Geo-information Science and Remote Sensing

Thesis Report GIRS-2019-04

AGENT BASED MODELING: FLEXIBLE SPATIAL MODEL TO PREDICT THE SPREAD OF MALARIA.

For the Anopheles Gambiae and Plasmodium Falciparum

Luuk van Gorp

24-02-2019

WAGENINGEN UNIVERSITY & RESEARCH

AGENT BASED MODELING: FLEXIBLE SPATIAL MODEL TO PREDICT THE SPREAD OF MALARIA

For the Anopheles Gambiae and Plasmodium Falciparum

Luuk van Gorp

Registration number 950302271100

Supervisors:

Arend Ligtenberg
Corné Vreugdenhil

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

24-02-2019
Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2019-04
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Acknowledgement
Thanks to Corné Vreugdenhil and Arend Ligtenberg for guiding me through the process of building a thesis and developing an agent-based model. Thank you Sander Koenraadt for answering all my questions on the Anopheles Gambiae and Plasmodium Falciparum. Thanks to all the other teachers and staff at GAIA for being approachable and willingness to help out. Thank you fellow thesis students for listening to me while I am complaining and making the most excellent cakes. Thank you security guards for being cool even if I leave the building late. Last but not least thank you free coffee card, thou hast energized me.

INDEX

1. Introduction

Malaria is a problem across the world. It has caused many serious illnesses and even lethality amongst children below five years old. This is caused by the Plasmodium parasite spread by female Anopheles mosquitoes [1]. There are about 216 million people that were infected with malaria in 2016, 445 thousand of which ended up lethal. This is half of what used to be the case in 2000 due to the counter measures such as the insecticide treated nets and pesticides [2].

Despite the progress being made towards stopping malaria, it is still prevalent and is spreading towards more northern areas as well. This is due to the increase of traveling and the increase of global temperature. The plasmodium parasite lives best in areas where the average daily temperature is 21 degrees Celsius and this causes areas closer to the equator with high humidity to be more likely to become infected with malaria as well [3, 4].

Therefore, understanding how malaria spreads across an area is highly relevant. A way to figure this out is with the use of spatial simulation. Spatial simulations allow the user to analyse where malaria spreads and what the most important factors of this spread would be [5].

1.1. Problem statement

Malaria is an urgent issue as the sixth millennium development goal of the United Nations suggest: "Combat HIV/AIDS, malaria and other diseases" [6]. Therefore, scientists have already looked into developing malaria simulations. Currently developed simulations are often not spatial however, like the study of Killeen et al. [1]. Those that are usually focus on one area and calibrate their model for that area specifically [7-10]. The spatial component could be of high value as it would visualize the problem and add an environment to the model to interact with. This environment can influence the decisions that mosquitoes and humans would make in such a model. This visualization often makes validation of outcome and communication of information much simpler and time efficient [5].

The spatial simulations are often performed using Agent Based Modelling (ABM). This is a method whereby the mosquitoes would act as individual agents, making independent decisions based on their environment and the other agents. This is a suitable method to simulate the behaviour of individual agents and to test the influence input variables in regards to these actions. This is especially interesting for the spread of infections as the environment of the individual infectors play a large role. For example, malaria mosquitoes are largely influenced by the location of surface water for the selection of ovipositions and the locations of humans for their blood meals [5, 11].

Gu and Novak developed an ABM whereby extra attention is paid to the influence of solutions to malaria and their impact, such as the bed nets. This model used a hypothetical area and assumed that all areas have the same interest for the mosquitoes. It did not differentiate other landscapes and only looked into buildings and water locations [7].

Bomblies et al. use the HYDREMATS model to determine the malaria spread, which estimates the surface water and soil moisture using field data, topographical data, satellite imagery and meteorological data. They state that HYDREMATS works almost perfectly, except for one of the measurements [12]. This model is originally not made as an ABM, but was later developed to an ABM with high accuracy. Bomblies' research seems centred around Niger however, and does not show simulations for other areas [8, 13].

Pizzitutti et al. looked into spread of malaria in the amazon area in which they also paid attention to the different landscapes. The landscapes such as farms were given as points for the purpose of human movement instead of potential increased ovipositions for mosquitoes. They concluded that human movement does not have a large impact on the spread of malaria with the exception of movement to high risk areas [9]. This is likely because anopheles mosquitoes are mostly active during the night [14].

Jindal and Rao developed an ABM that is publicly available on GitHub. This is the only agent-based model on Malaria of which the script could be found online. The special part of this model is that it is considering any mosquito spread diseases like chikungunya and dengue. It is also the most recent model developed. Just like all the other models, this focuses on one area. It has been tested whether this model can be applied for other areas and this was not possible without making significant changes to the structure and content of the script [10].

All aforementioned models also focus on one set of inputs, for example the type of species is usually Anopheles Gambiae and is not tested for other species. The different species have small differences in their flight speed, flight range, their oviposition and blood meal cycles [15, 16]. When the same mosquito is used for different models, other values were used for the same input variables. For example, Gu and Novak and Jindal and Rao disagree on the sensory range and maximum distance a mosquito can fly per day [7, 10].

Therefore, what currently is missing in the scientific community is a simulation model that is spatial and can be implemented for different areas. The spatial aspect would make it possible for the model to determine what environmental factors influence mosquito behaviour, human behaviour and the interaction between these two agent groups. A model such as this would provide malaria analysts the basis to calculate the spread of malaria within an area of choice and could be further specified if necessary. It would also provide basis on what input variables of mosquitoes change the most per landscape and which are the most influential. It could also potentially help establishing common behaviour of specific mosquitoes in different areas.

1.2. Objective

The objective of this research is to develop an agent-based model that simulates the spread of malaria from Anopheles Gambiae mosquitoes which can read flexible spatial inputs for different areas and values.

The following research questions will be answered in order to reach the goal:

1. What are the existing models and what attributes do they use?
2. Which spatial components and data need to be used as inputs to develop a malaria model?
3. What are the processes of malaria and how can they be represented in the model?
4. How can the model be best validated for its flexibility and accuracy?

1.3. Boundaries

This research will exclude the following things:

- Application on national or larger level;
- Analyses on attributes and input variables that are agreed upon by developers;
- Mosquito species other than the Anopheles Gambiae;
- Malaria species other than Plasmodium Falciparum;
- Discussion on GAMA functionality and explanation on direct coding;
- Data that cannot be directly put into the model.

This model focuses on an individual level as much as possible. Such detail makes it impossible to run on a national or larger level due to its intensity and demand from the system. Time also plays an important role, which is why further analysis on established attributes or on other species will not be performed either. Discussion on GAMA functionality will not be performed as this research is done with the intent of creating a functional model and not to reproduce it in GAMA specifically. As this research focuses on developing the model itself, only data that is readily available will be used. Data that needs to be generated from other data or collected on the field will be excluded from this research.

1.4. Reading guide

The following chapter describes what other ABM developers have done to analyse malaria, as well as the attributes and input variables they used. The third chapter goes into the methods of the model development. This means that this chapter also already describes some decision making on what is kept and what is excluded from the model. Fourth chapter describes different results produced for this model as well as some discussion. To go in more depth, an additional fifth chapter was added that discusses the model, results and the research process. Finally, a conclusion and advice chapter answers the objective of this research and gives recommendations for future research.

2. Literature study

The information collected for this research is from different malaria agent-based models, a discussion with an entomologist and various articles on the spread of malaria.

The different existing agent-based models were selected based on whether they are spatial, have mosquitoes as agents, have a clear description of their attributes and is the most recent model made by that developer and the most recent model of that type. The spatial aspect and mosquitoes as agents were criteria as they are the focus of this study. The description of attributes was relevant as it gives insight to what values were used and how they could be implemented for this research's model. Finally, there are several branches of models that build on one another as shown in figure 1. In order to get an understanding of the spectrum of ideas, different models have to be analysed. This was done by analysing different branches. Of course, if a publication does not follow one of the other criteria, the model would still not be included for this analysis. The most recent model by that developer was chosen as that would give the most insight on their experiences.

In the end, eight different agent-based models were chosen with this criteria from the Smith et al. [17] publication which reviews all existing agent-based models on malaria. Table 1 gives an overview of the different existing malaria ABMs, what they specialize in and their pros and cons. As many developers build on already existing models, the first article of that model was used as a reference as well. Despite their different purposes, their data on mosquitoes, malaria and blood meals as well as their take on inputs and data collection were insightful on how this model should be built.

Table 1: Model developers with the aim, pros and cons of their model.

Developer	**Focus on the impact of:**	**Pros**	**Cons**
Gu and Novak [7]	Mosquito habitat destruction	Clear goal and clear in input variable use	Hypothetical and with many assumptions for variables not specifically for the goal.
Bomblies et al. [8, 12] (Using HYDREMATS)	Climate and environment variability	Detailed in environmental and climate factors	Complex and requires many inputs
Pizzitutti et al.[9]	Human movement and risk areas	Analysed the impact of human movement and human demography	Looks into *Anopheles Darlingi* instead of *Anopheles Gambiae*
Jindal and Rao [10]	ABM to test detection range and modelling multiple vector diseases	Includes the SEIR cycle and has the model publicly available	Looks into *Aedes Aegypti* instead of *Anopheles Gambiae*
Linard et al. [18]	ABM to establish biting rate	Formulaic basis for most mosquito input variables	Looks into *Anopheles Hyrcanus* instead of *Anopheles Gambiae*
Arifin et al. [19, 20]	Vector control interventions (VCI)	Logical simplification of mosquito development and differentiation of mortality rates	Does not consider spatial properties
Gerardin et al. [21, 22] (Using EMOD)	Reactive Case Detection (RCD), which is a VCI	Detailed formulaic input variableization and has their model publicly available	Mostly uncertain and site-specific input variables. Requires many inputs
Zhu et al. [23, 24]	Insecticide Treated Nets (ITN) and sugar bait	Lays out clear differences and impacts for different VCI's	Small study area and hypothetical

The interview with the entomologist and other entomological articles will be used to fact check decisions made by the agent-based model developers and discuss the realism behind their decisions. This for the most part takes place when developers are unclear on what the processes are in their model and how they function, or when developers disagree on an attribute value or a process.

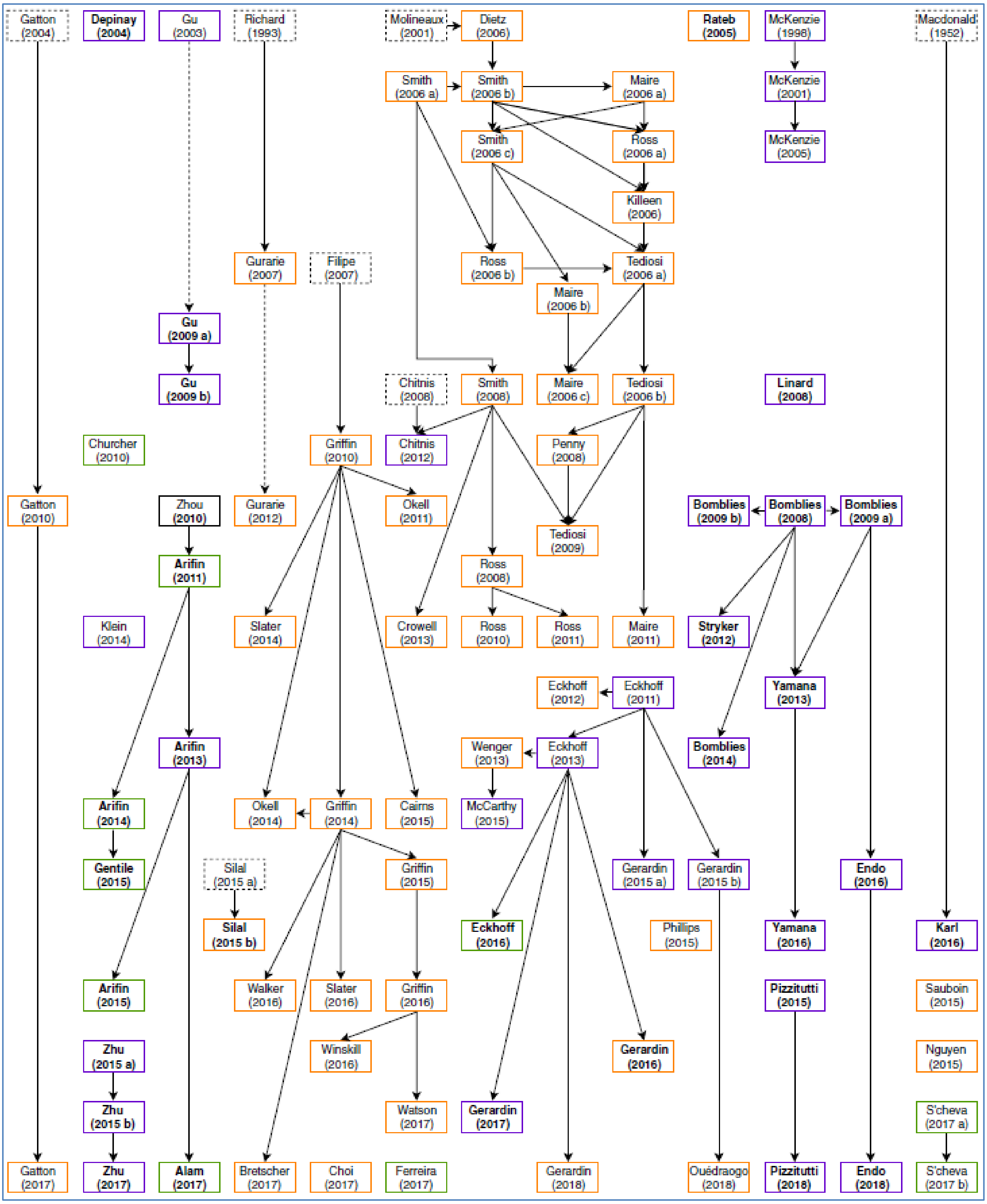

Figure 1: Different branches of malaria ABM development. [17] Dotted arrows mean there is a developer in common, yet have a different methodology. Black solid or dotted boxes means that they did not succeed their criteria or are not ABM's respectively and laid the ground for other ABM developers. Names that are not bold are not explicitly spatial. Boxes with orange, green or purple colour indicate respectively the use of human agents, mosquito agents or both in their model respectively.

Creating a new agent-based model requires an understanding of the existing attributes and the relevancy of these. Annex 2 shows all attributes used by the agent-based model developers. The values, formulas and conclusions discussed within this chapter are stated in this annex. Other mosquitoes than the Anopheles Gambiae are shown as well as a comparison, but it will not be used to conclude on the attributes used for this model.

The attributes in this research are divided into the adolescent mosquitoes, adult mosquitoes, malaria, prevention methods, environment, weather and model input variables. The first three could be considered universal for different areas and thus for these input variables' values are established if they were presented in their model. The prevention methods, environment and weather conditions were investigated to determine what input variables were often looked into and how they were applied into the model. Finally, the model input variables are there to give insight on what the foundations of the model.

Figure 2 shows a flowchart of the mosquito adolescent and adult states. Once the gravid mosquito lays eggs in an oviposition, those eggs will stay still until they are grown up to their adult states. The mosquitoes that are not yet mature will be referred to as adolescent mosquitoes in this research. The gravid phase will take place in water, but mosquitoes in any other phase can stay wherever, preferably out of the wind [25]. Every day the mosquito has a chance they will die, no matter what state the mosquito is in. This chance differentiates for the mosquito phases, but the differences do not go over 1% [20, 26].

For countries with high malaria prevalence, temperatures between 20-30 ºC are most common. This temperature range will be used for comparison in the rest of this chapter [2, 27]. The model itself will be unaffected by this decision.

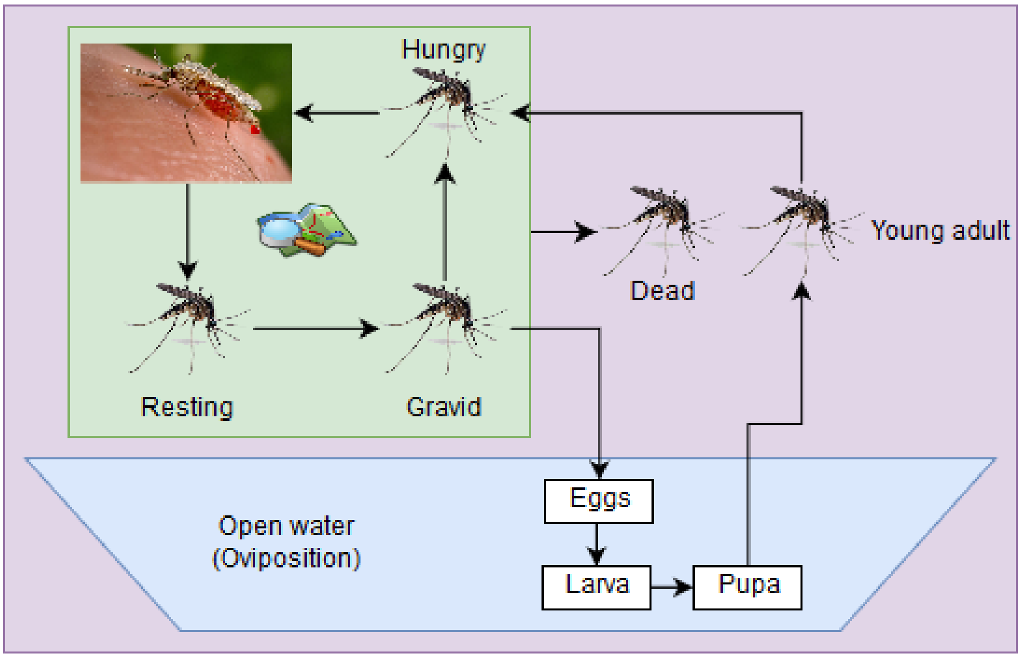

Figure 2: Model schematic for adolescent mosquitoes (purple and blue background) and adult mosquitoes (green background). The map with the magnifying glass symbolizes that these stats allow for spatial interaction.

2.1. Adolescent mosquitoes

The most important attribute for adolescent mosquitoes is how quickly they develop. The development of the mosquito is different for every model reviewed, but still share a general trend. Most of these follow a similar trajectory that is set up by Depinay et al. [28] who measured this in the field. The trajectory in figure 3 shows that for the total development the number of required days has an inverse relation with the temperature, with the exception of Gu and Novak [7] and Zhu et al. [23] who gave this a constant. The researchers agree that most of the adolescent development takes place in the larvae phase, which also has the most variety; the egg and pupa phases have much less variety, namely between 1-2 days with a similar inverse relationship. The total development time would usually be 10-20 days for countries with a 20-30 ºC temperature range, which holds up for most articles [29-32]. The method presented by

Arifin et al. [19] lines up the most with the Depinay's measurements whilst still being simplified into three formulas using only an additive value and a multiplicated value for each phase (equations 1 and 2).

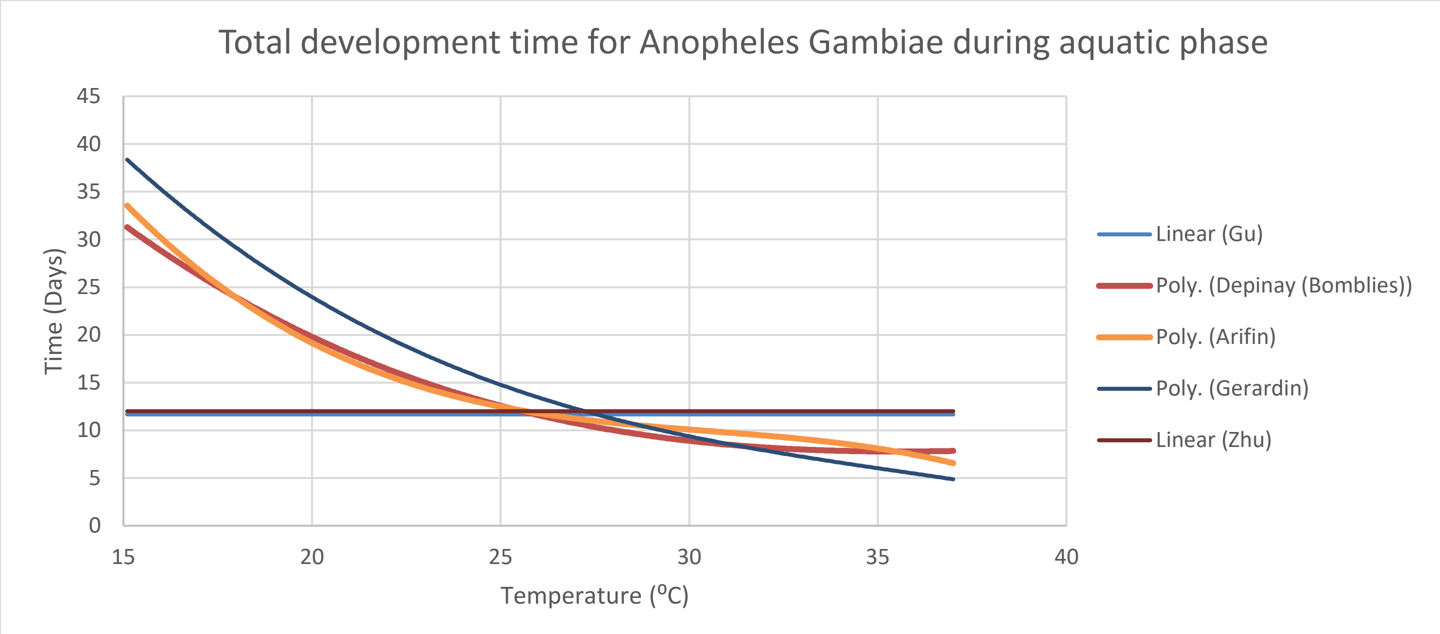

Figure 3: Trends of aquatic development given by different articles. The actual points are not given in this graphic to avoid a dense point cloud.

