh, Cambodia: Ministry of Health.

[54]. Mellor, S. (2013) Moving towards malaria elimination: developing innovative tools for malaria surveillance in Cambodia. *www.malariaconsortium.org/pages/learning-paper.htm*

[55]. Tango T, Yokoyama T, Takahashi K. (2007) *Introduction to spatial epidemiology*. Tokyo, Japan: Asakura publishing.

[56]. National Center for Parasitology, Entomology and Malaria Control. (2010) Cambodia malaria survey 2010. Phnom Penh, Cambodia: Ministry of Health.

[57]. Lowe R, Chirombo J, Tompkins AM. (2013) Relative importance of climatic, geographic and socio-economic determinants of malaria in Malawi. *Malaria Journal* **12**: 416.

[58]. Martuzzi M, Elliott P. (1996) Empirical Bayese estimation of small area prevalence of non-rare conditions. *Statistics in Medicine.* **15**: 1867-1873.

[59]. Mayer DJ, Tachikawa T, Abrams M, Crippen R, Krieger T, et al. (2012) Summary of the validation of the second version of the ASTER GDEM. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* XXXIX-B4: 291-293.

[60]. Nmor JC, Sunahara T, Goto K, Futami K, Sonye G, et al. (2013) Topographic models for predicting malaria vector breeding habitats: potential tools for vector control managers. *Parasites & Vectors* **6**: 14.

[61]. Gething PW, Boeckel TPV, Smith DL, Guerra CA, Patil AP, et al. (2011) Modeling the global constraints on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasites & Vectors* **4**: 92.

[62]. Moyes CL, Temperley WH, Henry AJ, Burgert CR, Hay SI. (2013) Providing open acess data online to advance malaria research and control. *Malaria Journal* **12**: 161.

[63]. Tatem AJ, Guerra CA, Kabaria CW, Noor AM, Hay SI. (2008) Human population, urban settlement patterns and their impact on *Plasmodium falciparum* malaria endemicity. *Malaria Journal* **7**: 218.

[64]. Tatem AJ, Gething PW, Smith DL, Hay SI. (2013) Urbanization and the global malaria recession. *Malaria Journal* **12**: 133.

[65]. National Center for Parasitology, Entomology and Malaria Control. (2014) Benefits to national malaria programs from regional support. Phnom Penh, Cambodia: Ministry of Health.

[66]. Kurshcke JK. (2014) *Doing Bayesian data analysis, 2nd edition.* Academic Press – Elsevier.

[67]. Brooks SP, Gelman A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7**(4): 434-455.

[68]. Médecins Sans Frontières (2005) Unpublished dataset.

[69]. Incadona S, Vong S, Chiv L, Lim P, Nhem S, et al. (2007) Large-scale malaria survey in Cambodia: Novel insights on species distribution and risk factors. *Malaria Journal* **6**: 37.

[70]. Altman DG. (1991) *Practical statistics for medical research.* London: Capman & Hall.

[71]. Welch BL. (1938) The significance of the difference between two means when the population variances are unequal. *Biometrica* **29**: 350-362.

[72]. Carpenter J, Bithell J, et al. (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 19: 1141-1164.

[73]. Trape J-F, Lefebvre-Zante E, Legros F, Ndiaye G, Bouganali H, et al. (1992) Vector density gradients and the epidemiology of urban malaria in Dakar. *American Journal of Tropical Medicine and Hygiene* **47**: 181-189.

[74]. Ghebreyesus TA, Haile M, Witten KH, Getachew A, Yohannes AM, et al. (1999) Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey. *British Medical Journal* **319**: 663-666.

[75]. Carter R, Mendis KN, Roberts D, et al. (2000) Spatial targeting of interventions against malaria. *Bulletin of the World Health Organization* **78**(12): 1401-1411.

[76]. National Center for Parasitology, Entomology and Malaria Control. (2005) Report of the Cambodia national malaria baseline survey 2004. Phnom Penh, Cambodia: Ministry of Health.

[77]. National Center for Parasitology, Entomology and Malaria Control. (2007) Cambodia malaria survey 2007 report. Phnom Penh, Cambodia: Ministry of Health.