Once the mosquito has matured, they need time before they will start collecting blood. This time is more commonly agreed upon by the different developers, where the consensus is 2-3 days [28, 33]. This again is included in the simple formula designed by Arifin (equation 3). For this research, this stage of life will be referred to as young adult. These equations for development are as follows:

$$d_{ep} = 1 / (-0.9*T + 61) \qquad (1)$$

$$d_l = T*0.000305 - 0.003285 \qquad (2)$$

$$d_y = 1 / (-2.67 * T + 120) \qquad (3)$$

In these equations the d_{ep}, d_l and d_y stand for the development per hour of the egg, pupa, larva and young adult respectively. To reach full development the total growth across hours need to add up to 1. T in this equation as well as equation 4 is temperature in ºC.

Table 2: Description and summary for the adolescent mosquito input variables.

Input variable	**Description**	**Summary from existing models**
Development as an egg	Time it takes for an egg to become larva	Often set in relation to temperature. Around 1-2 days.
Development as a larva	Time it takes for a larva to become pupa	Often set in relation to temperature. Around 7-30 days.
Development as a pupa	Time it takes for a pupa to become a young adult	Often set in relation to temperature. Around 1-2 days.
Time before first blood meal (young adult phase)	Time it takes for a young adult to be able to feed	Often set in relation to temperature. Around 2-3 days.
Mortality mosquito	Chance for death per time unit or set days to live	Around 10% chance per day, living around 10 days with a maximum of 30. Sometimes in relation to temperature.
Eggs per oviposition (fecundity)	Eggs laid at once by a gravid mosquito	80-100 female eggs
Reproduction success rate	Chance to find mate and reproduce	Often excluded
Includes male mosquitoes?	Male mosquitoes cannot spread malaria	Often excluded

Every day the mosquito has a chance to die which is usually given as a probability. Most researchers set this to a default 10% number or used a formula that is around this 10% [34-37]. Some also give a maximum survivability given of 30 days to avoid overstaying mosquitoes [38]. This is universal amongst all Anopheles species and the equation is as follows:

$$M = 1 - exp^{-1 / (-4.4 + 1.31 * T - 0.03 * T^2)} \qquad (4)\ [35]$$

In which M stands for chance for mortality chance per day and T the temperature in °C. Between 15-30 °C this gives about 10-11% mortality chance per day, and after which increasing rapidly to a mortality of 16% at 10 °C and 34 °C.

Each time a mosquito lays eggs, she will lay about 160-200 eggs, half of which are female. This seems to be the general consensus amongst the articles [39-42]. In Jindal and Rao's article [10] it was stated that there was a chance that reproduction might fail, but this is stated nowhere else. Male mosquitoes are often not included into the models as they cannot spread malaria and seem to not have significant influence on the behaviour of female mosquitoes [19]. Table 2 shows the descriptions and summaries for the different input variables once more in a summarized overview.

2.2. Adult mosquitoes

Adult mosquitoes are much more active than adolescents. They move, feed and lay eggs. For this the following attributes are relevant: Active period, resting time, flight speed, flight distance, detection range, maximum meals per day and the chance for a successful bite.

Anopheles mosquitoes are nocturnal and rest during the day. On this all developers agreed and the hours picked were around 6 pm until 6 am [43-45]. Not all articles include a resting phase for the mosquitoes, but those that do highly disagree with one another and thus a conclusion cannot be made from the developers [46-48]. After a discussion with an entomologist it was established that the resting is a phase that takes one day [25].

Despite the maximum flight range being around 2-3 kilometre per day, most Anopheles mosquitoes fly about 200-400 meter a day [49-55]. This is agreed upon by most, although Gerardin et al. [21] also developed a Gaussian distribution to include the extremes. They seem to imply a mean of about 200 meter with a standard deviation of 120 meter.

Table 3: Description and summary for the adult mosquito input variables.

Input variable	**Description**	**Summary from existing models**
Active period	Time in which a mosquito looks for oviposition or blood meal	As Anopheles Gambiae is a nocturnal insect, consensus is around 6pm to 6am.
Resting	Time a mosquito rests after feeding	Significant disagreement ranging from hours to days
Flight speed	How much distance per time a mosquito can take	Although it is inconsistent, it does always line up with the flight distance
Flight distance	The distance a mosquito can fly in a day	Anopheles Gambiae usually flies 200-400m a day, with a maximum of 3km.
Detection range (other)	Distance to which a mosquito recognises an oviposition or blood meal	For most it can be assumed that interaction only takes place when they're in the same cell. Else it's about 15m.
Detection range (CO2)	Distance to which a mosquito recognizes a blood meal	For most it can be assumed that interaction only takes place when they're in the same cell. Else it's about 15m.
Max meals per day	The amount of times a mosquito will feed	Most state it to be 1. Mosquitoes often rest after their blood meal which at least takes a few hours.
Successful bite	Chance a bite will succeed	Some variation, but all values are assumed and none are referenced.

The detection range can be either by visuals, by smell or by recognizing humidity; mosquitoes recognize blood meals by their exhalation of CO_2. Although most imply that mosquitoes will sting their target once they are within the same cell, this would cause inaccuracies once the cell size increases towards 50 meters or higher. Otherwise it is stated that the range would be about 15m. Some hypothetical models state that this would be 40 meters for recognizing blood meals [7, 24], but is not represented by models that analyse real locations. Until the mosquitoes recognize either an oviposition or a blood meal, they will fly randomly around shelter. This allows them to stay out of the wind [49, 56-58].

Table 3 shows the descriptions and summaries for the different input variables once more in a summarized overview. This includes the max meals per day and successful bite which was included in some of the models. Although for most it was concluded that the max meals per day would be 1 and the chance for a successful bite is always 100%. Mortality of the adult mosquitoes is the same for adolescent mosquitoes.

2.3. Malaria

The malaria input variables include transmission, incubation and the SEIR cycle. Malaria seems to be spread with a 50% chance from mosquito to human and about 15-20% from human to mosquito [59-61]. There is disagreement on the human to mosquito transmission values, but this difference is not severe [59, 60, 62, 63].

The external and internal incubation determine the amount of time needed for the parasite to develop in order to show the symptoms and to become infectious. The external in this case refers to the Anopheles mosquito and internal to the bite victim. For the mosquitoes there seems to be an agreement between Bomblies et al. [8] and Pizzitutti et al. [9] on 111-degree days above 18 ºC; which is 27.8-7.9 days for temperatures between 20-30 ºC. Arrhenius equation [22] on the other hand brings a range of 9.4-24.1 days for the same temperatures. Either fall within a similar range and could therefore both be true [40, 62, 64-66]. Malaria is not transmitted via birth and can only be obtained through mosquito bites for the mosquitoes [25].

The internal incubation time is only further developed by Gerardin et al. [21], as Pizzitutti et al. [9] only uses a simplified range for another Anopheles species and the others assume it has a default value [65]. For this only the Arrhenius method will be used which is the following equation:

$$t_{ii} = 1.17*10^{11}\, e^{(-8.4\ *\ 10\wedge3/Tk)} \qquad (5)$$

In which t_{ii} stands for internal incubation in days and Tk for temperature in Kelvin, resulting in the same range as it would be for the external incubation according to Eckhoff [22] who developed EMOD.

Table 4: Description, summary and conclusions for the blood meal input variables.

Input variable	**Description**	**Summary from existing models**
Transmission (mosquito to human)	Chance that an infected mosquito can infect a human	Both referenced values agree on it being 50 percent chance of success
Transmission (human to mosquito)	Chance that an infected human can infect a mosquito	The only Anopheles Gambiae models that referenced this number gave 15 and 20 percent chance.
External incubation	Time the disease develops inside a mosquito before being infectious	Often set in relation to temperature. It takes around 10 days for every method.
Internal incubation	Time the disease develops inside a human before being infectious	It takes around 9-14 days, likely with a similar relation as stated with the Arrhenius method
Susceptible	A blood meal is susceptible to being infected	Most models state that any human can become infected at any point in their life
Exposed	A blood meal is infected but not yet infectious	As this is lining up with the internal incubation, most assume this to be the same
Infectious	A blood meal can now infect mosquitoes	There is disagreement on this part, except that it stops when recovery has happened
Recovered	The blood meal is no longer infectious	The recovery phase seems to progress with about 1-1.2% per day
Includes cattle and/or domestic animals?	Whether the model includes alternative blood meals or not	Not many include cattle and those who do state they have little influence unless at a high density

The susceptible-exposed-infected-recovered (SEIR) method seems to be only implemented by Jindal and Rao [10]. This method states that once a blood meal has infected, it cannot be infected another time after recovery [67]; for that the SEIRS method exists in which they return to susceptible state. Other developers state that agents can be infected again after recovery, which lines up with WHO and CDC statistics. Exposed state is equal to the time it takes for the disease to develop, and thus is equalized to the incubation time. The infectious phase seems to be more unclear as there are only two articles are discussing this and they disagree with another. Therefore, it is assumed it will last until they are recovered. Researchers agree that this recovery would take something between 83-100 days [37, 68, 69].

Cattle can affect malaria spread as they can act as alternative blood meals. Anopheles mosquitoes most of all prefer human blood meals, but with high cattle density Anopheles mosquitoes might still choose for cattle [12]. There is no clear relation established between the chance humans get infected and the abundance of cattle. Therefore, a relation is developed in the methods chapter (3.5.3.) as an adult mosquito process.

Table 4 shows the descriptions and summaries for the different input variables once more in a summarized overview.

2.4. Malaria prevention methods

The most commonly included preventive method are the Insecticide Treated Nets (ITN), Indoor Residual Spraying (IRS) and Larval Source Management (LSM). The ITN are protective nets that are usually around beds, but can be used on windows and doors as well. Because of this, it can be assumed that when ITN's are in place, humans are only protected while at home [70-72]. IRS is the spraying of pesticides inside for the most part, which protects the human as long as they are inside [71, 73]. LSM is the destruction of mosquito habitat around human habitats to reduce the chance of mosquitoes reaching the humans. This method destroys any water bodies around human habitats and the studies that implement this usually focus on this method [7].

The models that include preventive methods predominantly use it as a testing input variable. For most of these models it is assumed that blood meals are safe from mosquitoes with these preventive methods installed [70-73].

2.5. Environment

The spatial information for the areas is most of all collected using multispectral imagery [8-10, 18]. The categorization is likely done using vegetation indices to classify different types of land [74]. Most articles define a maximum amount of eggs per oviposition, although there is disagreement on the value. Bomblies' method is the only one with source, in which they gave weights to the different phases of development for each mosquito [28]. These weights themselves are without source however, making Bomblies' method more unreliable than it seems at first glance.

2.6. Weather

Temperature is key to the development calculations and is therefore included by most developers. Not many state in what frequency they have this data, but it can be expected that most have this data available daily as most weather stations do not share their history in hours on public sources [75-77].

Precipitation is important for determining the number of ovipositions within an area and the increase of adolescent mosquito mortality. Although there is potentially increased mortality due to precipitation, it is highly inconsistent in reality and is not accurately defined by developers [78]. The creation of new ovipositions is only relevant if the water stays until the adolescent mosquitoes are fully developed [12]. Therefore, the water decreasing factors need to be taken into account, which are the infiltration rate and the evaporation rate. Both of these rates are complex to determine. The infiltration rate is influenced by the soil type, the ground water level, the amount of pressure the land receives and other soil characteristics [79]. The evaporation rate is determined by the humidity, temperature, wind speed and the area of water bodies [80]. Therefore, to determine the current water state, the current water on the surface due to rainfall can be subtracted with the infiltration rate and evaporation rate.

Almost none of the developers put wind into their model with the exception of Bomblies et al. [12] who used it in order to calculate the detection range. The wind also affects the evaporation rate and could potentially impact flight patterns. As such, winds that are greater than 3 km/h can significantly reduce host seeking flights. This is why mosquitoes are more often found in forests or households as the trees and walls can act as wind barriers [81]. Including wind into a model is complex however, as it is both has a temporal and a spatial component.

3. Methods

Four research questions were created in order to answer the main objective. The questions were about the used attributes by other developers, the required data, the processes behind malaria spread and on validating the model.

Before any conclusions were made on what attributes, inputs and processes were necessary to include, modelling principles were developed. These principles set the standard on how the model can reach its goal. After the principles were established, decisions were made on what attributes, inputs and processes were included. Finally, it was determined how these were implemented into the model. This chapter goes through the different steps and describes the decisions that were made.

This chapter mainly deals with the conceptual and formal design of the model. The technical description can be found in the ODD (annex 1).

3.1. Modelling principles

Before the attributes, inputs and processes are determined, it is important to reassure the model sticks to its goal. This is done by setting basic principles for this model. The main criteria for own developed parts are that:

1. It would not require inputs that were not generally available;
2. The required inputs are adaptable to different available data;
3. Only processes and attributes that are likely accurate or are highly relevant are included in the final model;
4. All assumed values with high relevancy will be determined using calibration;
5. All processes that can be spatially determined and that do not contradict the previous development principles, has been put into the model as a spatial component.

The first principle means that the necessary data should be accessible online or can be determined by own means. For example, data might be available on where people live and whom is protected.

The second principle ties in closely to the first principle, but this describes the input contents. As such the evaporation rate could have been calculated in the model based on wind speed, humidity and temperature. However, as wind is a complex input, the evaporation rate simplified to a number per day. If the data would be available, this could still be put into the input value of evaporation, but will not be processed by the model itself.

The third principle reassures no attributes or processes are made up, unless a likely estimate can be made or if it has a high impact on malaria. Other than speeding up the model, this also reassures that everything inside the model is built on known processes and interactions. For example, it is known that mosquitoes remember their blood meals up to three days, which has an impact on who they will feed on next time they are hungry. However, what impact it has and how it affects mosquito behaviour is not clearly defined [82]. Therefore, this process is not included into the final model.

The third principle can also be applied for attributes such as the detection range, which can be determined using either a complex formula or an estimation. As the complex formula requires inputs that cannot be considered generally available, an estimation had to be made on what this value is. This therefore also ties in with the fourth principle which states that relevant, yet assumed attributes have to be calibrated to the area unless a trend can be established.

The final principle reassures that this model stays spatial as was intended from the goal.

3.2. Identifying processes, attributes and inputs

Before the model could be built, the processes had to be established in order to determine what attributes and inputs were necessary. Sometimes estimates had to be made to what processes would be relevant, as this cannot always be determined beforehand. A process had to be either accurate or relevant to be included to the model. The processes that had an impact on spatial interactions between agents or their environment were in general considered highly relevant. Accuracy in this case refers to the number of developers agreeing on the process implementation.

This subchapter goes further into each type of agent and identifies their processes. The known attributes and processes also help to determine what area dependent attributes need to be used as an input for the model.

Attributes such as the initial percentage infected and the number of each agent will not be discussed in this subchapter as these are considered default attributes.

3.2.1. Adolescent mosquitoes

Adolescent mosquitoes are a simplistic agent set as they do not leave their water body until they grow up. While an adolescent mosquito is growing up, they will go through two processes: mortality rate and the development to their next phase of life. These processes are shown in figure 4.

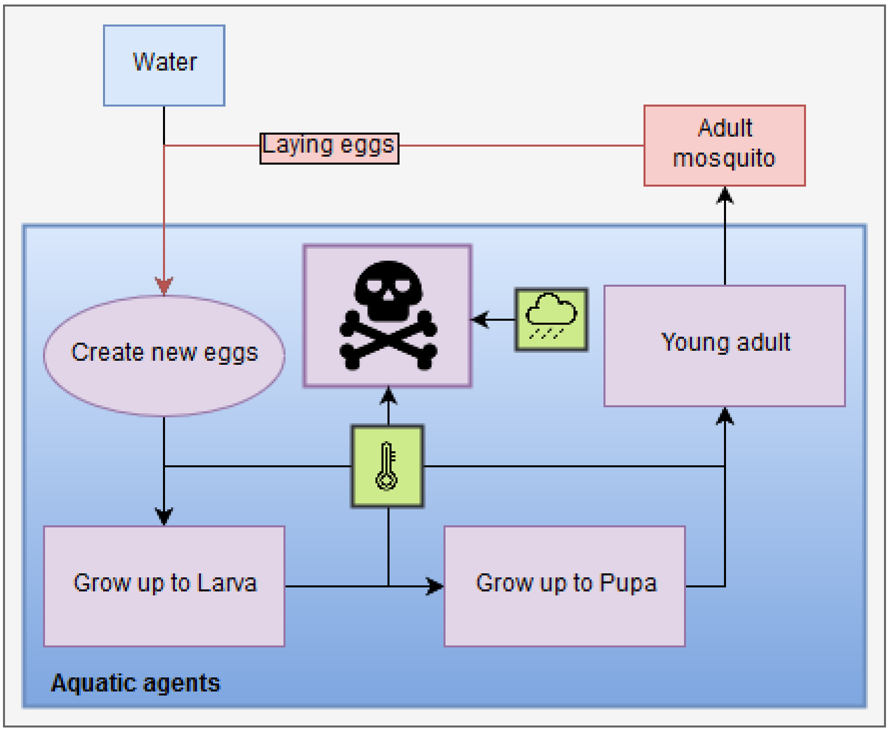

Figure 4: Adolescent mosquito processes and attributes.

The mortality rate is dependent on two attributes: the amount of rainfall and the temperature. While it is raining, droplets may kill the adolescent mosquitoes. This is due to the fact that eggs are usually on the water surface, making them vulnerable to the impact of droplets [9]. There are various animals who feed on mosquito eggs [83], but this is often combined with the temperature into a single input variable. Predation will therefore not be included separately. Finally, the mosquitoes are sensitive to temperature, thus the more it deviates from their preferred 25 ^{o}C the higher their mortality.

All developers agree that the process of development is only a temperature dependent process. This was even in a similar range as was confirmed in figure 3. Despite there being potentially other influential input variables, temperature is considered the most significant one as mosquitoes are poikilothermic [28].

Once the young adult has grown up, they will start off as a hungry mosquito and the adult mosquito processes will start taking place. At some point the adult mosquito will lay eggs in the water and thus the cycle repeats. The laying of eggs is considered an action of a gravid mosquito and will therefore be handled in the adult mosquito processes.

From these processes it can be concluded that the temperature, rainfall and predation rate are necessary in order to cover the mortality and development of adolescent mosquitoes. These attributes are considered spatially dependent and are therefore an input to this model. Temperature and rainfall are temporal, but regionally homogenous. This is why they are considered temporal inputs.

3.2.2. Adult mosquitoes

Adult mosquitoes are the most complex agent set of this model as they have the most interactivity with adolescent mosquitoes, humans and their own environment. As such they have movement, cycling of phases, feeding, laying eggs, transmission processes, external incubation and mortality rate. In figure 5 an overview is given of the adult mosquito processes.

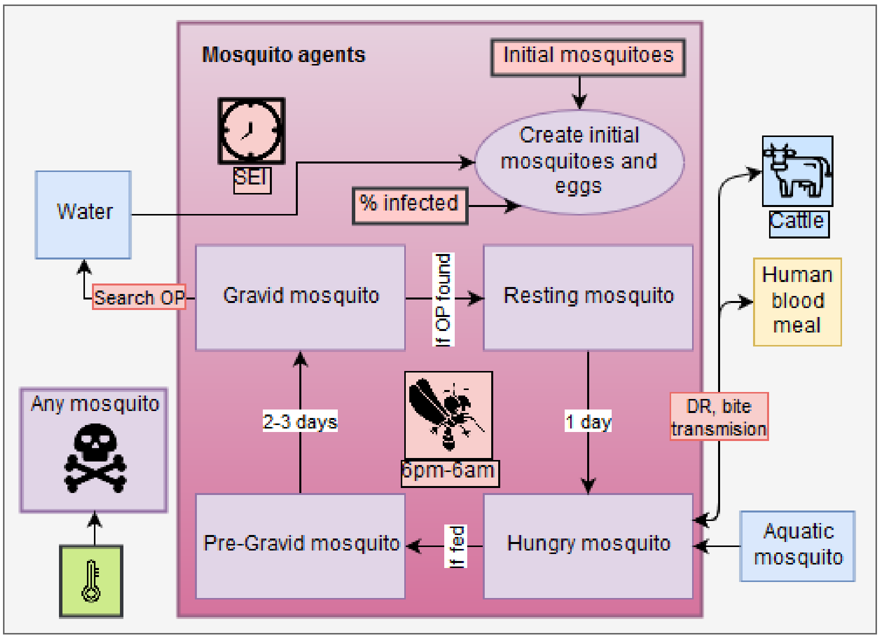

Figure 5: Adult mosquito processes. DR stands for Detection Range, SEI for Susceptible - Exposed – Infectious, the clock represents SEI being a general temporal process and OP for oviposition.

Movement of the mosquito is tied to many factors as it matters what time it is, what phase the mosquito is in, where the mosquito is located and what their speed is. Anopheles mosquitoes are active during the night and therefore will fly between dusk and dawn. Mosquitoes that are pre-gravid or resting will stay still, but the rest will move in this time. The mosquitoes that do fly will do their best to stay near vegetation or buildings as the mosquitoes prefer to stay out of the wind. Hungry mosquitoes will also stay close to humans as they are the preferred blood meal. They will stay within detection range of them during flight to make sure that they do not lose their preferred locations or targets.

Even if the mosquito is not resting, the phase of the mosquito is relevant for its movement. A gravid mosquito will search for an oviposition and a hungry mosquito will search for a blood meal. This phase transition for gravid mosquitoes and hungry mosquitoes is task dependent. Once a mosquito has fed or has laid eggs, their task for that phase is completed and they continue to their next phase. The resting and pre-gravid phases are time based and will continue once that time has passed. This means that this process does not require any attributes other than time to function, but does rely on other processes.

A grown-up adolescent mosquito starts off in the hungry phase. Here they will check any location within their detection range to find humans. Although humans are their preferred blood meals, there is a chance that they will feed on cattle instead if there are any in their neighbourhood. There is also a chance their attack fails when there are malaria preventive methods installed in that building. This may cause the mosquito to die. If a hungry mosquito has been able to find a target to bite successfully, the transmission process will take place.

Once an adult mosquito is in the gravid phase, they will look for an oviposition to lay eggs in. They will check any locations within their detection range to see whether there is water available to lay eggs in. This process does not really have a criterion, but rather a preference. Gravid mosquitoes will choose temporary pools of water over permanent ones. This as permanent pools of water have more often predators that may feed on the adolescent mosquitoes.

Transmission of malaria between mosquito and human are only dependent on whether one of the agents is infected and the other one is not. If both or neither are infected, no transmission can take place and therefore this process will be skipped.

Once a mosquito is infected with malaria, they will go through incubation until they are infectious. The external incubation time of malaria is solely dependent on temperature. After this time has passed the mosquito will stay infectious until they die.

Adult mosquito mortality functions similarly to adolescent mosquito mortality, except that they are not threatened by rain. Temperature here plays the same role and the input variable also includes potential predation.

The adult mosquito processes require environmental, but also general attributes. Temperature, available egg space, predation rate and locations of buildings and vegetation are environmentally dependent data. Flying speed, detection range, fecundity, incubation time and chances for transmission can be expected to be the same for the entire Anopheles Gambiae species. Therefore, standard values can be given to the general attributes.

Available egg space is the only environmental input that can be considered a mosquito input variable. The adolescent mosquitoes have standard weights and a maximum weight per square meter. For this, standard values can be found as well in combination with the area of the water body.

3.2.3. Humans

Human agents are simplified in most models with only movement between home and work, the implementation of malaria preventive methods and the SEIRS cycle. Figure 6 shows the processes the human agents follow.

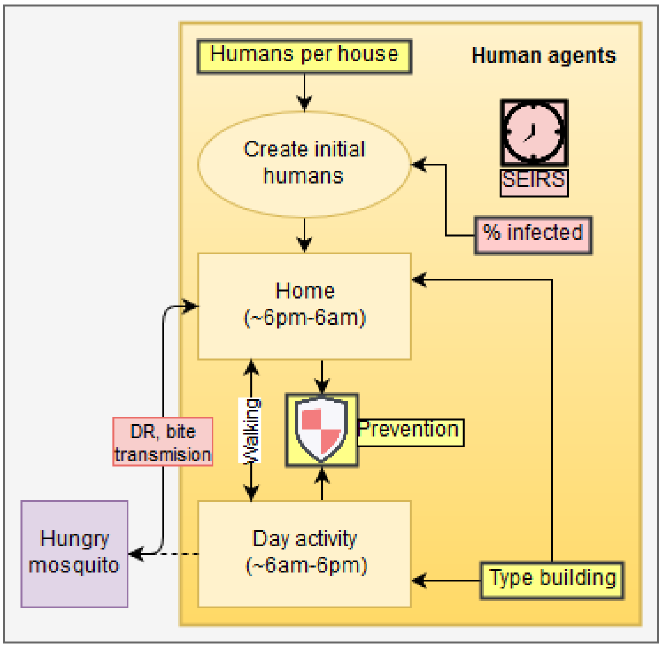

Figure 6: All human agent processes used for this model. DR stands for detection range and the clock represents SEIRS being a general temporal process.

Just like mosquitoes, human movement depends on their speed, their location, their starting time and their end time. The time that they start and stop walking depends on the society they live in; some cultures have a lot of late-night activities whereas others sleep early.

Malaria preventive methods reassure that humans inside buildings with protection cannot obtain malaria and have a chance of killing attacking mosquitoes. Although this is not directly a human method, the human may decide to leave their protection at random for any reason given. Example of reasons may include bed bugs, overheating or night activities [25]. This also depends on the area, the society and even the person. For this reason, an estimate has to be given to this attribute.

A human can obtain malaria during the day despite Anopheles mosquitoes being nocturnal. This is only when the humans work in risk areas such as farms, where much surface water is present on a regular basis. This makes the humans so close to the mosquitoes that they are able to sting them during the day. Preventive methods are also never available in these risk areas.

Once a human does obtain malaria, they will go through the SEIRS cycle. The human will only be able to obtain malaria in a susceptible state and the cycle will start continuously from that point forward. Malaria cannot be transmitted to a mosquito until the human reaches the infectious state, which is obtained after incubation. Unlike mosquitoes, humans have a chance of recovering from malaria. Once they have recovered, they are once again susceptible. The exposed state and infectious state have input variables attached to them that indicate how long it takes before they reach the next state. The exposed state is dependent on the temperature as it works highly similar as the external incubation. The infectious state is more dependent on how quickly a human can recover from malaria and this seems to be given an attribute rather than a input variable.

From these processes, it can be concluded that the human speed, walking times, temperature, recovery rate, type of building and chance people will leave their protective states are the necessary attributes to compute these processes. Most of these are environmental inputs, with the exception of human speed and the recovery rate which are expected to be area independent.

3.2.4. Environmental process

The environment has only one process in this model, namely the water cycle. It is not expected that the location or the values of other environmental aspects will change within a few years, which is the most common duration of a model. Aspects that might change within that time, such as the construction of new buildings or the digging of large holes, are considered highly unpredictable. Only when these changes are known beforehand, these can be considered during analysis.

To determine the water cycle, it is required to know the amount of rain per day, the evaporation rate and the infiltration rate. Rain per day will of course determine how much water will hit the surface, while the evaporation rate and the infiltration rate determine how much decreases per day. The evaporation rate and the infiltration rate can both be calculated, but sometimes these are also documented. For infiltration rates estimates can be made based on land classes and soil types. The reason why these values are not calculated by default, as most other processes are, is because the calculations may require data from field testing, which is against the first modelling principle of this research.

Because of this, the rain per day, evaporation rate and infiltration rate will be used as direct inputs. The rates may be filled in depending on what data there is available.

3.3. Attribute values

In the literature research, common malaria ABM attributes and input variables were discussed. Following the principles of 3.1., the number of attributes necessary to complete these processes have been reduced and now conclusions can be made for the values.

There is a clear distinction between attributes that are consistent at all times, input variables and area dependent attributes. The static attributes are expected to be the same for Anopheles Gambiae and Plasmodium Falciparum regardless of a different space or time. Area dependent attributes can be considered the least consistent, as they are dependent on the local culture. An assessment has to be made on how these values play out within that area, which may especially be difficult to do without field observations.

3.3.1. Static attributes

As established, there were a few static attributes that can be determined from the identified processes, namely: Adolescent mosquito weights, fecundity, active period, flight distance, detection range, malaria transmission chances, recovery time and human movement speed. Table 6 shows the determined values for these attributes.

Adolescent mosquito weights were difficult to establish as the best way to limit the amount of adolescent mosquito per square meter had to be determined. It was decided to go with Bomblies et al. [12] method as their maximum eggs per water body was referenced and their method was reproducible. The other models either set arbitrary values as maximums or were unclear in their implementation. The weights of the adolescent mosquitoes are not referenced however, and are estimates. This makes sense as the weight of a mosquito varies significantly between each individual [84]. It is not expected that this potential variance will have a significant impact and the actual ranges of weights are unknown. Therefore, these values were chosen.

Adolescent mosquitoes also have an increased mortality during the rain, which differs for each mosquito phase [85]. This is generally consistent throughout different intensities of rain.

Table 5: Attributes used in the model that are expected to be consistent regardless of time and place. Yellow is used to identify human attributes, purple for adult mosquitoes and blue for adolescent mosquitoes.

Attribute	Value	Unit	Source
Egg weight	0.02	mg	[12]
Larva weight	0.16	mg	[12]
Pupa weight	0.30	mg	[12]
Young adult mosquito weight	0.45	mg	[12]
Impact mortality rain for egg	17.5	% increased	[85]
Impact mortality rain for larva	13	% increased	[85]
Impact mortality rain for pupa	9	% increased	[85]
Impact mortality rain for young adult	4.8	% increased	[85]
Maximum eggs per water body	300	mg / m^2	[12, 28]
Eggs per oviposition (fecundity)	80-100	Female eggs	[39, 41, 42]
Active period	18:00-06:00	h	[45]
Flight distance	200-400	m/day	[50, 55]
Detection range	15	m	[25, 56]
Transmission (mosquito to human)	50	% chance / bite	[60, 61]
Transmission (human to mosquito)	17.5	% chance / bite	[60, 62]
Human movement speed	1.4	m/s	[86]
Time until recovered	0.04167-0.05	% development / h	[37, 68]

Fecundity and flight distance were sometimes defined as a normal distribution, yet most developers kept these values as ranges. As such these attributes are kept as linear ranges within this model.

Detection range was difficult to establish. Even though it is an attribute that can be considered important, there are significant differences between the point of views of the developers. After further research it has been decided 15m would be most realistic [25, 56]. This is still an approximation and therefore has to be tested further before this can truly be defined. This attribute is considered highly relevant as it directly relates to the interaction between mosquitoes, other agents and their environment.

Human to mosquito transmission had a small disagreement on whether it would be 15% or 20% chance. Although this could have been added as a range like the flight distance and the fecundity, it was decided to take an average for this. This was because the decision was made to represent more than one mosquito with a single agent to speed up the model. Changing this also made sure that the transmission values had to be resolved mathematically to realistically represent what the chances of a swarm of mosquitoes would have as opposed to a single mosquito. Because of this, the average was taken to prevent recalculation each time a mosquito had the chance of obtaining malaria. Just like the detection range, the human to mosquito transmission could use further testing to reassure the right value has been given to this attribute.

Human movement is consistent for different cultures and body types. Although developers did not discuss this attribute, it was quickly found to be 1.4 m/s [86]. This does not differ at any point throughout the model.

Recovery time was given as 1-1.2% development per day, which would be the same as 0.04167-0.05% per hour. This was given as two different values and unlike the transmission, from these two values a range was made. This is because the difference is insignificant, ranging between 83-100 days. It is unlikely that the 17 days difference will have a large impact, especially because there is a chance to be stung and a chance for transmission. This is used as a linear range as there is no defined standard deviation.

Because there are multiple ranges of values for the area and time independent attributes, there is a chance that this would result in inaccurate results. Therefore, sensitivity analysis will be performed on these attributes using the Monte Carlo method to determine whether they have significant stochasticity.

3.3.2. Input variables

Input variables used in this model are only dependent on the temperature. This includes the adolescent mosquito development rate, mortality and incubation rates. These are shown in table 7.

Development of adolescent mosquitoes is calibrated by Arifin et al. based on the findings of Depinay et al. This works for most cases but will falter for temperatures below 11 degrees Celsius. Then the larva development will be brought to minus. As this is not a temperature in which Anopheles mosquitoes are prevalent, it is expected that this temperature will not occur in malaria areas.

Table 6: Input variables used in this model. T stands for temperature in Celsius. Tk for temperature in Kelvin.

Attribute	Value	Unit	Source
Development as an egg	1 / (-0.9*T + 61)	% development / hour	[29]
Development as a larva	T*0.000305–0.003285	% development / hour	[30]
Development as a pupa	1 / (-0.9*T + 61)	% development / hour	[29]
Development as a young adult	1 / (-2.67*T + 120)	% development / hour	[28]
Mortality mosquito	$1 - e^{-1 / (-4.4 + 1.31 * T - 0.03 * T^2)}$ and maximum 30 days	% chance/day	[35]
External incubation	$1.17 * 10^{11} * e^{(-8.4 * 10^3/Tk)}$	days	[64, 65]
Intrinsic incubation	$1.17 * 10^{11} * e^{(-8.4 * 10^3/Tk)}$	days	[65]

The mortality of the mosquito is considered to be all encompassing. This means it would supposedly include the predation rate, influence of humidity and temperature. Although this is an assumed trend, it is most often used by other developers. A maximum of 30 days is given to the mortality of the mosquito as well. This is to prevent some mosquitoes to live indefinitely by chance, especially as malaria is not transmitted through birth for the mosquito.

Incubation of the Plasmodium Falciparum is the same for humans and mosquitoes. For this Arrhenius method will be used. The values here are fitted by Eckhoff and are similar to the degree days method used by Reeves. Degree days method is not used here as Eckhoff confirmed that his method can be applied for intrinsic incubation as well whereas Reeves did not.

3.3.3. Area dependent attributes

Some of the attributes depend on the area it is set in, but do not differ much regionally. These most of all depend on the culture of that area and human behaviour. This is represented in table 7. All values in this table are used for the hypothetical situation and have to be revaluated each time a different study area is used.

The number of adolescent and adult mosquitoes as well as the initial number of infected mosquitoes and humans are general indicators. These values are used as an input for an initial run of the model. Initial run means that the model will first run for an entire year and the result of that will be used as an input for the actual model. Using this method, the mosquitoes and malaria would start in a realistic state for that area. It would not randomize the starting positions of the mosquitoes and malaria each time it would run.

Table 7: Area dependant attributes. Values given for the hypothetical situation.

Attribute	Value	Unit
Number of adolescent mosquitoes	Two times the adult mosquitoes	-
Number of adult mosquitoes	10.000	-
Initial infected mosquitoes	10	%
Initial infected humans	30	%
Start of the day	6-10	Hour
End of the day	20-24	Hour
Chance of leaving protection	10	% / bite

The start and end of the day and the chance of leaving protection are attributes that last throughout the model. The start and end of day are about when a person would enter or leave their house's malaria preventive methods. Chance of leaving protection is an additional attribute that keeps this in mind, but this would be while the human should be safe within their protection. This is due do whether a human decides to leave their protection because of night activities or discomfort within their protection.

3.4. Input files

The attributes discussed can be put in the model itself either as default values. However, there are also attributes that differ depending on the space and time. These need to be loaded into the model before they can be interacted with. Although there are many ways to include the remaining data as inputs, it has been decided to do this via three files: a building shapefile, a spatial environment shapefile and a temporal environment csv.

3.4.1. Buildings shapefile

The building data includes the following attributes: number of people per building, what the people work/living places are and whether this building is protected from malaria yes or no. Although cattle are closely related to humans for the sake of being stung by mosquitoes, the cattle are expected to play more an environmental role.

The working and living places are important to determine where humans are during the night and during the day. Some of the work places can also fall under the category of being risky, such as farmland. These areas are given permanent surface water in the model itself to allow breeding to be done at the farm at all times. This in turn also means that mosquitoes will be more often near this environment.

The protection methods differ in approach. As such bed nets protect people during their sleep and indoor residual spraying protect people as long as they are inside. Both ITN and IRS are added as separate attributes.

3.4.2. Spatial environment shapefile

The spatial environment includes the infiltration rate, cattle to human ratio, the area and whether it is a forest or water body yes or no. The spatial environment is as big as the study area and is separated in sections that differ from one another. For example, if two areas are known to be the same for every attribute except for, they would still be separated into two different polygons. This environment can be made as detailed as desired.

The infiltration rate is important for determining how quickly water disappears from the surface. Although the infiltration can be considered complex as it depends on many factors, it has been simplified here to single value per polygon. This factor can keep in mind the soil texture and the height of the land. For example, if there is a significant amount of soil dug up from the ground, the infiltration rate is expected to decrease because of that.

The cattle to human ratio determines how much cattle is commonly in that area. This rate has a direct relationship to the chance that a human would be stung. This relationship will further be discussed in 3.5.3. The reason as to why this is part of the spatial environment and not the buildings is because cattle are not bound by buildings and can reduce human victims even when they are walking towards their destination. Although cattle do have day and night cycles like humans and mosquitoes, they are simplified here towards a single number for an area. This as cattle behaviour is more unpredictable than humans and less relevant than mosquito behaviour.

The area has to be included as an input variable as GAMA has difficulty recognizing area directly. This input variable is relevant for determining how many mosquitoes and eggs will be initially there and how many eggs can be laid maximum per polygon.

As mosquitoes prefer areas that are out of wind, they prefer to be in vegetated area or next to buildings. The building shapefiles are included already separately, but the forests have to be included in the spatial environment.

The water input variable is set as well to represent permanent water bodies. This is included separately as these water bodies have different properties. If a polygon is water, all other input variables associated with the polygon are put to 0.

3.4.3. Temporal environment csv

The temporal environment is for the most part weather, and therefore includes minimum temperature, maximum temperature, rainfall and evaporation rate on a daily base. This means hourly detail cannot be put in without changing the model.

The minimum and maximum temperature for each day is included to make sure there is a difference between night and day. These temperatures are important to determine the input variables discussed in 3.1. and therefore, knowing this is of high relevance.

The rainfall determines how much water reaches the surface. This water will stay on the surface until it has been fully depleted. While it is raining, it is known that evaporation does not take place as the humidity will be at 100% at that point. Infiltration rate is also reduced while it is raining as the added water to the ground water will go above the soil at some point. For these reasons, it is assumed that while it is raining, the depletion rate is 10% of what it would normally be.

Although wind is mentioned to be relevant for the residence of mosquitoes during the day, it is not directly included into the model. This as wind is both spatial and temporal, making it complex to implement this into the model without harming the model speed. Other than the residence of the mosquitoes, wind can be considered relevant for the evaporation rate and the detection range. The evaporation rate is therefore included as a separate attribute. Wind could still be used to calculate it if it is available, if not some measurements of evaporation rate in that region could be used.

Some models also include humidity to determine mortality or the evaporation rates. However, as the evaporation rates are already directly included and the equation developed by Martens [35] indirectly includes humidity, this value remains excluded.

3.4.4. Summary

Table 8 shows the inputs described in this subchapter summarized into a singular table. This table describes the data file, what it is for, the attributes inside, the type of the attribute and what the attribute is used for. Here it is important to note that the type of building is included into the model as an int instead of a string. This is to prevent misspelling and convenient randomization.

Table 8: Input data for the malaria model.

Input file	**With**	**For**
Buildings (Shapefile) (Human species)	Number of people (int)	Creating the right amount of people per building
	Type of building (int)	Determining whether it will be a location for the day or the night
	ITN protected (Boolean), IRS protected (Boolean)	Determining whether humans are safe while asleep or when inside the building
Spatial environment (Shapefile) (Land species)	Area in square meters (float)	Determining how much water can be created during precipitation
	Infiltration rate in mm/d (float)	Calculating the temporary stay of water from rain
	Cattle to human ratio (float)	Calculating the chance of the mosquito stinging a human instead
	Water body (Boolean)	Whether it is a water body by default. If so, all other values are 0 except for area.
	Forest (Boolean)	Whether it is an area where wind is being blocked.
Temporal environment (csv) (global values)	Day or max temperature in ^{0}C (float)	Calculating the input variables
	Night or min temperature in ^{0}C (float)	Calculating the input variables
	Evaporation rate in mm/day (float)	Calculating the temporary stay of water from rain
	Precipitation in mm (float)	Determining the amount of potentially create ovipositions from rainfall

Unlike the other input files, real weather data is used as the temporal environment in the hypothetical model. This is to represent a realistic climate. Giving random values might give significant inconsistencies in the temporal data and therefore it is always recommended to have a semblance of realism for the weather input, even when doing hypothetical runs.

As the number of humans, type of building and preventive methods are building dependent rather than based per person, these are considered a separate input from the spatial environment.

3.5. Process implementation

This subchapter will elaborate on the sub-models, the initialization and the development of the results. The model works hourly and the processes that are daily are performed at the 24th hour of the day. All processes are implemented in GAMA due to its advanced spatial integration.

Figure 7 shows a summary of the full model. This is based on the figures 4, 5 and 6 combined. The inputs and environmental processes are shown here as well. The environmental processes are limited to the water cycle as shown in the Open surface water box.