[78]. Cohen JM, Ernst KC, Lindblade KA, Vulule JM, John CC, Wilson ML. (2008) Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands. *Malaria Journal* **7**: 40.

[79]. Targett GA, Yeung S, Tanner M. (2009) Identifying the gaps – what we need to know. *A Prospectus on Malaria Elimination.*

[80]. Tsilipanos K, Neokosmidis I, Varoutas D. (2013) A system of systems framework for the reliability assessment of telecommunications networks. *IEEE Systems Journal* **7**:114-124.

[81]. Boardman J and Sauser B. (2006) System of systems: The meaning of of. *Proc. IEEE/SMC Int. Conf. Syst. Syst. Eng.*, Apr. p. 6.

[82]. Baldwin W.C. and Sauser B. (2009) Modeling the characteristics of system of systems. Proc. *IEEE Int. Conf. SoSE*, May–Jun. pp. 1–6.

[83]. Center for Diease Control and Prevention. (2016) 2014 Ebora outbreak in west Africa [Online]. Available: http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/.

[84]. Center for Disease Congtrol and Prevention. (2004) Fact sheet for SARS patients and their close contacts. Atlanta, Georgea: CDC.

[85]. Ashton K. (2009) That 'Internet of Things' Thing. *RFID Journal* 22 June. New York: RFID Journal LLC.

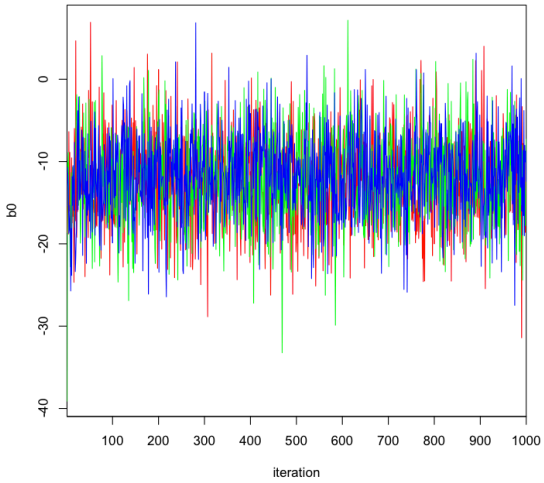
[86]. Kagermann H, Wahlster W, Helbig J, et al. (2013) Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group. Germany: The industry-science research alliance.

*Referenced website URLs are available as of January 2016.

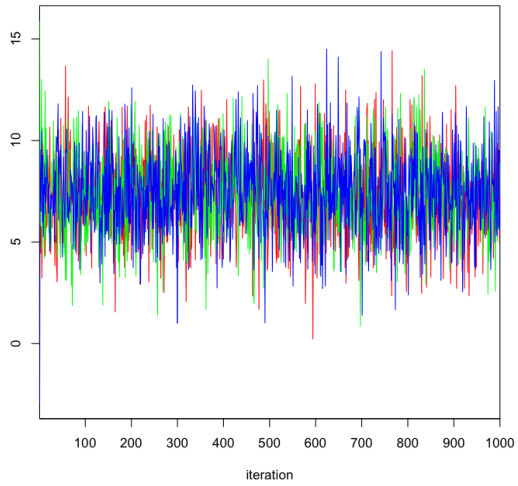
13. APPENDIX

Following figures are the trace plots of the Bayesian modeling frame to examine the convergence of each parameter.