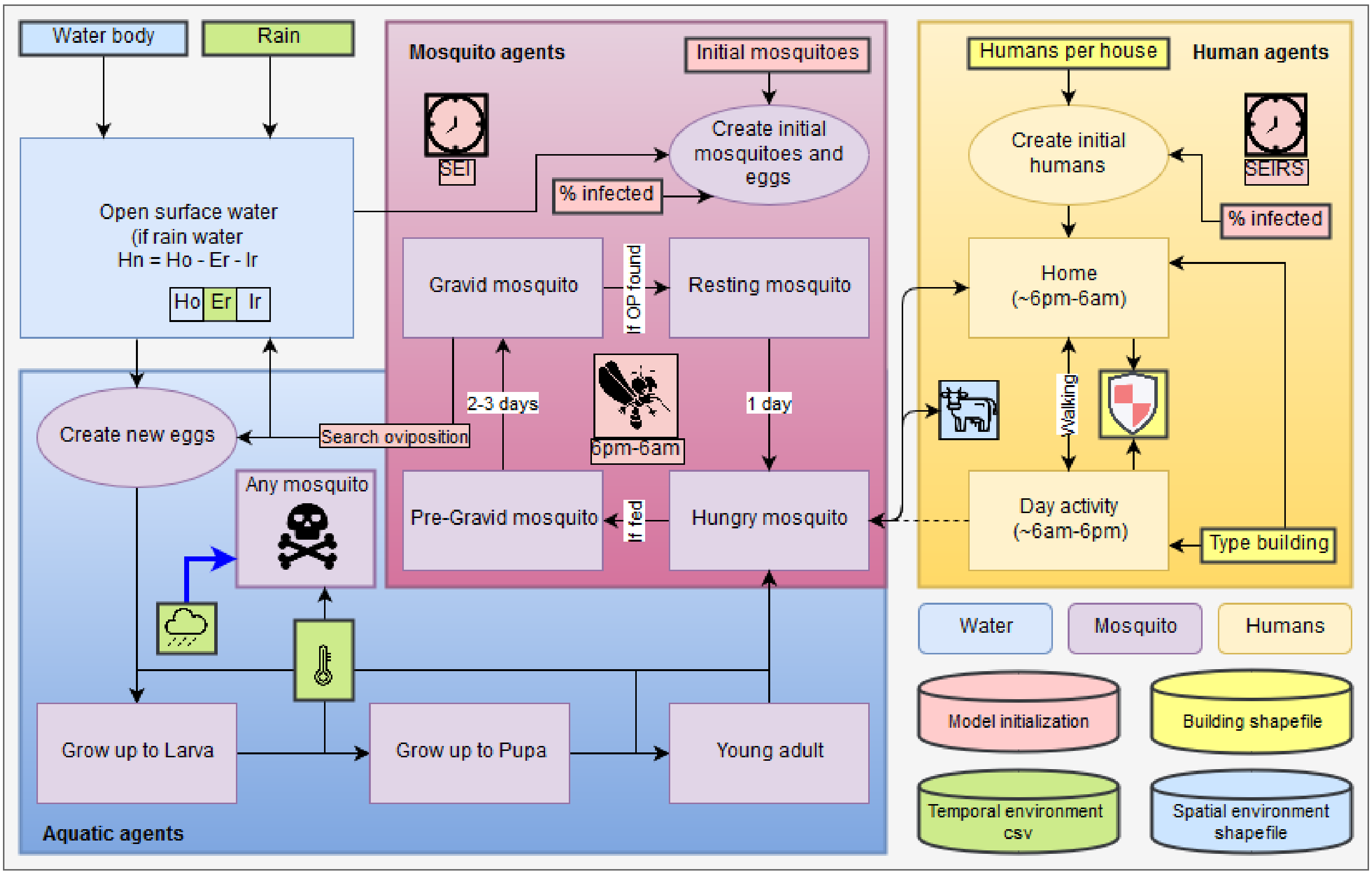

Figure 7: Full model implementation. Hn stands for water height new in mm, Ho stands for water height old in mm, Er for evaporation rate in mm/h and Ir for infiltration rate in mm/h. The blue arrow indicates it only influences adolescent mosquitoes. The bottom right shows what the colours of rectangles mean. Thick black borders mean they are from the inputs and borders similar to its content are processes and input variables.

3.5.1. Global processes

The global processes of the model include any part that is not spatial or has to initialize a spatial component. These processes therefore include the initialization, weather and error detection.

The initialization imports the input data, gives the appropriate values to the imported data as well as creating the agents. Other than that, it also gives values to the initial mosquitoes and infection as seen in table 5 and 7. All mosquitoes are spread equally across the area in its initial state and for each adolescent mosquito it is assumed there is 0.9 larva, 0.05 pupa and 0.05 egg for the initial run. These numbers are based on the number of days each phase takes at 20 ^{0}C. The initial infected humans and adult mosquitoes are a percent of the total. The infection and the percent of people with a risk day activity are randomly determined. All this means there is variation in the initial state. For this reason, an initial run is done to reassure all Monte Carlo runs start with the same state.

The initial mosquito and human data are imported in with the attributes for the main runs. Buildings that are considered risky get a 1x1 meter pool of permanent water to simulate the open water in farms or other risky buildings.

Weather is also included in the global process and takes the csv data one row per day. As the weather input is daily, the evaporation rate and rainfall are divided by 24. Temperature is calculated assuming that 0:00 is the coldest moment of the day and 12:00 the warmest. Therefore, every hour further from 12:00 is closer to the minimum temperature of that day. As the data is per day, rain is assumed to be consistent throughout the day.

Error recognition is also put into the model to communicate any problems that may occur before the full simulation has run. The model will check whether malaria has completely disappeared from the area. This warning is given by text message and ends that run.

3.5.2. Adolescent mosquito processes

Adolescent mosquitoes have only two processes which are death and development. Development is an hourly process that calculates how much development is being performed that hour using the formulas given in table 6. This development is being added to a progressing variable. Once this progressing variable reaches 100%, the adolescent will go to the next development phase in their life. A young adult mosquito that is fully developed creates a hungry mosquito agent at its location.

Each adolescent mosquito agent represents 20 adolescents by default. This is done to speed up the model. This should not have an impact as the process's adolescents go through are identical to one another. Gravid mosquitoes also lay about 80-100 eggs at once, which is why it is chosen to divide by 20. Other numbers could be chosen, but the GAMA scripting language rounds down so it would either have to be 20 or a smaller number.

When a pool has too many adolescents residing in it, all eggs within the pool die as well as 10% of larva. This process gets repeated every hour the weight is above the pool limit. The area of water caused by rainfall is set to be 10% of the actual area. This is an estimated number that simulates the small pools created by the rain. Adolescents that are laid in permanent water bodies are almost killed instantly due to predation.

3.5.3. Adult mosquito processes

Adult mosquitoes go through two universal processes and four phase specific processes. Mortality and SEI are universal and do not change for the different phases. Mortality is a process that checks the age of the mosquito and applies the mortality process daily. SEI activates after a mosquito is infected. Once they are, they add the development of the incubation to an attribute until that attribute reaches 100%. After which the mosquito is permanently infectious.

Movement, day in phase count, procreation and feeding processes are phase specific. Movement can only be performed by gravid and hungry mosquitoes in this model. Other mosquitoes may move in their phases as well, but as this would likely have no impact to the result of the model this is left out. Especially as movement is the most time intensive process of the model. The mosquitoes are active between 6 pm and 6 am. The hungry mosquitoes will first check for any humans in the neighbourhood and fly to them. If they cannot find any humans or if the mosquitoes are gravid, they instead stay within detection range of vegetation or buildings. If they cannot find anything they will wander aimlessly.

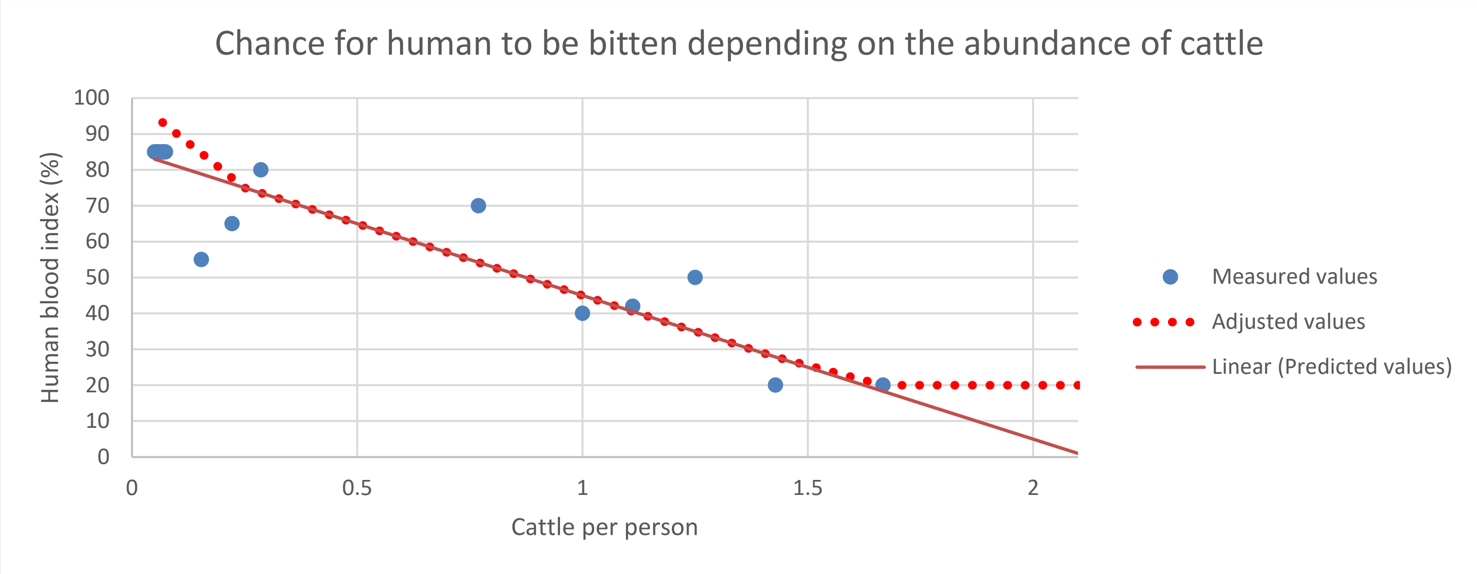

Figure 7: Chance for human to be bitten against cattle abundance. Including measured, predicted and adjusted values [87].

Day in phase count is for the pre-gravid and the resting phase. This process is simply keeping count on how many days they are in their phase and when they should go on to their next phase.

Procreation phase makes the adult mosquito search for any water within detection range. Once they have found surface water in which they can lay eggs, they lay 80-100 eggs in that water and move on to their next phase.

The feeding process is the most complicated out of all processes as it is built out of sub processes. First, they pick a random human within detection range. Then they decide whether they would prefer to feed on cattle or on a human target. If they do decide to feed on a human target, it is checked whether malaria can be transmitted and what the odds are that they are being transmitted.

There is no clear established relation between cattle and humans. However, there have been measurements made on how many humans are stung with a certain amount of cattle in that area. Figure 7 shows these measurements as made by Lindsay et al. [87]. Human blood index refers to the chance a human gets stung.

This figure shows the relation between the chance to be stung more clearly and a predicted trend can be made from this. The linear predicted trend has an R^2 of 0.87, which should make it accurate to follow. However, it shows two problems: Cattle have an almost total preference over 1.625 cattle per person and still have some preference when the cattle per person is close to 0. As humans are always the preferred target, a minimum human blood index is set. Also, when the human blood index reaches above 75%, every 0.01 less cattle equals 1% increased chance for a human to be bitten. This results in the following equation:

$$HBI = (-40\% * R_{cth} + 85\%) \qquad if\ R_{cth} <= 1.625\ and\ R_{cth} >= 0.25$$

$$HBI = 20\% \qquad if\ R_{cth} > 1.625 \quad (6)$$

$$HBI = 100\% - R_{cth} \qquad if\ R_{cth} < 0.25$$

HBI is the human blood index and R_{cth} the cattle to human ratio.

After the chance of being stung successfully is determined, the transmission can be calculated. Which is done by checking whether either the victim or the attacker is infected and infectious. After feeding successfully the mosquito goes to the next phase.

Adult mosquito agents may represent multiple adult mosquitoes at once, which is referred to as Mosquitoes per Agent (MpA) in this model. This significantly increases the speed of the model. However, this has a larger effect on the results than combining the adolescent mosquitoes does. The transmission numbers are changed as if all mosquitoes from that agent sting the same target. An additional input variable is added as well that represents how many of the mosquitoes are infected. As a single agent represents a group of mosquitoes, the weight of the adolescents is multiplied by this number as well. This decision does not have much impact on their movement as mosquitoes often stay within the same neighbourhood.

3.5.4. Human processes

Humans are simulated with two processes: movement and SEIRS. The SEIRS cycle starts the moment a human gets infected. After which the progress of each part is being kept track off, upon reaching 100% going through their next phase. In reality, only the exposed and infectious state are being kept track of. The exposed duration is equal to the intrinsic incubation time and the recovery speed determines the length of the infectious state. Once a human has recovered from malaria, they will become susceptible again.

Movement works different for humans than for mosquitoes. Each day they have a home and a day activity. Once the start time is reached, they practically teleport to their day activity. Upon reaching the end time, they teleport back. This instant movement is implemented as the average movement of a human was large enough to reach any spot on the map within an hour. This starts becoming a problem once the study area is reaching over 5x5 kilometres; which case it should be implemented as regular movement. This was implemented because instant movement reduced run time significantly. Now it does not have to calculate movement for each agent, but rather change location. Such an implementation is much more time efficient and will not harm the results in study areas below 5x5 kilometre.

Every human agent represents one human. After implementing the instant movement of the humans, they do not require much time to be processed thus further simplification will unlikely be necessary.

3.5.5. Spatial environment and building processes

The spatial environment has one process which is to keep track of the surface water. Rain increases the millimetre of water on the land and the evaporation rate combined with the infiltration rate reduce the millimetre of water again. It is assumed that the water availability is consistent throughout the polygon. Once the water level reaches zero, all adolescent mosquitoes inside that polygon die instantly.

Building processes are limited to whether the humans will be safe while they are protected. They will look into the chances a human has to leave their protection as well as the associated time for that kind of protection. ITN would only be active during sleeping hours whereas IRS is active the moment they are inside the building.

Both the spatial environment and the buildings save how much mosquitoes and humans are infected respectively. Saving it as polygons makes the eventual result more insightful as the map will not be cluttered with points. The points could alternatively be saved to then be interpolated, but in the end, this is faster and shows the results just as well.

3.5.6. Running the model

Before the model should be run, it first should go through an initial run. This initial run will be a single run for a year. This is done to stabilize the first positions of the mosquitoes instead of having them spread over the area for the actual run. Another benefit is that for sensitivity analysis, it also gives a consistent initial state.

After the initial run, the main run of the model will be using the Monte Carlo method. This means that the model will run many times to predict the most likely values and calculate the standard deviation. Due to the amount of stochastic values in this model, this is a necessity. It will run 100 simulations at default and cannot be interrupted while it is going through these simulations. Only at the end all simulations the results will be saved. No live simulations are visible for the model in order to increase processing speed. The eventual results are the spatial environment polygons and building polygons. The spatial environment polygons contain information on how many mosquitoes were there and how many of these were infected. The building polygons contains the same information for humans, but then also how many people were infected in total, including those who got recovered.

The eventual results can be further processed in R or another statistical analysis script.

4. Results

This chapter goes through each result, discussing what the results imply and whether this is as it was expected. Results include a Monte Carlo, spatial attribute, stochastic attributes and spatial environment analysis. The Monte Carlo analysis looks into different number of runs, to determine the stability of the model. Spatial attribute analysis is to test the detection range and the MpA on their influence on the model at different levels. Stochastic attribute analysis goes into the different static attributes mentioned in chapter 3.1.1. Finally, the spatial environment analysis looks into the attributes of the associated input and their influence on the malaria spread. This chapter will also often mention a default run. In this run, all attributes and inputs stay as they were discussed in chapter 3.

All runs take place for a hypothetical scenario and validation data was not available within the time of this project.

4.1. Initial state

As the model is developed for a hypothetical situation, inputs are generated using estimated values and the basic topography of the Rusinga area. This area was chosen as it is known to be infested by malaria.

The buildings had the people, protection and the its type randomly generated. The type of building had influence on its protection and its people. It was arbitrarily decided that 10% of the buildings would be risk day activities and another 10% safe day activities; 20% of the people would have risky day activities and the rest safe day activities. All other buildings were households and would only be used outside day activities. People were associated with the households and it was estimated that for every 15 m^2 of building, one person could live or less. ITN protection was only implemented for 20% of the households. IRS could be implemented anywhere with the same chance. An overview of the resulting numbers is given in table 9.

Table 9: Overview of default initial state from the developed inputs and the initial run.

	Initial attribute (number of:)	**Result input development**	**Result initial run**
Input data	People	1,244	994
	Buildings	666	
	Buildings protected	111 (16.7%)	
	People protected	207 (16.6%)	
Model initialization	People infected	373 (30%)	194 (19.5%)
	Adult mosquitoes	10,000	1,001,800
	Mosquitoes infected	10,000 (100%)	400 (<1%)
	Adolescent mosquitoes	400,000	52,916,000

The table shows that the initial mosquitoes infected were set at 100% and the people infected at 30%. This is significantly high and was done to reassure that malaria would stay until the end of the model runs. Despite this, malaria often disappeared from the model. Therefore, 20% of the humans are set to be infected randomly at the end of the initial run. The reason why malaria decreases is further discussed chapter 5. The number of adolescent mosquitoes might seem significant as well, but most of these adolescent mosquitoes die before they grow up. The main contributor to this mortality is the depletion of surface water.

Oddly the number of people decreases after the initial run. This is due to an error in the preparation script. People were assigned to live in buildings that were not classified as homes, causing the model to not recognize these people. This initial result is the only one consistent throughout all initial states and is from now on used as the reference for the number of people in that area. This will no longer be a problem for future use.

A random state was developed for the environment as well. Here the infiltration rate was randomly generated between 0 and 24 mm/day, the cattle to human ratio between 0 and 1 and the vegetation had a 20% chance to occur. The infiltration rate was generated as if it would be clay, which most of Rusinga is. The cattle to human ratio and the vegetation are estimates on what could be realistic for that area.

The initial state described here is applied for the default state, the Monte Carlo analysis and the stochastic attribute analysis. Detection range, MpA and spatial environment runs had the same inputs, but a different initial run. The initial run already included the changes that the main runs would have. The default initial state is also visualized in figure 8. This figure also shows the locations of the risk day activities, which is also randomly spread, although consisting of minor clusters.

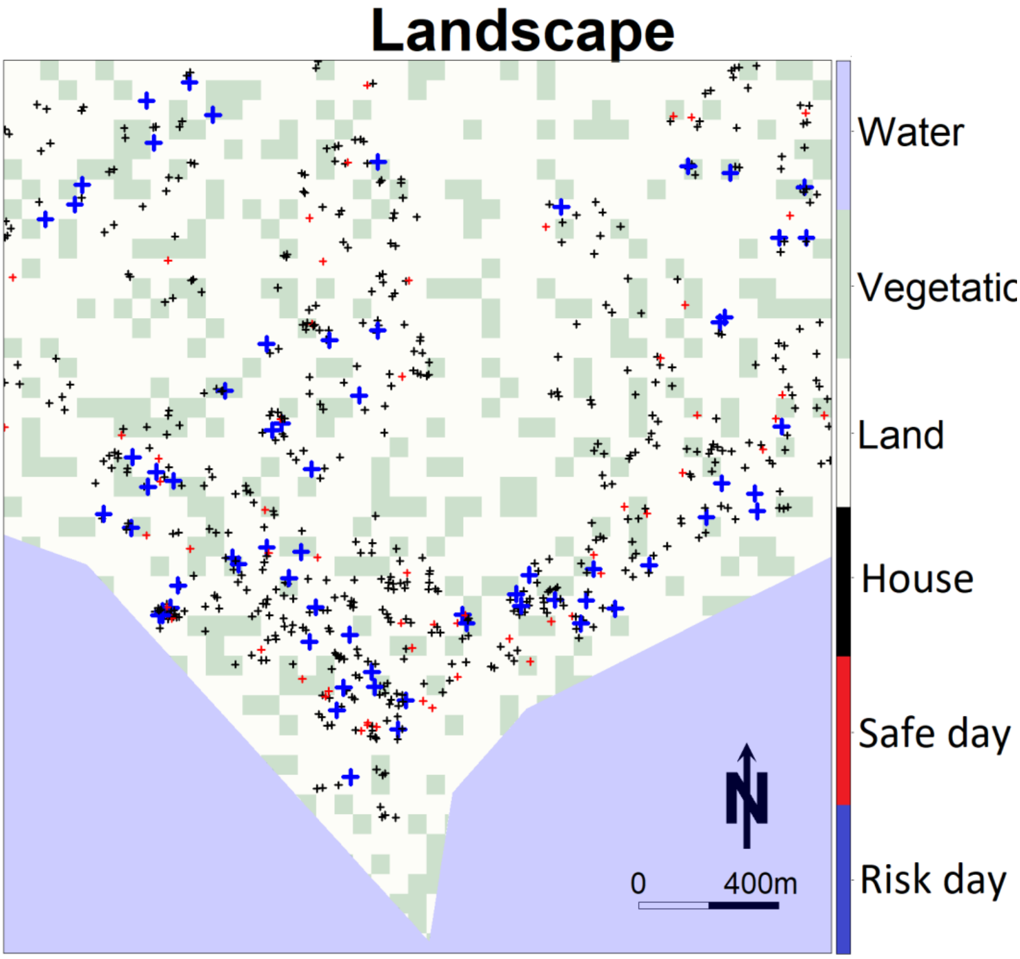

Figure 8: Input state of the vegetation, water and building locations for the default runs. The strong colours indicate buildings and the softer colours the spatial environment.

Not every run uses the default initial state and this has some variation as seen in table 10. As agents were directly imported to the main runs, agents with significant different location or with different counts were made as separate initial runs. The count plays a major role for the MpA, due to their representation of mosquitoes. As the MpA is halved for the analysis, the number of agents would have to be doubled for the mosquitoes. The detection range and spatial environment input attributes, the location was significantly more important. A significantly different starting point from what they would have would require the mosquitoes to stabilize for longer in the main run. This is why they were still kept separate.

Only one initial run was performed for each run shown in table 10. This was because the location of mosquito agents cannot be averaged out over multiple runs. The mosquito agents could be averaged out over their spatial environment and be exported as such. However, this would make exporting and importing mosquitoes more complex. Therefore, it was decided to directly import and export this as agents. As the initial states do not differ much and the resulting plots showed significant spatial relations, this decision will likely an insignificant adverse impact.

Most are in the same line though as the standard deviations are relatively low with the exception of the number infected mosquitoes (see bottom row of table 10). This is closely related to the effects the different attributes have, which will be further explained in their respective chapters. Important to note is that the number infected mosquitoes have a higher standard deviation than a mean. This can only happen if there are a lot of 0 values in the data range, which happens more often in other results as well.

Table 10: Initial state statistics for the entire area.

For chapter	Run name	Number adolescents	Number mosquitoes	Number infected mosquitoes	Number infected humans
4.2.-4.5.	Default	52,916,000	1,001,800	400	185
4.3.	Detection range 10m	53,896,000	1,000,100	0	184
4.3.	Detection range 12m	66,432,000	1,001,800	0	190
4.3.	Detection range 14m	64,296,000	1,001,500	100	200
4.3.	Detection range 16m	56,108,000	999,900	900	214
4.3.	Detection range 18m	60,840,000	1,000,300	2000	209
4.3.	Detection range 20m	62,516,000	1,000,500	2000	190
4.4.	50 mosquitoes per agent	73,888,000	1,000,500	300	194
4.5.	Cattle to human corner	58,840,000	1,000,200	0	191
4.5.	Cattle to human middle	56,864,000	1,000,400	0	197
4.5.	Vegetation corner	63,044,000	1,000,500	800	199
4.5.	Vegetation middle	58,104,000	999,800	0	187
4.5.	Infiltration rate corner	63,016,000	1,000,200	200	188
4.5.	Infiltration rate middle	60,488,000	1,001,300	500	187
	Average of all the above	60,803,429	1,000,629	514	194
	Standard deviation (SD)	5,255,381	656	671	9

The eggs will also not hatch if more than 1,000,000 mosquitoes are present in this model. This was done initially to prevent mosquitoes to go above a hundred times their expected amount. Unfortunately, the mosquitoes count grew significantly more than was expected and this happened frequently. To investigate the number of mosquitoes that would be there without a limit, three additional initial runs were performed. These were done at different MpA values, as shown in table 11. Values above 100 MpA were chosen to speed up computation time. As the numbers show, the mosquitoes are expected to be double the set maximum mosquitoes consistently. Mosquito locations might be affected by this as eggs randomly died due to the mosquito limitation. As malaria is not transmitted through mosquito offspring, it is not expected that this random occurrence will have a significant impact on the results presented in this chapter. The infection in table 11 is varying, but conclusions cannot be made on this many runs. The spatial attribute subchapter will go deeper into the relation of MpA and infection.

Table 11: initial runs without limits on the number of mosquitoes active in the model.

MpA initial run	Number mosquitoes	Number infected humans	Number infected mosquitoes
150	1,830,900	11	1
200	1,996,400	8	0
250	2,011,250	1	0

To avoid wasting time on bad results, the model ends the run when there is no longer malaria in the model. These runs were being counted to be able to analyse in the future how often malaria would disappear from the model. When removing outliers from the model for analysis, these runs without malaria cause much more values to be considered extreme. For this reason, the runs without malaria will not be used in this chapter for the analysis.

The total infected mosquitoes over the year will be used for analysis. The mosquitoes and infections at the end were recorded as well, but these results ended up being spatially homogeneous. This homogeneity is likely because they are only recorded at the end of the year and because of the mosquito limit. Together with the fact that the mosquito densities were only recorded at the end, mosquito locations would average each other out over the different runs. Due to the insight that the total infected in the year has added, another attribute to represent the total mosquitoes over the year would likely grant much more insight on the situation.

It is also chosen to use the total infected mosquitoes as an indication, as they also represent where they receive infection most often. For this to happen, infected humans need to be present there. Mosquitoes cannot obtain malaria from cattle in this model.

4.2. Monte Carlo analysis

The first analysis done is on the number of Monte Carlo runs. The key factor to analyse in this case is to determine how many runs would be necessary in order to have a consistent result. For this, determining the standard deviation for the different input variables at less runs will be compared to those with many runs. The runs could be done in bulk, depending on how many processors are available on the device. A maximum of 7 runs could be done at the same time on one computer, which needed an hour to complete per bulk. This was consistent throughout all different runs.

As only limited time was available, the amount of Monte Carlo runs had to be low by default. A number of 100 runs was chosen as the standard for all the runs and a number of 50 was chosen as a comparison. A Monte Carlo with 200 runs was also tested, but ended up for naught as the process took too long and demanded too much from the system to run. The details of the system used for this research are given in table 12.

Table 12: The specs of the system used for this research.

System component	**Product**
Operating system	Windows 7
Processor	Intel® Core™ i7-6700 CPU @ 3.40GHz 3.40 GHz
RAM	16.0 GB
System type	64-bit OS
Graphics card	Intel® HD Graphics 530

The initial values of these runs were randomized for the spatial environment and the buildings. Although this initial state is randomized, it was the same for both runs. There was no difference between the starting point of both runs.

Table 13: Statistics on the Monte Carlo runs for the entire area.

Name run	**Human**				**Mosquito**						**Total runs**
	Infected at end		Infected in year		Total mosquitoes		Infected at end		Infected in year		With malaria in end state
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
100 runs	1.88	1.25	40	12.46	1000763	540	6.75	13.39	98904	13272	8
50 runs	1.57	0.53	41.71	8.88	500286	193	1.29	3.40	55801	6851	8

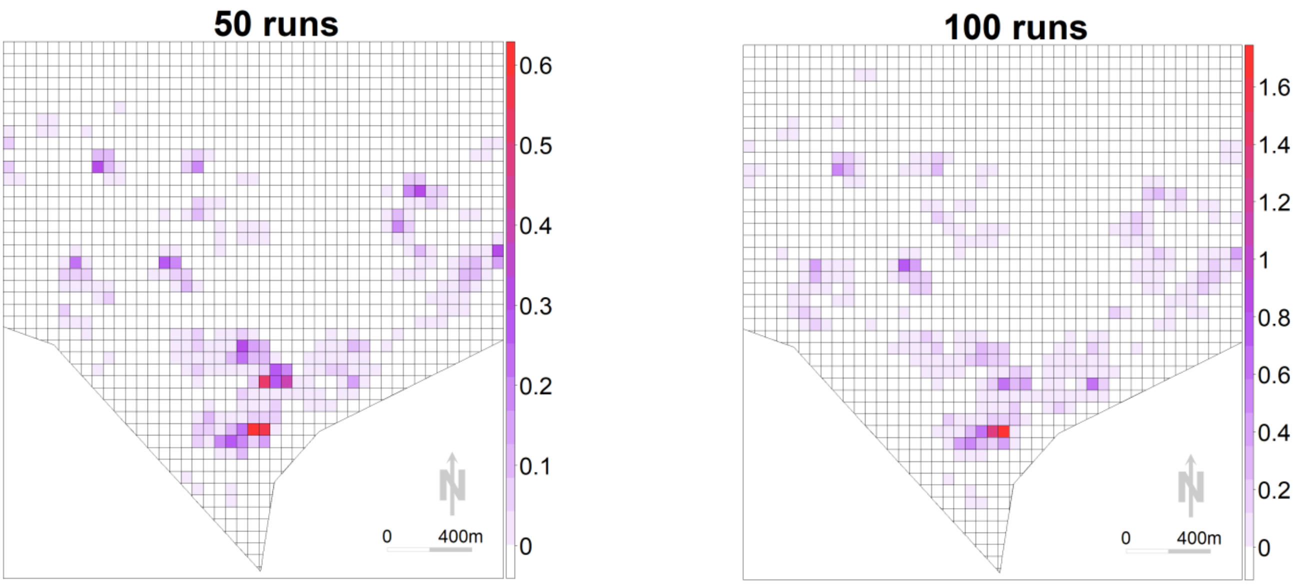

Figure 9: Monte Carlo comparison between 50 runs and 100 runs. This map shows the total incidences that a mosquito gets infected per square meter over the entire year.

Table 13 shows the results for the entire area. The values of each produced result are summed for the entire area and then an average and a standard deviation is calculated over all the Monte Carlo runs. The human infection rates seem to be in common for the 100 and 50 runs. Although logically the standard deviation of the 50 runs should be higher, the amount of runs with malaria in the end state is the same. This means that for both only 8 runs are used to calculate the mean and standard deviation. The 50 runs can therefore be considered an exception; it has double the runs ending with malaria than it should have.

The results on the mosquitoes seem off as the 100 runs seem to have doubled values from the 50 runs. At first glance, it might look as if the results therefore are summed up from all the runs instead of taking the mean. However, this was a mistake caused by the maximum mosquitoes in the model. The limit here was set at 500,000 instead of a million. This is now fixed for future use. Despite the double numbers, they do have spatial patterns in common. This is visualized in figure 9.

As the model has similar patterns between 50 and 100 runs, the 50 runs would likely provide accurate results after the removal of the maximum mosquitoes and the solution to the malaria extinction. As this was found out after the models have run, all other results except the MpA still use 100 runs.

4.3. Spatial attribute analysis

There are two attributes that were assumed to be of high importance when testing the model: detection range and mosquitoes per agent. Representing more mosquitoes in one agent could have adverse effects on the end result and therefore an analysis was done to see what the difference would be at different levels of MpA. The detection range is an attribute that is expected to have a high influence, but also with high uncertainty. Not many developers defined when the mosquitoes would recognize their targets and those who did disagreed with one another. The detection range is tested at increments of 2m between 10m and 20m. This is in addition to the default run which runs the detection range at 15m. This was kept close around 15m as this was confirmed to be a realistic number by the entomologist [25].

Table 14 shows the results for the entire area. The number of infected at the end of the year seem mostly consistent throughout the runs. On the other hand, the total infected humans and mosquitoes is rising with the detection range. This is likely because mosquitoes with high detection range should be able to go through their phases quicker. They were able to find human blood meals and ovipositions much quicker than the mosquitoes with a 10-meter detection range.

Table 14: Detection range statistical analysis for the entire area. DR stands for detection range. This table represents 100 Monte Carlo runs.

DR	**Human**				**Mosquito**						**Total runs**
	Infected at end		Infected in year		Total mosquitoes		Infected at end		Infected in year		With malaria in end state
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
10m	2.00	1.41	14.50	4.95	1001250	636	9.00	12.73	37353	5846	2
12m	1.00	0.00	10.00	0.00	1000000	0	0.00	0.00	27693	0	1
14m	2.20	1.30	32.00	9.41	1000720	476	14.40	23.47	72339	6550	5
15m	1.88	1.25	40.00	12.47	1000763	540	6.75	13.39	98904	13272	8
16m	2.72	1.90	52.56	13.73	1000867	653	29.83	32.06	119624	10662	18
18m	3.17	3.41	63.10	12.45	1000390	801	23.40	35.10	161982	18521	30
20m	2.80	2.48	75.84	15.26	999833	1615	22.00	27.91	194534	23376	45

The run with 12m detection range has a standard deviation of 0, because there was only one run out of the hundred with actual malaria remaining in the model. As discussed before, these runs are excluded from the analysis. The extinction of malaria for the 12m detection range is likely due to chance.

Figure 10 shows that the mosquitoes most of all stay around the coast line. This is likely because it is the most densely populated zone in the area. There is also more population present on the sides of the study area. The low amount of infections there is likely related to the design of the model. Borders cannot be crossed by mosquitoes in this model. Mosquitoes would only be able to reach these populated areas specifically from the middle of the area whereas the middle area can be reached from anywhere. This modelling phenomenon is especially visible in the lower detection ranges. As mosquitoes there cannot find new areas to live in as easily as the high detection ranges, making them more likely to stay where they are.

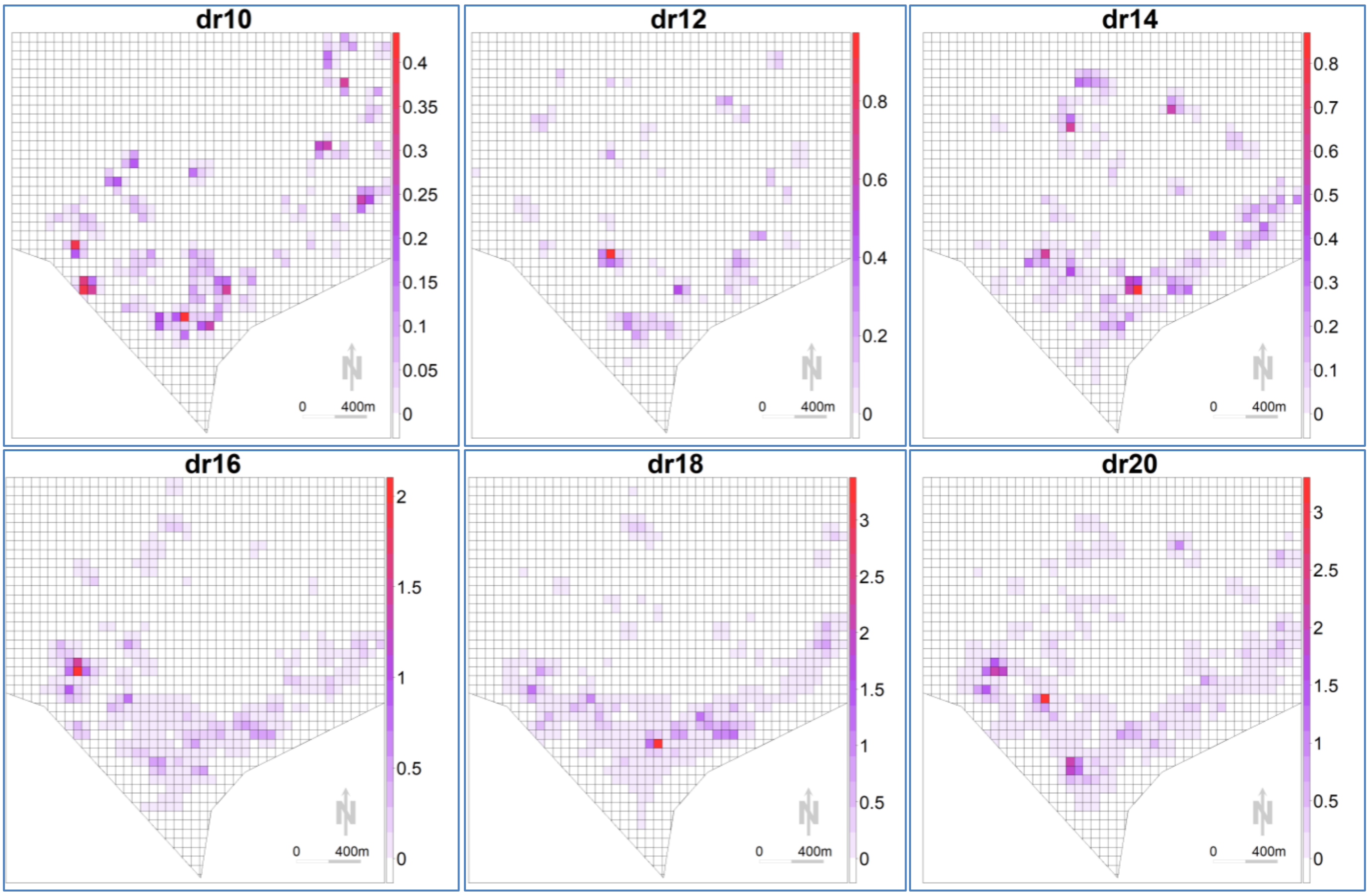

Figure 10: Detection range comparison between 10m, 12m, 14m, 16m, 18m and 20m detection. This map shows the total incidences that a mosquito gets infected per square meter over the entire year.

The MpA analysis was done at 50 MpA and at 100 MpA, the latter being the default value used for the other runs. 50 MpA uses only 50 Monte Carlo runs due to its intense runtime and demand from the system. This is also compared to the 50 MC runs from the previous paragraph. A run for 1 MpA could not be done due to time constraints.

Table 15 shows the difference between 100 MpA and 50 MpA. As expected, the 100 MpA has a higher standard deviation in general compared to the 50 agents. The means seem to be about the same for both runs, which means the overall difference is insignificant. Figure 11 further establishes this as well, with similar maximum densities and spread on the area. This is likely because a higher MpA also relates to a higher total weight of adolescent mosquitoes. It was found that at extreme MpA's, mosquitoes could not lay eggs as the eggs would die instantly due to the maximum number of adolescents per square meter. This may play a role here as well. Offspring laid smaller pools would die quicker at 100 MpA. Smaller pools in this model are predominantly from risk day activity buildings. This is likely why the 100 MpA densities are around an area that is crowded with people and risk type buildings. Some difference might also occur from the initial state, but it would likely not be significant.

Table 15: Statistical analysis for the mosquitoes per agent runs on the entire area. This table represents 50 Monte Carlo runs.

Name run	**Human**				**Mosquito**						**Total runs**
	Infected at end		Infected in year		Total mosquitoes		Infected at end		Infected in year		With malaria at end state
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
50 MpA	1.57	0.53	41.71	8.88	500286	193	1.29	3.40	55801	6851	7
100 MpA	2.00	1.31	38.13	6.75	500244	645	3.38	4.66	49745	4839	8

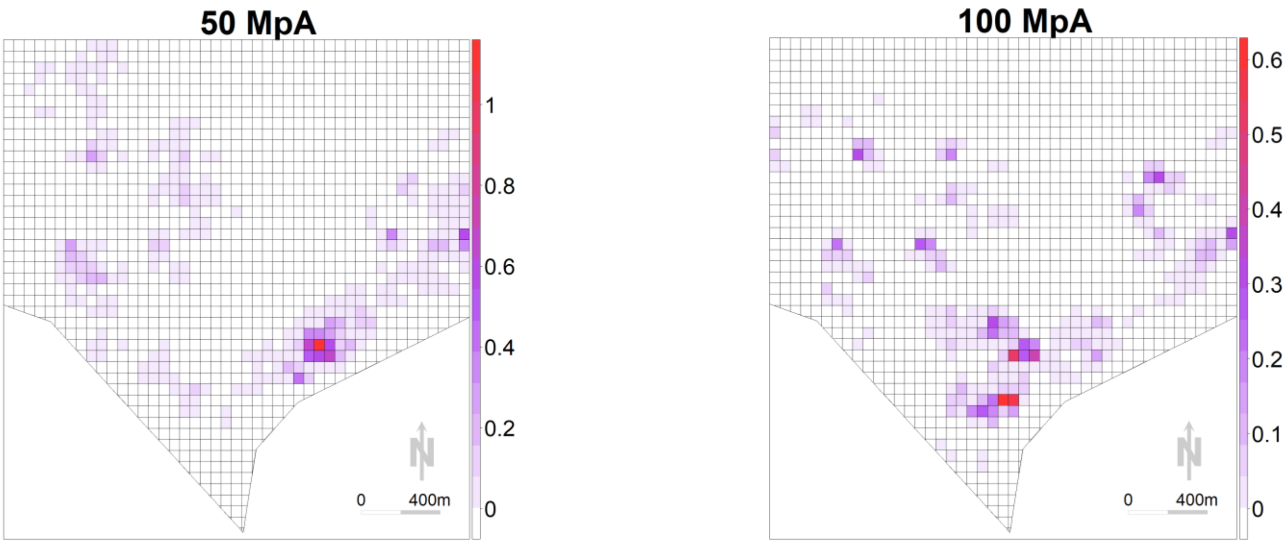

Figure 11: Total mosquito infection spread for 100 MpA and 50 MpA. This map shows the total incidences that a mosquito gets infected per square meter over the entire year.

4.4. Stochastic attribute analysis

There are a few stochastic attributes in the model, namely the fecundity, start and end time of human movement, the recovery rate and the mosquito movement speed. Although it is agreed upon by different developers that these value ranges are correct, it is unknown how much they would contribute to the variability of the model. For these runs, all of the stochastic attributes are fixed to a standard value with the exception of the attribute the run is named after. As an example, speed would be set at 300 meter per day unless the run was named “speed”. In that case it would be a random value between 200-400 meter per day. In the default run, all these attributes remain stochastic. As a comparison, a minimum randomness run was done as well. In this run all stochastic attributes are fixed to their standard value.

Table 16: Stochastic values and their statistics for the entire area. This table represents 100 Monte Carlo runs.

Name run	Human				Mosquito						Total runs
	Infected at end		Infected in year		Total mosquitoes		Infected at end		Infected in year		With malaria at end state
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Default	1.88	1.25	40.00	12.47	1000763	540	6.8	13.4	98904	13272	8
Minimum	1.40	0.55	25.80	6.53	1000760	757	7.2	16.1	76413	2968	5
Fecundity	2.00	1.67	28.17	11.02	1001167	472	6.0	9.3	81523	8742	6
Start day	1.40	0.70	29.60	7.09	1000820	800	34.2	55.9	87026	6833	10
Recovery	1.40	0.55	22.20	9.76	1000220	415	7.2	16.1	79119	11724	5
Mosq. speed	1.00	0.00	25.00	0.00	1000600	0	0.0	0.0	95206	0	1

The initial states are all as they would be for the default run. This is to prevent potential variation from the initial state itself for a better comparison with the default state.

Table 16 shows the statistics for the stochastic attributes. The default state does have the highest standard deviation in most cases which is as expected and the minimum randomness indeed has the lowest. For some results the start day attribute also has a high standard deviation, which is likely because this attribute is directly related to the interactivity between humans and mosquitoes. As interactivity between these two agents determines the malaria spread, it is high likely that this attribute is one of the main contributors to the uncertainty of the model.