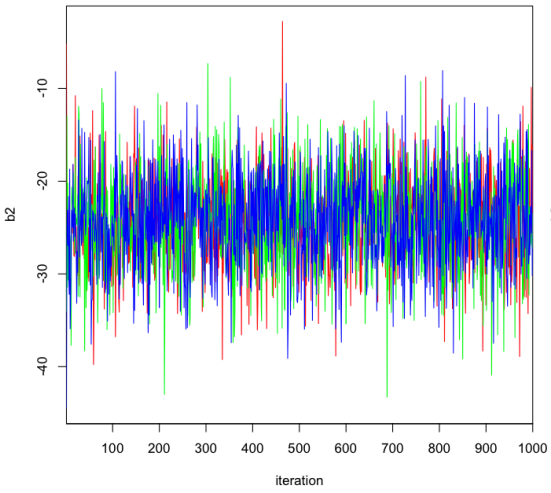
A



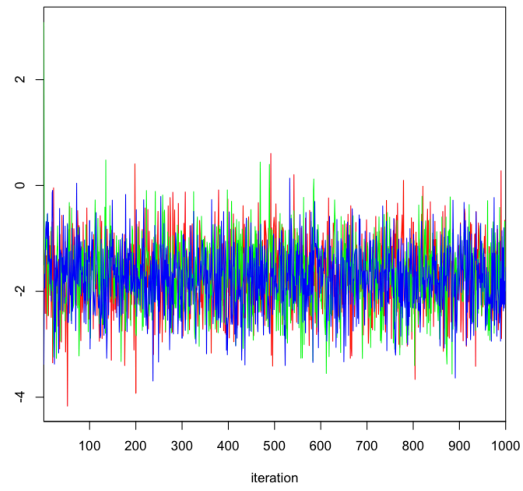
B



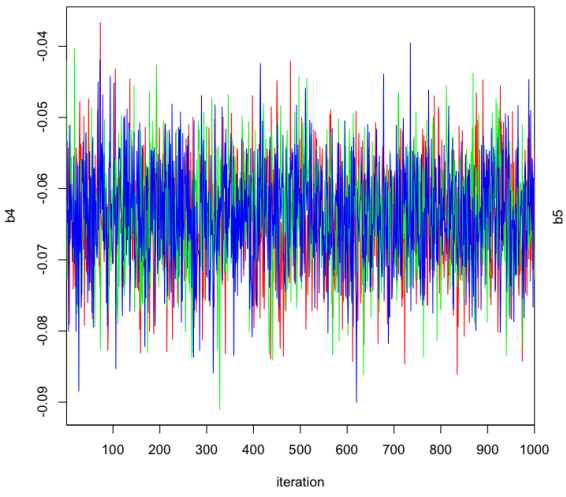
C



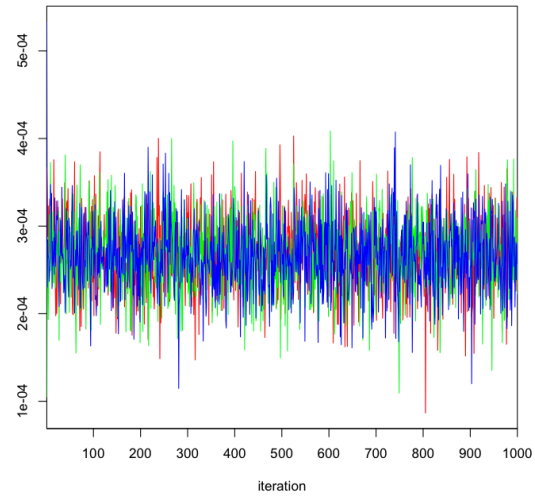
D



E



F



G

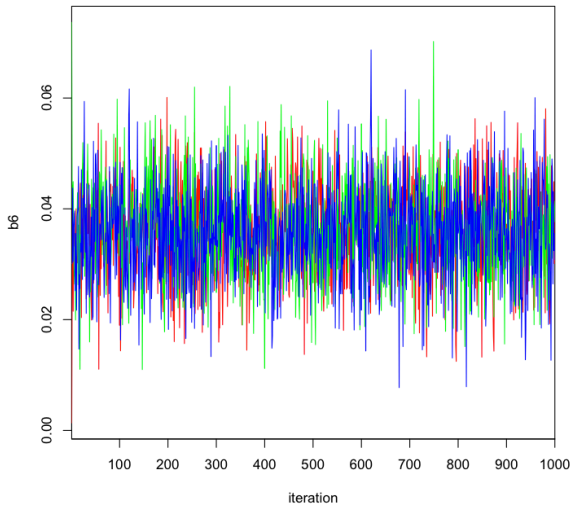


Figure 40 Trace plot of the parameter in the Bayesian modeling for cross prediction for fine-scale malaria risk