The reason as to why start day likely has the most total runs with malaria is because humans have an additional chance to be more active in the night than the other runs. The standard values for these runs would be 8:00 and 22:00, whereas mosquitoes in this run will also be able to find blood meals later in that night. This is relevant as those hours they would be out of their ITN's and IRS, thus being more likely to be stung without having the mosquito die. Despite this fact, the difference is not significant.

Mosquito speed has the lowest amount of runs with malaria at the end state. This might be because the stability of 300 meter per day gives other models more interaction between humans and mosquitoes. If a mosquito would have 200 meter per day a couple days in a row, this would already significantly reduce the interaction. Whereas the 400 meter per day would not necessarily increase the interaction.

Figure 12 shows that these stochastic values do not per se have an influence on the spatial distribution. The intensities and the spread are about the same as they are for the default run.

Figure 12: Stochastic attribute maps. This map shows the total incidences that a mosquito gets infected per square meter over the entire year.

4.5. Spatial environment analysis

The spatial environment in this model consists of three varying attributes: Cattle to human ratio, infiltration rate and vegetation. To understand the impact of these different attributes, a Monte Carlo run is done for three scenarios: a vegetation scenario (Veg), a cattle scenario (Cth) and an infiltration rate scenario (Inf). These scenarios are tested at two locations in the environment (corner and middle). Figure 13 shows the initial values for the scenarios. A circle was chosen so it would be positioned between most buildings, which makes it easier to recognize the effects of the spatial environment specifically. For the corner a square was used as that would lower the effects of the border avoidance agents have in this model. The bottom left corner of the green square is closer to the middle after all. This is represented in figure 13. Both green areas are considered one polygon for this analysis.

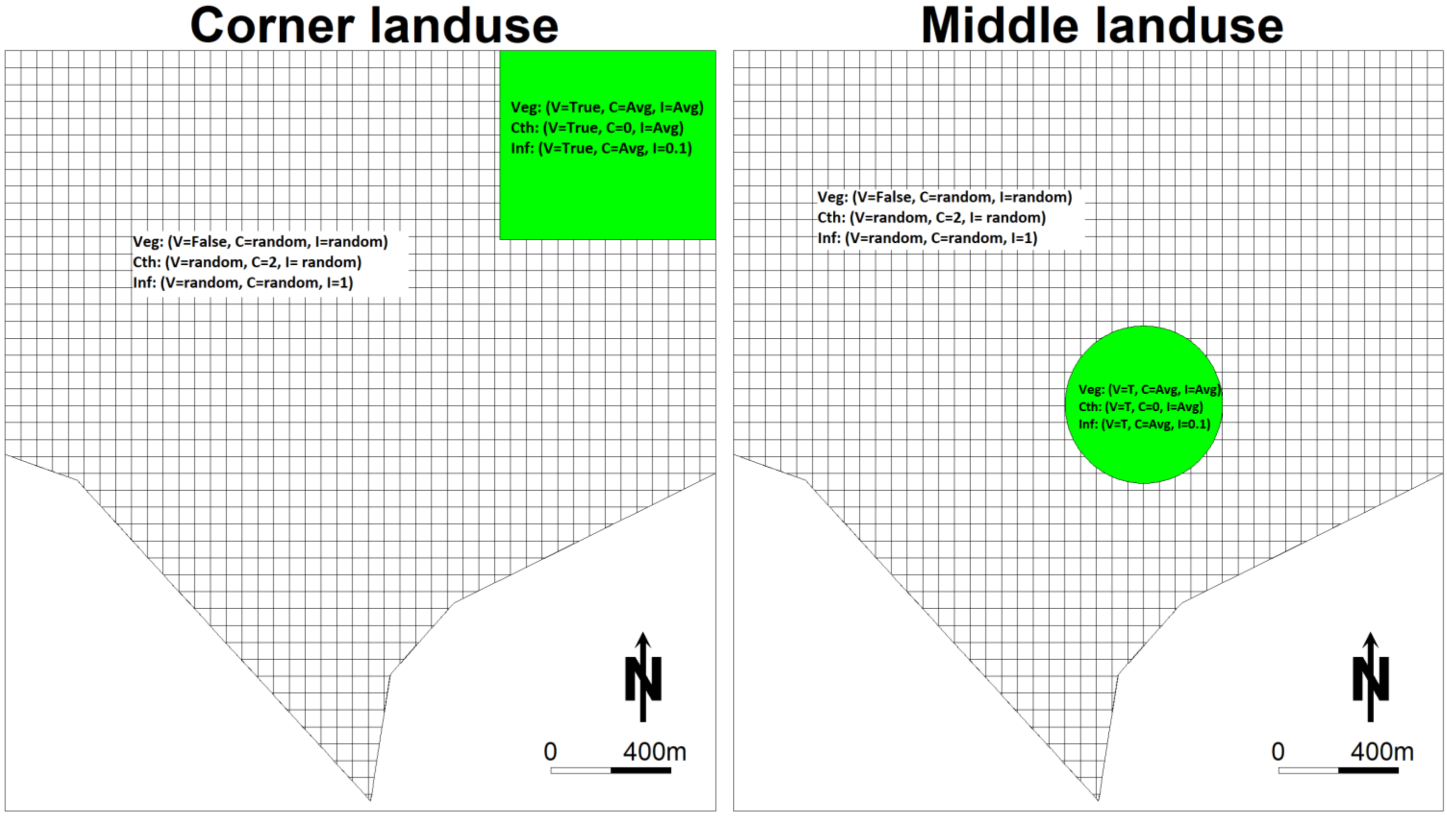

Figure 13: Corner and middle land states for the initial and the main runs. Veg stands for vegetation, Cth for cattle to human ratio, Inf for infiltration rate and Avg for average. True is abbreviated to T in the middle land use plot to save space. Text on the white grid counts for the entire white grid.

For these scenarios, it is chosen that the green areas would be the preferred area for mosquitoes. Vegetation has impact on the position of the mosquito, cattle to human ratio an impact on the spread of malaria and the infiltration rate on the possibilities for eggs to develop. Therefore, malaria is always be most prevalent in the green areas of figure 13.

Table 17 shows the statistics for the different spatial environment runs. All the runs with the green areas in the middle have a higher infection than those in the corner. This makes sense as mosquitoes cannot enter from outside the boundaries. Most notable is the disappearance of malaria for the cattle runs (CtH in table 16). This is due to there being only 20% chance mosquitoes will sting humans in the majority of the area. This proves that cattle are definitely an important factor to take into account, but it would probably require further analysis to what extent this is the case.

Infiltration rate seems to act mostly as expected. Although the majority of the area has a significantly higher depletion rate, there is not much decrease on the total mosquitoes compared to other runs. This is likely due to the spread of risk day activities across the area.

The vegetation scenario for the location in the middle has the most infections throughout the year. This is likely because there were no extremes in the cattle nor the infiltration rates. There is more total vegetation present in the CtH and Inf runs, which might be a potential reason for mosquitoes to be more infectious in the Veg runs. As mosquitoes will not be able to find vegetation to hide in, they would more likely decide to hide within buildings instead. This would cause them to interact more with humans. Figure 14 also shows that the mosquitoes are barely present in the green areas shown in figure 13 and actually much more so in the other areas.

Table 17: Statistics of the entire area for the different spatial environment scenarios. Mosquitoes for the cattle are removed due to the error recognition of malaria becoming 0. For all the runs, cattle ended up with 0 malaria. This table represents 100 Monte Carlo runs.

Name run	**Human**				**Mosquito**						**Total runs**
	Infected at end		Infected in year		Total mosquitoes		Infected at end		Infected in year		With malaria at end state
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Default	1.88	1.25	40.00	12.47	1000763	540	6.75	13.39	98904	13272	8
Veg corner	1.85	1.72	31.85	7.74	1000738	744	9.69	29.95	80822	5353	13
Veg middle	1.92	0.95	44.08	10.77	1000800	723	5.54	8.65	110875	14544	13
CtH corner	0.00	0.00	1.55	1.29	-	-	0.00	0.00	15537	2808	0
CtH middle	0.00	0.00	1.82	1.32	-	-	0.00	0.00	16535	2566	0
Inf corner	1.83	1.53	32.00	8.57	1000625	982	19.50	17.93	58031	7800	5
Inf middle	0.80	0.45	27.40	11.87	1000680	832	3.60	8.05	83767	13438	12

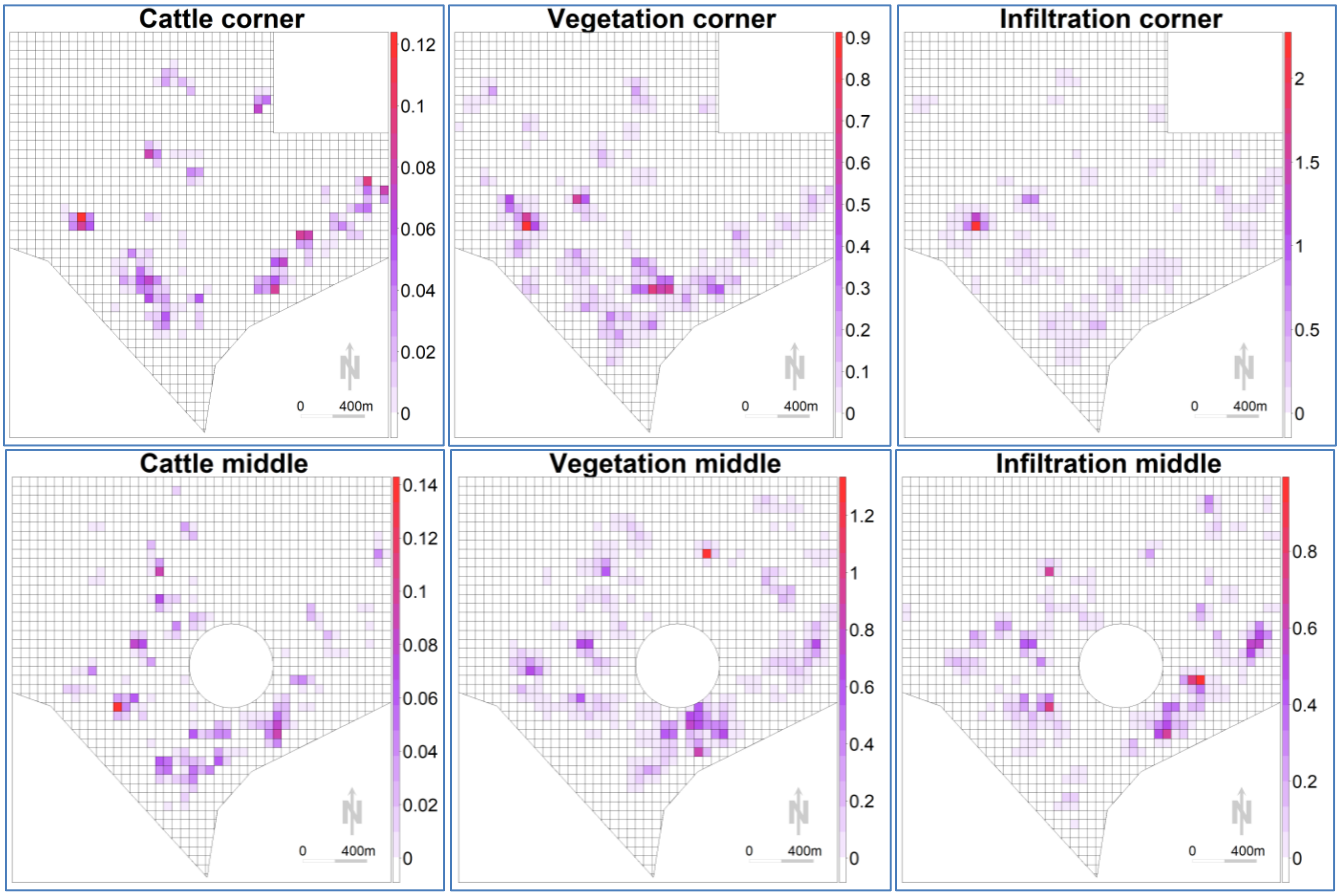

Figure 14: The different spatial environment variations.

Figure 14 shows that the infection spread is like most other spreads. The plots with the middle part show that most of the infection is indeed nearby the circle and seems about equally spread around it. The corner plots show that the infection is predominantly opposite of what would be the preferred place for mosquitoes. This is likely because there is a much higher population there as shown in figure 15.

The middle plots show surprisingly more activity in the corner than the actual corner plots. There does not seem to be a clear relation between the green area of figure 13 and the infection of figure 14. This counts for all figures, even human infection and mosquito positions.

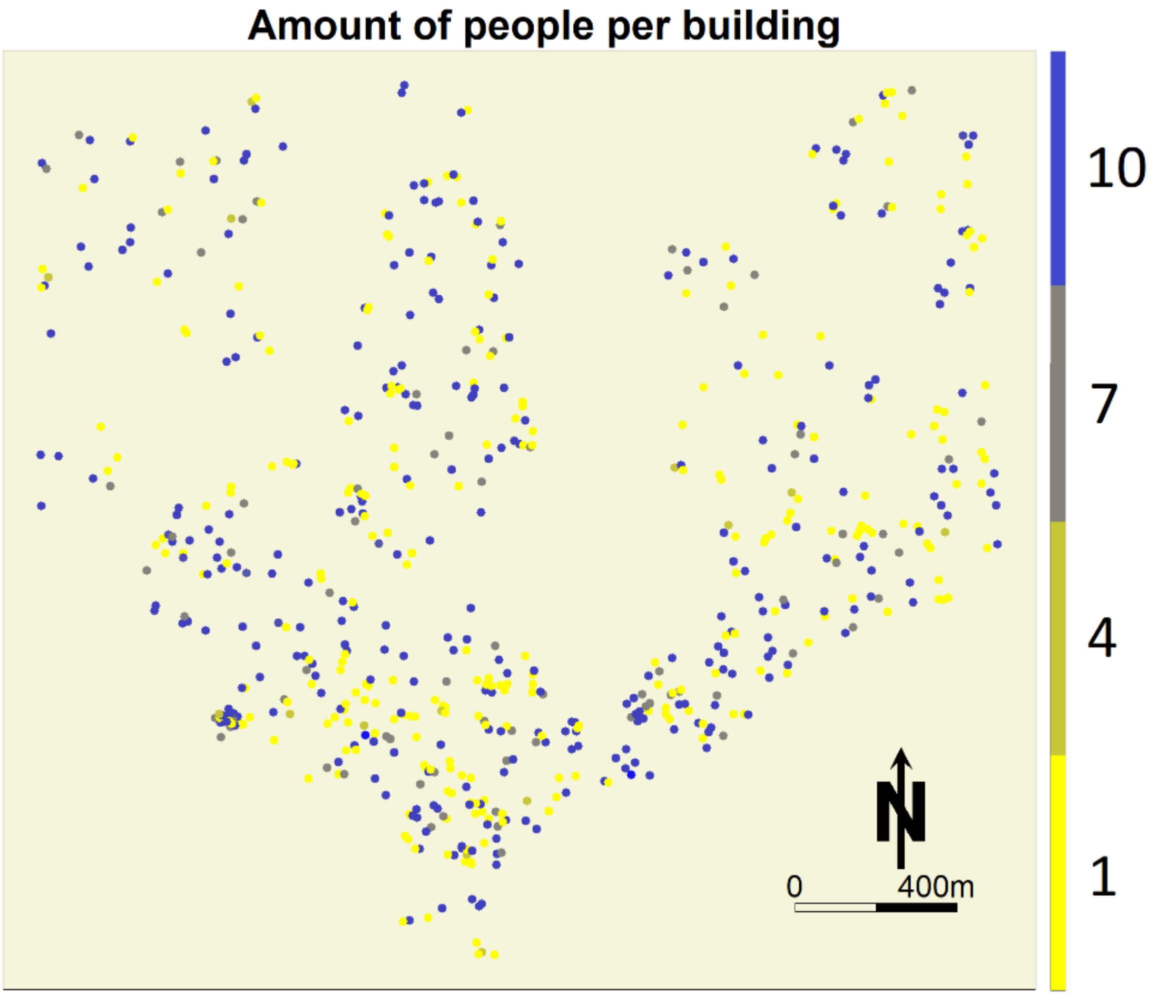

Figure 15: Amount of people per building.

5. Discussion

After analysing all the results for the different comparisons, some general observations can be made: very few runs have malaria at the end, the inputs are randomly generated and most infection densities are away from the borders and nearby populated areas. The low number of infections near the borders is due to the model restrictions. Opening the borders would allow to simulate the migration of the mosquitoes as well. This border opening is called a torus [88]. This allows the mosquitoes to go from one side of the border to the other as if they would teleport, which would simulate migration. Most developers do not share whether they use torus with the exception of Gu and Novak [7] who did not use it. This research also stated that study areas should be above 2x2 km, which is likely directly related to this. As there is little information available on the effects of torus, a conclusion cannot be made on its effects [17]. For this study area specifically, it would likely not have worked anyhow, as Anopheles mosquitoes do not usually come from large water bodies [25].

What is more concerning is the number of runs that do not have malaria at the end. The hypothesized area is based on the Rusinga area where malaria is prevalent, meaning that this should not be the case. How this would have to be resolved entirely depends on calibration based on a validation area. The necessary data was being requested halfway during the research, but unfortunately due to communicative problems this was delayed. After the realization that the data that would be available might not be in a workable format, the validation was cancelled. Entomologists often measure data non-spatially, requiring a significant amount of time to implement this properly.

Then the input problem ties in with the validation, but also the model itself. Almost all spatial information of the hypothesized area is estimated with the exception of the location of the buildings and water bodies. This goes against the first principle of the model development (see page 10). Cattle densities, infiltration rates and malaria prevention methods are not often recorded for areas such as this. Almost always these have to be estimated, unless a field study takes place. It could be argued that the model is flexible at least because of these estimations. For example, one could link land use with different cattle to human ratios and infiltration rates.

Although there is no validation available, it can be hypothesized why malaria was disappearing. The most likely reason is the detection range, as many researchers disagreed on the actual value of this attribute even though the attribute has a significant influence on the end result (see chapter 4.3.). Gu and Novak [7] also made a model for a hypothetical area and they calibrated the detection range at 40 meter. Zhu et al. [23] also used a hypothetical area, except his detection range was 40 meter for recognizing humans and 5 meter for everything else. However, Bomblies et al. [12] state that the detection range for finding blood meals is dependent on the density of cattle, wind speed and wind direction. It is most likely that the latter would be correct as more studies show that this would be the case [89]. These variables change for every study area and therefore it is likely that detection range needs to become an input variable or an area dependent attribute for this model. In hindsight, the value of 40 meter detection range should also have been tested. This was only skipped over as it seemed extremely different from the expected 15 meter value.

Another potential problem could be the human to mosquito transmission rate and the influence of cattle. It was estimated that the transmission would be 17.5%, but could potentially be 20%. This, of course would increase the chance malaria would spread. The role of cattle is unclear as not many general statements are made on the topic. Often the cattle is separated by different types of cattle, such as goat or sheep [90, 91]. They have different influences on cattle but have been simplified in this model to a single number.

All the resulting plots suffered from the maximum mosquito limit. The analysis of represented in table 11 shows that the mosquito count would likely have doubled. In which case, malaria would likely be more prevalent as well [37]. The maximum also likely contributed to the homogenous mosquito spread. Information on the mosquito's whereabouts could have been especially relevant for determining which spots would be best for Larvae Source Management (LSM). Adolescent mosquito results could have been used for this as well. This was worked on during development, but did not produce results. The saving did not work due to an error in defining the attribute that keeps track of this value.

While working on this model, two main problems became apparent: the lack of knowledge on entomology and the lack of experience with modelling in GAMA. This is the first research in which I am working with either entomology or GAMA. This inexperience caused many decisions that were made early on in the project that would slow down the rest of the project. Processes that were simplified by a developer or were not clarified in their meaning were still implemented into the model. Of course, this resulted in time inefficient implementation of processes.

Lack of entomological knowledge also caused difficulties in analysing the results. One can assume what the results could mean in the context of reality, but without the understanding of entomology this is more likely mistaken. As such it is unknown what realistic mosquito counts would be or what realistic malaria spreads are. Analysing a hypothetical scenario is extra challenging as there are no results to compare to. Zhu et al. [23] as well as Gu and Novak [7] are confident in their scenarios, but their goal is different from this model. This model is designed to be flexible and used for different areas to represent a realistic malaria spread. Their models are made to test the effectiveness of malaria reduction methods. This also meant that all contributing factors towards the spread of malaria are of higher relevance here than with their hypothetical models [88].

Having little experience with GAMA made efficient implementation of processes challenging. Most decisions that were made had to do with time efficiency and reducing the demands on the system. Understanding what the most efficient processes were and how they could be improved took a significant amount of time. This caused the model development to take longer than expected and less time was available for analysing results themselves. If more time for results were available, the adult and adolescent mosquito locations could have been analysed as well. Unnecessary mistakes such as the disappearance of humans and the maximum number of mosquitoes could have completely been avoided.

6. Conclusion and advice

The goal of this research was to create a flexible agent-based model for analysing malaria spread with spatial components. This chapter goes into what conclusions can be made on this research and is recommended to do further after this study.

6.1. Conclusion

This research consists of four research questions, related to: the attributes, inputs, processes and validation. The attributes were mostly agreed upon in the reviewed research and can therefore be considered accurate. Attributes that developers disagreed on were settled by consulting an entomological expert and can also be considered correct. Detection range and human to mosquito transmission values are the largest uncertainties. There is disagreement on these values both by developers and entomologists. It is likely that the detection range differs depending on the area. The influence of transmission was not tested in this research and could use further testing as well.

All inputs and their attributes are considered relevant for the model. From the spatial environment analysis, it appears most inputs are relevant and that the change of attributes shows the expected outcome (table 17). Buildings and the temporal input attributes were not investigated as they were commonly agreed upon by the developers. Evaporation rate is the exception to this rule, but was necessary to calculate the depletion rate of water. It is certain that all these inputs are necessary for creating a spatial malaria agent-based model. Wind and humidity could be added to the model as well, which would improve the detection range, evaporation rate and the mosquito movement patterns. Adding these mean that more calculations and processes would be added to the model. Therefore, it should only be done once the model is made more time efficient.

The model was supposedly meant to be flexible, by having multiple ways of accommodating inputs. However, this was limited as a spatial model requires inputs that are not always easily accessible. These include for example, the evaporation rate, malaria preventive methods, infiltration rate and cattle to human ratio. There are also human behaviour processes that require field observations such as their movement times and the chance they will not use their malaria preventive methods.

The processes implemented were generally considered relevant, but some are uncertain in their implementation. The influence of cattle is uncertain as there has only been one research found with actual measurements on the human blood index at different cattle abundancies. Mosquitoes have been bundled into less agents as well to speed up the the model. This did not show much influence, but could potentially still cause inaccuracies. Especially as this only has been tested at higher MpA's. The other processes are not expected to cause uncertainties. Human movement might cause problems at study areas larger than 5x5 kilometre.

Although the models have been run and the results have been generated, it is uncertain what the accuracy of the model is. There was not enough time to validate the model. When a validation method would be applied, the problems could further be analysed and a more realistic model can be developed.

6.2. Advice

The goal of the research was partially reached, thus an advice for further research is set up. A distinction is made between two types of advices: actions and research. Actions go into the current problems the model has and that definitely should be improved. Research goes into topics that could be interesting to add to the model.

6.2.1. Actions

It is recommended to perform the following actions to further develop the model:

- Add a total adult and adolescent mosquitoes within area throughout time as a result of the model;
- Apply the model for two different areas to test its flexibility;
- Change the detection range to an input variable or an area dependent attribute;
- Apply a proper sensitivity analysis;
- Test the model with a higher number of Monte Carlo runs;
- Further test the influence of mosquitoes per agent and the human to mosquito transmission.

The total infected mosquitoes within the area are a consistent way of showing results that also highlights hotspots of infection. However, it is currently the only result that is being kept track of for the whole year. To add the total adult and adolescent mosquitoes, hotspots of where mosquitoes would live could give a lot of insight on how the spread of malaria could potentially be controlled. The implementation of this statistic would be similar to that of the total infected mosquitoes, so it would be simple to implement.

Validation of the model is the most important factor that was missing from this research. Without the validation it cannot be determined whether the simulated spread of malaria is representative of a real scenario. As this model was designed to be implementable for multiple areas, it is recommended to look into two different areas. If the model would produce realistic results for both models with only a change of area dependent attributes and inputs, the model can be considered flexible.

The detection range is likely to be area dependent and thus will likely have to be changed for that. This could either be dependent on the area and has to be calibrated for each area that is testing. Alternatively, it could be used as an input variable in which the detection range would adapt to the weather in that area. Either options should be tested for computation speed and reliability.

Another research result that helps understanding the quality of the model is the sensitivity analysis. This would give insight on the stochasticity of the model. There are many ways to perform a sensitivity analysis, so it would also have to be researched what the best way to perform this would be for this model. The building and temporal inputs for example were not analysed in this study.

The current results are based on 100 runs. Although there are patterns visible already in this number of runs, some of the differences remain unclear whether they are caused because of the difference in set up or whether it happened because of stochasticity of the model. The simple comparison between a 100 run and 50 run Monte Carlo is a clear example of such a difference in stochasticity and that this is still apparent. Once the model runs quicker, it would be much more interesting to run this at 1000 runs for example.

Mosquito per agent is an attribute designed in this model to reduce the necessary computation time. Although this attribute is tested at 100 and at 50 MpA, it is potentially necessary to also look at this at 1 MpA. This would give insight on whether more mosquitoes could be simplified to one agent in the first place. This parameter is also directly related to the human to mosquito transmission which should be further investigated as well.

These actions reassure that the model can be tested and developed to be a flexible and realistic model.

6.2.2. Research

While further developing this model, the following are recommended to research:

- Create a better comparison for the infiltration rates and cattle to human ratios;
- How the model could be made to be more time efficient;
- Test the influence of open borders within the model;
- Find appliances and communicate with interested model users.

Although a spatial environment analysis has been performed, the infiltration rate and cattle to human ratios can probably use more testing. Preferably with more extreme cases for the infiltration rates and less extreme for the cattle to human ratios. This in combination with the total mosquitoes could give a better insight on how these attributes affect the results.

Most decisions made while building the model were in relation to time efficiency. The MpA and 100 Monte Carlo runs are prime examples of this. If this model could be made more time efficient, more Monte Carlo runs can be performed and the 1 MpA test can be run. Making this more time efficient will also demand less from the system.

One of the noticeable things of the results is that the mosquitoes stay relatively away from the borders. This is due to the closed borders of the model. The borders could be opened to simulate migration, causing the mosquitoes to teleport from one end to the other when leaving the study area. This phenomenon could help make the model realistic, but can also potentially create new problems. For example, the mosquitoes leaving the area north would result in them entering the area from the water. This is not realistic as Anopheles Gambiae does not usually come from larger water bodies.

Finally, this research was set up to add a new type of model to the currently existing ones. However, it was not made with a user in mind. It would be good to find interested users for the model so that the model can be further developed for that target audience. Sander Koenraadt has shown interest in further experimenting with the model, thinking of doing more climate analysis. This can of course be performed, but closer communication would be valuable as it guarantees the use of this model in the future.

Bibliography

1. Killeen, G.F., et al., *Preventing Childhood Malaria in Africa by Protecting Adults from Mosquitoes with Insecticide-Treated Nets.* PLOS Medicine, 2007. **4**(7): p. e229.
2. World Health Organization, *Monitoring health for the sustainable development goals.*, in *World health statistics*. 2018, WHO: Geneva. p. 21.
3. Patz, J., et al., *Impact of Regional Climate Change on Human Health. Nature 438: 310-317 (17 November)*. Vol. 438. 2005. 310-7.
4. Paaijmans, K., et al., *Influence of climate on malaria transmission depends on daily temperature variation*. Vol. 107. 2010. 15135-9.
5. T. Crooks, A., et al., *Perspectives on Agent-Based Models and Geographical Systems*. 2012.
6. United Nations, *The Millennium Development Goals Report*. 2015, UN: New York.
7. Gu, W. and R. Novak, *Agent-based modelling of mosquito foraging behaviour for malaria control*. Vol. 103. 2009. 1105-12.
8. Bomblies, A., *Agent-based modeling of malaria vectors: The importance of spatial simulation*. Vol. 7. 2014. 308.
9. Pizzitutti, F., et al., *Out of the net: An agent-based model to study human movements influence on local-scale malaria transmission.* PLOS ONE, 2018. **13**(3): p. e0193493.
10. Jindal, A. and S. Rao, *Agent-Based Modeling and Simulation of Mosquito-Borne Disease Transmission*, in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 2017, International Foundation for Autonomous Agents and Multiagent Systems: São Paulo, Brazil. p. 426-435.
11. Matthews, R.B., et al., *Agent-based land-use models: a review of applications.* Landscape Ecology, 2007. **22**(10): p. 1447-1459.
12. Bomblies, A., J.-B. Duchemin, and E.A.B. Eltahir, *Hydrology of malaria: Model development and application to a Sahelian village.* Water Resources Research, 2008. **44**(12).
13. Asare, E. and L. Amekudzi, *Assessing Climate Driven Malaria Variability in Ghana Using a Regional Scale Dynamical Model.* 2017. **5**(1): p. 20.
14. Center of Disease Control and prevention. *Division of Parasitic Diseases and Malaria.* . 2015; Available from: https://www.cdc.gov/malaria/about/biology/mosquitoes/.
15. Briegel, H., *Fecundity, metabolism, and body size in Anopheles (Diptera: Culicidae), vectors of malaria.* J Med Entomol, 1990. **27**(5): p. 839-50.
16. Kaufmann, C. and H. Briegel, *Flight performance of the malaria vectors Anopheles gambiae and Anopheles atroparvus.* J Vector Ecol, 2004. **29**(1): p. 140-53.
17. Smith, N.R., et al., *Agent-based models of malaria transmission: a systematic review.* Malaria Journal, 2018. **17**(1): p. 299.
18. Linard, C., et al., *A multi-agent simulation to assess the risk of malaria re-emergence in southern France.* Ecological Modelling, 2009. **220**(2): p. 160-174.
19. Arifin, S.M.N., et al., *An agent-based model of the population dynamics of Anopheles gambiae.* Malaria Journal, 2014. **13**(1): p. 424.
20. Zhou, Y., et al., *An agent-based model of the Anopheles gambiae mosquito life cycle*. 2010. 201-208.
21. Gerardin, J., et al., *Effectiveness of reactive case detection for malaria elimination in three archetypical transmission settings: a modelling study.* Malaria Journal, 2017. **16**(1): p. 248.
22. Eckhoff, P.A., *A malaria transmission-directed model of mosquito life cycle and ecology.* Malaria Journal, 2011. **10**(1): p. 303.
23. Zhu, L., et al., *Is outdoor vector control needed for malaria elimination? An individual-based modelling study.* Malaria Journal, 2017. **16**(1): p. 266.
24. Zhu, L., et al., *A spatial individual-based model predicting a great impact of copious sugar sources and resting sites on survival of Anopheles gambiae and malaria parasite transmission.* Malaria Journal, 2015. **14**(1): p. 59.
25. Koenraadt, S., *Behaviour of Anopheles Gambiae*, L.v. Gorp, Editor. 2018.

26. Beck-Johnson, L.M., et al., *The Effect of Temperature on Anopheles Mosquito Population Dynamics and the Potential for Malaria Transmission.* PLOS ONE, 2013. **8**(11): p. e79276.
27. Gommes, R. *Global Climate maps: Tour Guide*. 2014; Available from: http://www.fao.org/nr/climpag/climate/index_en.asp.
28. Depinay, J.-M.O., et al., *A simulation model of African Anopheles ecology and population dynamics for the analysis of malaria transmission.* Malaria Journal, 2004. **3**(1): p. 29.
29. Yaro, A.S., et al., *The distribution of hatching time in Anopheles gambiae.* Malaria Journal, 2006. **5**(1): p. 19.
30. Schoolfield, R.M., P.J. Sharpe, and C.E. Magnuson, *Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory.* J Theor Biol, 1981. **88**(4): p. 719-31.
31. Beier, J.C., *Malaria parasite development in mosquitoes.* Annu Rev Entomol, 1998. **43**: p. 519-43.
32. Knols, B.G.J., et al., *MalariaSphere: A greenhouse-enclosed simulation of a natural Anopheles gambiae (Diptera: Culicidae) ecosystem in western Kenya.* Malaria Journal, 2002. **1**(1): p. 19.
33. Charlwood, J.D., et al., *'A mate or a meal' – Pre-gravid behaviour of female Anopheles gambiae from the islands of São Tomé and Príncipe, West Africa.* Malaria Journal, 2003. **2**: p. 9-9.
34. Charlwood, J., et al., *Survival and infection probabilities of anthropophagic Anophelines from an area of high prevalence of Plasmodium falciparum in humans*. Vol. 87. 1997.
35. Martens, P., *Health Impacts of Climate Change and Ozone Depletion: An Ecoepidemiologic Modeling Approach*. Vol. 106 Suppl 1. 1998. 241-51.
36. M Styer, L., et al., *Mosquitoes do senesce: Departure from the paradigm of constant mortality*. Vol. 76. 2007. 111-7.
37. Smith, D. and F. Ellis McKenzie, *Statics and dynamics of malaria infection in Anopheles mosquitoes*. Vol. 3. 2004. 13.
38. Okech, B.A., et al., *Influence of sugar availability and indoor microclimate on survival of Anopheles gambiae (Diptera: Culicidae) under semifield conditions in western Kenya.* J Med Entomol, 2003. **40**(5): p. 657-63.
39. Roitberg, B. and I. Gordon, *Does the Anopheles blood meal - Fecundity curve, curve?* Vol. 30. 2005. 83-6.
40. S Detinova, T. and M. T Gillies, *Observations on the Determination of the Age Composition and Epidemiological Importance of Populations of Anopheles gambiae Giles and Anopheles funestus Giles in Tanganyika*. Vol. 30. 1964. 23-8.
41. Hogg, J.C. and H. Hurd, *The effects of natural Plasmodium falciparum infection on the fecundity and mortality of Anopheles gambiae s.l. in north east Tanzania*. Vol. 114 (Pt 4). 1997. 325-31.
42. Lyimo, E.O. and W. Takken, *Effects of adult body size on fecundity and the pre-gravid rate of Anopheles gambiae females in Tanzania.* Med Vet Entomol, 1993. **7**(4): p. 328-32.
43. Webb, C.E., *Mosquito Ecology: Field Sampling Methods.* Australian Journal of Entomology, 2008. **47**(4): p. 382-383.
44. Turell, M.J., et al., *Seasonal distribution, biology, and human attraction patterns of mosquitoes (Diptera: Culicidae) in a rural village and adjacent forested site near Iquitos, Peru.* J Med Entomol, 2008. **45**(6): p. 1165-72.
45. McCrae, A.W., *Oviposition by African malaria vector mosquitoes. I. Temporal activity patterns of caged, wild-caught, freshwater Anopheles gambiae Giles sensu lato.* Ann Trop Med Parasitol, 1983. **77**(6): p. 615-25.
46. Killeen, G.F., et al., *A simplified model for predicting malaria entomologic inoculation rates based on entomologic and parasitologic parameters relevant to control.* Am J Trop Med Hyg, 2000. **62**(5): p. 535-44.
47. Rubio-Palis, Y. and C. F Curtis, *Biting and resting behaviour of Anophelines in Western Venezuela and implications for control of malaria transmission*. Vol. 6. 1992. 325-34.
48. Rozendaal, J.A., *Biting and resting behavior of Anopheles darlingi in the Suriname rainforest.* J Am Mosq Control Assoc, 1989. **5**(3): p. 351-8.
49. Gillies, M.T., *Studies on the dispersion and survival of Anopheles gambiae Giles in East Africa, by means of marking and release experiments.* Bulletin of Entomological Research, 1961. **52**(1): p. 99-127.
50. Midega, J.T., et al., *Estimating dispersal and survival of Anopheles gambiae and Anopheles funestus along the Kenyan coast by using mark-release-recapture methods.* J Med Entomol, 2007. **44**(6): p. 923-9.
51. Costantini, C., et al., *Density, survival and dispersal of Anopheles gambiae complex mosquitoes in a West African Sudan savanna village*. Vol. 10. 1996. 203-19.

52. Thomas, C.J., D.E. Cross, and C. Bøgh, *Landscape Movements of Anopheles gambiae Malaria Vector Mosquitoes in Rural Gambia.* PLOS ONE, 2013. **8**(7): p. e68679.
53. Bargielowski, I., et al., *Flight Performance and Teneral Energy Reserves of Two Genetically-Modified and One Wild-Type Strain of the Yellow Fever Mosquito Aedes aegypti.* Vector Borne and Zoonotic Diseases, 2012. **12**(12): p. 1053-1058.
54. Service, M.W., *Mosquito (Diptera: Culicidae) dispersal--the long and short of it.* J Med Entomol, 1997. **34**(6): p. 579-88.
55. Bousema, T. and C. Drakeley, *Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination.* Clin Microbiol Rev, 2011. **24**(2): p. 377-410.
56. T. Gillies, M., *The role of carbon dioxide in host-finding by mosquitoes (Diptera: Culicidae): a review*. Vol. 70. 1980. 525-532.
57. Takken, W. and B.G.J. Knols, *ODOR-MEDIATED BEHAVIOR OF AFROTROPICAL MALARIA MOSQUITOES.* Annual Review of Entomology, 1999. **44**(1): p. 131-157.
58. Kettle, D.S., *Medical and Veterinary Entomology*. 1984, Beckenham: Croom Helm Ltd. vi + 658pp.
59. Mohammed, A. and D. Chadee, *Effects of different temperature regimens on the development of Aedes aegypti (L.) (Diptera: Culicidae) mosquitoes*. Vol. 119. 2011. 38-43.
60. Gu, W., et al., *Low recovery rates stabilize malaria endemicity in areas of low transmission in coastal Kenya*. Vol. 86. 2003. 71-81.
61. Smith, D., et al., *The entomological inoculation rate and Plasmodium falciparum infection in African children*. Vol. 438. 2005. 492-5.
62. Lines, J., T. Wilkes, and E.O.K. Lyimo, *Human malaria infectiousness measured by age-specific sporozoite rates in Anopheles gambiae in Tanzania*. Vol. 102 Pt 2. 1991. 167-77.
63. Bharti, A.R., et al., *Experimental infection of the neotropical malaria vector Anopheles darlingi by human patient-derived Plasmodium vivax in the Peruvian Amazon.* Am J Trop Med Hyg, 2006. **75**(4): p. 610-6.
64. Reeves, W.C., *AGE-GROUPING METHODS IN DIPTERA OF MEDICAL IMPORTANCE: With Special Reference to Some Vectors of Malaria.* California Medicine, 1963. **98**(3): p. 185-185.
65. Craig, M.H., R.W. Snow, and D. le Sueur, *A climate-based distribution model of malaria transmission in sub-Saharan Africa.* Parasitol Today, 1999. **15**(3): p. 105-11.
66. Githeko, A.K., et al., *The reservoir of Plasmodium falciparum malaria in a holoendemic area of western Kenya.* Trans R Soc Trop Med Hyg, 1992. **86**(4): p. 355-8.
67. World Health Organization, *Chikungunya.* 2016.
68. Hoshen, M.B. and A.P. Morse, *A weather-driven model of malaria transmission.* Malaria Journal, 2004. **3**(1): p. 32.
69. Bleijs, D., *Chikungunya signs clinical symptoms.* . 2016.
70. Loewenberg, S., *Niger welcomes largest bednet distribution in history.* Lancet, 2006. **367**(9521): p. 1473.
71. Jamet, H. and C. Curtis, *Mosquito Behavior and Vector Control*. Vol. 50. 2005. 53-70.
72. GiveWell, *Mass distribution of long-lasting insecticide-treated nets (LLINs).* 2016.
73. HiCare, *Make your home safe today. expert mosquito control service.* 2016.
74. Tucker, C.J., J.R.G. Townshend, and T.E. Goff, *African Land-Cover Classification Using Satellite Data.* 1985. **227**(4685): p. 369-375.
75. ChampionTraveler. *The Best Time to Visit Rusinga Island, Kenya for Weather, Safety, & Tourism*. 2018; Available from: https://championtraveler.com/dates/best-time-to-visit-rusinga-island-ke/.
76. Holiday Weather. *Mombasa, Kenya: Annual Weather Averages*. 2018; Available from: https://www.holiday-weather.com/mombasa/averages/.
77. World Weather & Climate Information. *CLIMATE: AVERAGE MONTHLY WEATHER IN Kisumu, Kenya*. 2016; Available from: https://weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,kisumu-ke,Kenya.
78. Abiodun, G.J., et al., *Modelling the influence of temperature and rainfall on the population dynamics of Anopheles arabiensis.* Malaria Journal, 2016. **15**(1): p. 364.
79. Khatri, K.L. and R.J. Smith, *Evaluation of methods for determining infiltration parameters from irrigation advance data.* Irrigation and Drainage, 2005. **54**(4): p. 467-482.