Parameter distributions for (A): Intercept, (B): NDVI, (C): NDWI,
(D): TWI, (E): LLIN, (F): Temperature and (G): TF
NDWI, Normalized difference water index; NDVI, Normalized difference vegetation index;
TWI, Topographical wetness index; LLIN, Long lasting insecticide-treated net

ACRONYM LIST

API	Annual Parasite Incidence
CNM	National Center for Parasitology, Entomology and Malaria Control
CRS	Coordinate Reference System
CVCA	Customer Chain Value Analysis
DEM	Digital Elevation Model
EBSMR	Standardized Morbidity Ratio Calculated by Empirical Bayese Method
FFBD	Functional Flow Block Diagram
GIS	Geographical Information System
HC	Health Center
HIS	Health Information System
HOD	Health Operational District
IoT	Internet of Things
LLIN	Long-Lasting Insecticide Treated Mosquito Net
LSWI	Land Surface Water Index
MCMC	Marcov Chain Monte Carlo
MIS	Malaria Information System
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NGO	Non Governmental Organization
PR	Parasite Rate
*Pf*TSI	*Prasmodium falciparum* Temperature Suitability Index
RDT	Rapid Diagnostic Testing Kit
SMR	Standardized Morbidity Ratio
SoS	System of Systems
TWI	Topograhical Wetness Index
VMW	Village Malaria Worker
WHO	World Health Organization

R-code for spatial modeling

```
library(R2jags)
library(bayesm)
library(mgcv)

#Data load
#Regression modeling
data <- read.table("R_Analysis.csv", sep = ",", header = T)
EBSMR.lm <- lm(log(EBSMR)~NDVI_mean_5000+NDWI_1000+TWI_5000+Temp+LLIN_Suf+TF_rate, data =
 data)
summary(EBSMR.lm)

pred.EBSMR <- predict(EBSRR.lm)
plot(log(data1$EBSRR), pred.EBSRR, xlab = "actual value", ylab = "predicted value")
abline(0, 1)

#Preparing Data frame
y <- as.vector(log(data$EBSMR))
n <- length(y)
x <- as.matrix(cbind(data$NDVI_mean_5000, data$NDWI_1000, data$TWI_5000, data$LLIN_Suf,
 data$Temp, data$TF_rate))
model.lm <- lm(y~x)
summary(model.lm)

data <- list("n", "y", "x")
in1 <- list(b0 = model.lm$coefficients[1], b1 = model.lm$coefficients[2], b2 =
 model.lm$coefficients[3], b3 = model.lm$coefficients[4], b4 = model.lm$coefficinets[5], b5
 = model.lm$coefficients[6], b6 = model.lm$coefficients[7], tau = 1)
in2 <- list(b0 = model.lm$coefficients[1], b1 = model.lm$coefficients[2], b2 =
 model.lm$coefficients[3], b3 = model.lm$coefficients[4], b4 = model.lm$coefficinets[5], b5
 = model.lm$coefficients[6], b6 = model.lm$coefficients[7], tau = 1)
in3 <- list(b0 = model.lm$coefficients[1], b1 = model.lm$coefficients[2], b2 =
 model.lm$coefficients[3], b3 = model.lm$coefficients[4], b4 = model.lm$coefficinets[5], b5
 = model.lm$coefficients[6], b6 = model.lm$coefficients[7], tau = 1)

inits <- list(in1, in2, in3)

parameters <- c("b0", "b1", "b2", "b3", "b4", "b5", "b6", "tau", "sigma")
model.file <- system.file(package = "R2jags", "model", "Spatial_modeling_jags")
```

```r
#Bayesian regression modeling
lm.jags <- jags(data = data, inits = inits, parameters, n.iter = 80000, n.burnin = 50000,
 n.chains = 3, model.file = model.file)

print(lm.jags, digits = 5)
plot(lm.jags)
traceplot(lm.jags)

lm.fit <- update(lm.jags)
print(lm.fit, digits = 5)

#Plot samples
jags_samples <- as.mcmc(lm.jags)
plot(jags_samples[, 'b0'])
plot(jags_samples[, 'b1'])
plot(jags_samples[, 'b2'])
plot(jags_samples[, 'b3'])
plot(jags_samples[, 'b4'])
plot(jags_samples[, 'b5'])
plot(jags_samples[, 'b6'])
plot(jags_samples[, 'tau'])
plot(jags_samples[, 'sigma'])
```

JAGS code for the Bayesian modeling frame

```
model{
for(i in 1:n){
y[i]~dnorm(mu[i], tau)
mu[i] <- b0 + b1*x[i, 1] + b2*x[i, 2] + b3*x[i, 3] + b4*x[i, 4] + b5*x[i, 5] + b6*x[i, 6]
}

b0~dnorm(0, 1.0E-6)
b1~dnorm(0, 1.0E-6)
b2~dnorm(0, 1.0E-6)
b3~dnorm(0, 1.0E-6)
b4~dnorm(0, 1.0E-6)
b5~dnorm(0, 1.0E-6)
b6~dnorm(0, 1.0E-6)
tau~dgamma(0.01, 0.01)
sigma <- 1/tau
}
```

R-code for interpolation for fine-scale mapping

```
library(spdep)
library(gstat)
library(maptools)
library(spsurvey)
library(ggplot2)
library(rgdal)

#Data load
Pailin <- read.shape("Pailin_village_32648.shp")
mesh.grid <- read.shape("Pailin_mesh.shp")
mesh.grid <- as(mesh.grid, "SpatialPixelsDataFrame")
Pol <- read.shape("Pailin_poly.shp")

#Plot variogram
var0 <- variogram(SMR~1, data = Pailin)
plot(var0)
var1 <- variogram(SMR~X+Y, data = Pailin)
plot(var1)
plot(variogram(SMR~X+Y, data = Pailin, cloud = TRUE))
var2 <- variogram(SMR~X+Y, data = Pailin, alpha = 0:4*90)
plot(var2)

#Application of variogram (Gauss)
model <- vgm(psill = 0.2, model = "Gau", range = 15000, nugget = 0.02)
plot(var1, model)
model.fit <- fit.variogram(var1, model)
plot(var1, model)

#IDW
idw1 <- idw(SMR~1, Pailin, mesh.grid, idp = 2)
IDW <- as.data.frame(idw1)
names(IDW)[1:3] <- c("long", "lat", "SMR")
ggplot() + geom_tile(data = IDW, aes(x = long, y = lat, fill = SMR)) + stat_contour(data =
 IDW, aes(x = long, y = lat, z = SMR), size = 0.1) + geom_path(data = Pol, aes(x = long, y
 = lat, group = group)) + ggtitle("IDW") + scale_fill_gradient2(low = "white", mid = "yellow",
 high = "red")

#Write Geotiff
writeGDAL(idw1, fname = "IDW.tif", drivername = "GTiff", type = "Float32")
```

```
#Plot
spplot(idw1["var1.pred"], main = "IDW", scales = list(draw=T), xlab = "X Coord", ylab = "Y
 Coord")

#Contour
plot(Pol, lwd = 2)
contour(idw1["var1.pred"], add = TRUE)

#Ordinal kriging
go <- gstat(id="ID", formula = SMR~1, data = Pailin, model = model)
po <- predict(go, mesh.grid)

#plot
spplot(po[1], main = "Ordinal Kriging",  scales = list(draw=T), xlab = "X Coord", ylab = "Y
 Coord")

#Write GeoTiff
writeGDAL(po, fname = "Ordinal_Kriging.tif", drivername = "GTiff", type = "Float32")

#Display map
PO <- as.data.frame(po)
names(PO)[1:3] <- c("long", "lat", "SMR")
ggplot() + geom_tile(data = PO, aes(x = long, y = lat, fill = SMR)) + stat_contour(data =
 PO, aes(x = long, y = lat, z = SMR), size = 0.1) + geom_path(data = Pol, aes(x = long, y
 = lat, group = group)) + ggtitle("Ordinal_Kriging") + scale_fill_gradient(low = "yellow",
 high = "red")

#Contour
plot(Pol, lwd = 2)
contour(po, add = TRUE)
```