80. Engineering ToolBox. *Humidity Ratio of Air.* 2004; Available from: https://www.engineeringtoolbox.com/humidity-ratio-air-d_686.html.
81. Service, M.W., *Effects of wind on the behaviour and distribution of mosquitoes and blackflies.* International Journal of Biometeorology, 1980. **24**(4): p. 347-353.
82. Chilaka, N., E. Perkins, and F. Tripet, *Visual and olfactory associative learning in the malaria vector Anopheles gambiae sensu stricto.* Malaria Journal, 2012. **11**(1): p. 27.
83. CHASE, J.M. and R.S. SHULMAN, *Wetland isolation facilitates larval mosquito density through the reduction of predators.* Ecological Entomology, 2009. **34**(6): p. 741-747.
84. Lyimo, E.O., W. Takken, and J.C. Koella, *Effect of rearing temperature and larval density on larval survival, age at pupation and adult size of Anopheles gambiae.* Entomologia Experimentalis et Applicata, 1992. **63**(3): p. 265-271.
85. Paaijmans, K.P., et al., *Unexpected High Losses of Anopheles gambiae Larvae Due to Rainfall.* PLOS ONE, 2007. **2**(11): p. e1146.
86. Browning, R.C., et al., *Effects of obesity and sex on the energetic cost and preferred speed of walking.* Journal of Applied Physiology, 2006. **100**(2): p. 390-398.
87. Lindsay, S.W., et al., *A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of The Gambia, west Africa. 3. Entomological characteristics of the study area.* Trans R Soc Trop Med Hyg, 1993. **87 Suppl 2**: p. 19-23.
88. Crooks, A., C. Castle, and M. Batty, *Key challenges in agent-based modelling for geo-spatial simulation.* Computers, Environment and Urban Systems, 2008. **32**(6): p. 417-430.
89. Spitzen, J., R.C. Smallegange, and W. Takken, *Effect of human odours and positioning of CO2 release point on trap catches of the malaria mosquito Anopheles gambiae sensu stricto in an olfactometer.* Physiological Entomology, 2008. **33**(2): p. 116-122.
90. Mahande, A., et al., *Feeding and resting behaviour of malaria vector, Anopheles arabiensis with reference to zooprophylaxis.* Malaria Journal, 2007. **6**(1): p. 100.
91. Diatta, M., et al., *Similar feeding preferences of Anopheles gambiae and A. arabiensis in Senegal.* Transactions of The Royal Society of Tropical Medicine and Hygiene, 1998. **92**(3): p. 270-272.
92. Grimm, V., et al., *The ODD protocol: A review and first update.* Ecological Modelling, 2010. **221**(23): p. 2760-2768.
93. de Barros, F.S., N.A. Honorio, and M.E. Arruda, *Survivorship of Anopheles darlingi (Diptera: Culicidae) in relation with malaria incidence in the Brazilian Amazon.* PLoS One, 2011. **6**(8): p. e22388.
94. Brady, O., et al., *Modelling adult Aedes aegypti and Aedes albopictus survival at different temperatures in laboratory and field settings*. Vol. 6. 2013. 351.
95. Jetten, T.H. and W. Takken, *Anophelism Without Malaria in Europe. A Review of the Ecology and Distribution of the Genus Anopheles in Europe*. Vol. 94. 1994.
96. Craig, M., R. Snow, and D. le Sueur, *A Climate-based Distribution Model of Malaria Transmission in Sub-Saharan Africa*. Vol. 15. 1999. 105-11.
97. Ponçon, N., et al., *A quantitative risk assessment approach for mosquito-borne diseases: malaria re-emergence in southern France.* Malaria Journal, 2008. **7**(1): p. 147.
98. Center of Disease Control and prevention, *Dengue and the Aedes aegypti mosquito*. 2016.
99. Bleijs, D., *Aedes aegypti*. 2016.
100. Gary, R.E., Jr. and W.A. Foster, *Diel timing and frequency of sugar feeding in the mosquito Anopheles gambiae, depending on sex, gonotrophic state and resource availability.* Med Vet Entomol, 2006. **20**(3): p. 308-16.
101. Price, R.N., et al., *Vivax malaria: neglected and not benign.* Am J Trop Med Hyg, 2007. **77**(6 Suppl): p. 79-87.
102. Takken, W., et al., *Dispersal and survival of Anopheles funestus and A. gambiae s.l. (Diptera: Culicidae) during the rainy season in southeast Tanzania*. Vol. 88. 1998.
103. Kabaria, C.W., et al., *The impact of urbanization and population density on childhood Plasmodium falciparum parasite prevalence rates in Africa.* Malaria Journal, 2017. **16**(1): p. 49.

Annex 1: ODD Protocol

The methods for this research follows the review of Overview, Design concepts and Details (ODD) protocol developed to standardize ABM descriptions [92].

1. Purpose

The purpose of this model is to be used or further developed as a tool for representing the malaria spread, with a flexible area input in order to assist determining the values of spatial independent input variables.

2. Entities, state variables, and scales

As described in chapter 2 this study differentiates between seven different categories of input variables and only three of these have spatial independent values, namely: Mosquito life cycle, mosquito day cycle and malaria. In table 18 all input variables that are included as stated in chapter 2 are given an overview as well as some assumptions that have not yet been discussed.

For the development input variables, Arifin et al.'s [19] method is chosen and for the mortality Martens [35] method is used. This development method goes best with the real situation as measured by Depinay et al. [28], as was discussed in chapter 2.1. The mortality equation is used as a standard for different Anopheles mosquitoes, which makes the equation flexible and thus could also be used for Anopheles Funestus or Anopheles Arabiensis for example. Although there is a difference between mortality for different life phases, it is assumed to be the same as the difference is insignificant.

The fecundity, active period and flight distance is generally agreed upon by the developers and thus their values will be used for these input variables. Although the fecundity and the flight distance do not necessarily have a linear range, it is not specified anywhere what kind of range it would in reality be. For this model it is assumed that it will have a linear range.

Table 18: Input variables used for the model. The purple represents mosquito input variables and yellow malaria input variables. All temperature (T) values in this table are in ºC.

Input variable	Value	Source
Development as an egg	% development / hour = 1 / (-0.9*T + 61)	[29]
Development as a larva	% development / hour = T*0.000305–0.003285	[30]
Development as a pupa	% development / hour = 1 / (-0.9*T + 61)	[29]
Development as a young adult	% development / hour = 1 / (-2.67*T + 120)	[28]
Mortality mosquito	% chance/day = 1 - $e^{-1 / (-4.4 + 1.31 * T - 0.03 * T^2)}$ and maximum 30 days	[35]
Mortality of adolescent mosquitoes	Same as mortality mosquito	
Eggs per oviposition (fecundity)	80-100 female eggs	[39, 41, 42]
Active period	6 pm to 6 am	[45]
Flight distance	200-400 m/day	[50, 55]
Detection range	15 m	[25, 56]
Transmission (mosquito to human)	50 % chance / bite	[60, 61]
Transmission (human to mosquito)	17.5 % chance / bite	[60, 62]
External incubation	111 degree days above 18 degrees and Arrhenius method	[64, 65]
Intrinsic incubation	Arrhenius method	[65]
Time until recovered	0.04167-0.05 % development / hour	[37, 68]
Human movement speed	1.4 m/s	[86]
Human movement time	At 6-10 am until 18-22 pm	Assumed
Safety when prevention methods are installed	90% chance	Assumed

The detection range is considered 15m as this was the only value that was referenced and was realistic to implement. Bomblies et al. [12] suggest a relation with the wind, but this would also include calculations for the exhalation rate, the emission distribution and the height of the people and the mosquito. This is a complex addition to the model for what seems to be a small contributor to the end result. Therefore, for the sake of time it will be simplified to a single number.

The transmission and incubation times seem to be generally agreed upon and therefore these methods will be used. The human to mosquito transmission was minor disagreement in, so it is assumed that both values might be valid. Therefore, an average was taken of 17.5%. The external incubation rates between the Arrhenius and the degree days methods are similar, with the exception that the degree days show more extreme values. Both methods will be tested to test its viability. There was no mention of incubation taking place during the aquatic phase, therefore incubation will not take place for infected adolescent mosquitoes. The intrinsic incubation will use Arrhenius method as they are the only one who gave a clear range and formula for their incubation times.

The expected time needed for recovery would be the 1-1.2% development per day, however as this model will work on an hourly rate, this value is also put in per hour.

It is uncertain how humans would work and this usually differences entirely per person and per area. However, as decisions have to be made on how humans act, it is expected that they will walk an average 1.4 m/s at all times and are active between 6-10 am until 17-21 pm. As these numbers are arbitrary, they can be changed based on field observations. It is also expected that the prevention methods will always exceed, except for 10% of the people who decide to leave their protection during the night for whatever reason.

As the assumed values created for the humans are dependent on the culture and their awareness of the mosquitoes, they can be treated as initialization values and should be changed for every study area if the values can be assessed.

3. Process overview and scheduling

This model’s functions and processes work are updated each hour. This includes mosquito and human activity, mosquito mortality, mosquito development, incubation times, recovery time, water drainage and temperature updates. Mosquito and human activity in this case refers to movement and the actions in relation to their current status. For example, gravid mosquitoes laying eggs.

It is constructed in GAMA due to their good GIS integration and design for agent-based modelling. It is also a language often used in modelling, making it easy to adapt for other modellers.

In figure 16 an overview is given to these processes, which are broken down into three main agents: human agents, aquatic agents and mosquito agents. The human agents move between work and home at two points during the day, which are chosen randomly as discussed in the previous chapter. During any moment of a full day, the SEIRS cycle will progress and the humans might leave their protected areas. At any point while they are inside the building, they might open the water container present. The aquatic agents do not move in the model and will remain in the same place. The only actions aquatic agents can take is to develop to the next phase of their life, or die due to random chance or due to a dried-out habitat.

The mosquito agents are created with a numeric input and randomly spread across all water in the study area. The initial number of adolescent mosquitoes is assumed to be the same as the initial adult mosquitoes spread across the egg, larva and pupa phases. Once the adult mosquitoes are initialized or developed, they start off in the hungry phase in which they will search for blood means using random movement during the night. They will either feed off humans during the day, or with high cattle intensity, they will feed on cattle instead. They may also feed during daylight if the humans are within detection range of their habitat. After having fed, they will move to the resting phase in which they will do nothing for a day. Then the next day they are ready to lay eggs and thus they will search for a new oviposition in either a water body, a water container or a pool created by the rain. They will prioritize water bodies as ovipositions as they are available consistently, then areas around forests or buildings as they block wind flow and then any other area with water. This resting, gravid and hungry cycle will switch every day. During the whole process, the mosquito might die at any point due to random chance. Once an adult mosquito becomes infected, it will first go through external incubation before the disease will transmit during their hungry phase. As the mosquitoes can only be susceptible, exposed or infectious this will be considered a SEI cycle.

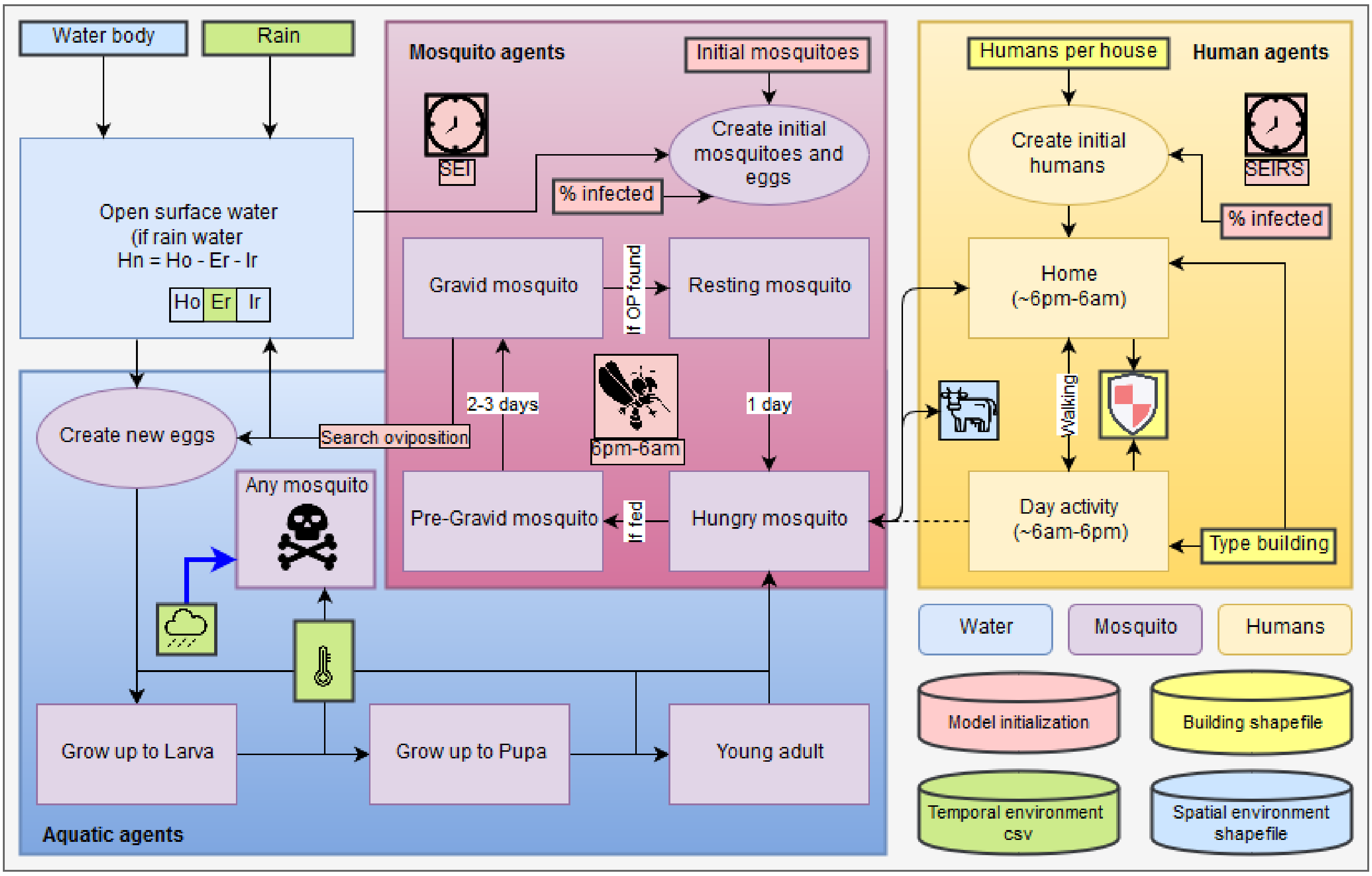

Figure 16: The model inputs and interactions. The colours used within the model refer to the colours shown in the bottom right. The rectangles with thick black borders represent inputs. When a singular process is not attached to any other process (for example SEIRS) it counts for the entire background box. DR stands for detection range.

Water is dynamic in this model, as the rain affects whether new ovipositions would be created. Waterbodies created by rain are assumed to be homogenously spread across a spatial environment polygon and for each a calculation is being done as is shown in the "open surface water" process. The formula shown there is as follows:

$$Hn = Ho - Er - Ir \qquad (6)$$

In which Hn is the new water height in mm, Ho is the old water height in mm, Er is the evaporation rate in mm/h and Ir is the infiltration rate in mm/h. Human construction as well as trash can collect water which is represented by the container rate. This container rate implies that the infiltration rate would not be effective for parts of the area and therefore is reduced by a higher container rate.

This model will be run a hundred times for each simulation according to the Monte Carlo method. This will be run different times to determine the different stochastic values and the different uncertain input variables (table 19). The stochastic input variables will be tested for their influence on the variance and the averages of the malaria spread outputs. The uncertain input variables will be tested for their accuracy. These will have to be compared to a ground truth. For this a study area will be chosen for comparison. The stochastic simulations will suffice with one Monte Carlo run per input variable plus a run where all these input variables are as the test values. The uncertain input variables will have to be tested for different possible values and therefore the total amount of runs for this is currently unknown.

Table 19: Testing values for stochastic input variables and description why some uncertain input variables are being tested.

Stochastic input variable	**Test**	**Default**
Fecundity	90 female eggs	80-100 female eggs
Flight distance	300 m/day	200-400 m/day
Time until recovered	0.045835 % development / hour	0.04167-0.05 % development / hour
Human active time	At 8 am until 20 pm	At 6-10 am until 18-22 pm
Uncertain input variable	**Description**	
Detection range	This input variable is influential to almost all human and mosquito interactions as well as the procreation of the mosquitoes. Therefore, determining the point where it is certain this value is correct is of high importance.	

4. Design concepts

While constructing the model, there were several principles kept in mind in order to reach the goal of this research.

4.1. Basic principles

The model is built upon the foundations set by already existing models as well as implementing parts using testing. The inspiration of existing models is for the most part based on publicly available model created by Jindal and Rao [10] and the general model structures summarized by Smith et al. [17]. There were also many parts that were personally developed. The main criteria for own developed parts are that:

1. It would not require inputs that were not generally available;
2. The required inputs are adaptable to different available data;
3. Only processes and attributes that are likely accurate or are highly relevant are included in the final model;
4. All assumed values with high relevancy will be determined using calibration;
5. All processes that can be spatially determined and that do not contradict the previous development principles, has been put into the model as a spatial component.

With the principle for the inputs to be generally available it is meant that the necessary data should be accessible online or can be determined by own means or assumption. For example, data might be available on where people live and whom is protected.

The second principle ties in closely to the first principle, but this describes the input contents. As such the evaporation rate could have been broken into wind speed, humidity and temperature to be calculated inside the model. However, as wind is a complex input the evaporation rate is set as the input only. If the data would be available, this can still be put into the input but will not be processed by the model itself.

The principle about relevancy and correctness reassures no attributes or processes are made up unless a likely correct guess can be made or that it has a known high impact on malaria spread. Other than speeding up the model, this also reassures that everything inside the model is only built on known processes and interactions. For example, it is known that mosquitoes remember their blood meals up to three days, which has an impact on who they will feed on next time they are hungry. However, what impact it has and how it affects mosquito behaviour is not clearly defined. Therefore, this process is not included into the final model.

The third principle can also be applied for attributes such as the detection range, which can be determined using either a complex formula or an estimation. As the complex formula requires inputs that cannot be considered generally available, an estimation had to be made on what this value is. This therefore also ties in with the fourth principle which states that relevant, yet assumed attributes have to be calibrated to the area unless a trend can be established.

The final principle reassures that this model stays spatial as was intended from the goal.

4.2. Emergence

The main output of the model is the malaria spread across the area, specifically what number of people gets infected and which areas in particular get infected. This can be affected by multiple input variables, as this model is built on multiple stochastic values (see table 6). Currently it is uncertain which input variables would be most defining towards the end result as determining this is part of the objective. However, it is expected that especially detection range will play a significant role in the results of this model as it has a significant impact on the mosquito movement behaviour.

4.3. Adaptation and objective

The changing components are separated in spatial and temporal components. Spatial components include the position of blood meals, mosquitoes and ovipositions. These impact the mosquito behaviour as the mosquitoes have limited detection range and this is how the mosquito will adapt to their situation. Temporal components are for the most part weather. Different temperatures influence mosquito development and mortality. Precipitation and evaporation rate have an effect on the available surface water. This new water makes the mosquito interested in residing there.

The main adaptation for the mosquito is that it will not have the same habitat as the mosquito would lay eggs in. As such the mosquito will always try to find a new oviposition to rest or lay eggs into.

Although there are more changing components to which the mosquito can adapt, it will not necessarily recognize a worse or better target. It will simply take the one that is the easiest to reach. This as the mosquito movement per day is limited.

4.4. Learning

Mosquitoes remember the location of their blood meal and search for new potential blood meals within that area, but it is not expected that this will have a relevant impact on the model results. Especially as the mosquitoes usually stay within the area of the humans that they have bitten. The memory of mosquitoes is also a study that is still in development and the actual impacts memory would have on malaria spread is not yet established.

4.5. Prediction

The prediction is limited in this model as it is assumed that the mosquitoes will not be able to remember paths or patterns and only recognize blood meals or objects once they are within detection range.

4.6. Sensing

Mosquitoes will sense their targets using the detection range and will stay within that range to buildings and vegetation. They also fly to humans once they are in the hungry state or to oviposition once becoming gravid. Humans are expected to not have any senses that influence the malaria spread.

4.7. Interaction

The mosquitoes will not interact with each other as in this model they will only be present as female. There is interaction between the different agents in the form of mosquitoes laying the eggs to create adolescent mosquitoes and mosquitoes biting humans.

4.8. Stochasticity

The stochasticity is described by the input variables in table 17. The value ranges given there are linear in all situations and this randomness could impact the end result. In total there are five stochastic values and four values are uncertain. These values need to be adapted if necessary.

Other than the given values, there are still some other uncertainties present bound to the modelling process itself. As such the initial agent locations are random inside the appropriate spatial environments and the initial percentage infected are spread randomly amongst the agents. The mosquitoes also perform a random movement until they recognize vegetation, buildings or people in which their movement will be limited. This as they prefer areas that are out of the wind and close to blood meals.

4.10. Collectives

The only collective attribute is the maximum eggs per oviposition. All other input variables do not cause restrictions in each other's behaviour.

4.11. Observation

In the end the model will investigate the malaria spread over multiple runs to investigate its accuracy. The results will include how many people are infected and where they are located.

5. Initialization

The model initialization is for the most part defined in table 20, which describes the input variables as if the mosquito in question is the Anopheles Gambiae and the malaria parasite Plasmodium Falciparum. There are some other model initializations which are shown in table 21. These are area specific input variables and have to be adapted for each new case study.

Table 20: Initial values for the input variables inside the malaria model.

Input variable	**Value**	**Description**
Number of adult mosquitoes	10000	Arbitrary value
Aquatic to adult ratio	2	How many adolescent mosquitoes exist per initial adult mosquito
Percent infected mosquitoes	100%	How many mosquitoes are infected by malaria initially
Percent infected humans	30%	How many humans are infected by malaria initially

6. Input data

One of the main parts of this model is that it would require as little inputs as possible. For this reason, only three inputs are required: buildings, spatial environment and temporal environment. The buildings contain data on how many people are present inside the building and what type of building it is as well as how many water containers would be present. These data are necessary for the behaviour of humans as well as potential ovipositions for the mosquitoes.

The spatial environment is a dataset which separates each area where one of the values would be different from the one next to it. The input variables are the infiltration rate, the cattle to human ratio, the potential water containers and the water body. The water body is there to simplify land and water by putting them into one shapefile. The potential water containers refer to waste, open barrels, buckets and anything else that could potentially hold water. This will assist in finding out how long rain will stay after it has fallen. For the same reason infiltration rate is included. This varies mostly on the type of soil and the activities that partake on the soil.

Although multispectral data was included for the majority of reviewed ABM's for determining the environment, for this model this will not be used directly. Setting this is considered optional for determining the spatial environment and building positions for its input.

Temporal environment is the only csv dataset that includes weather across the whole study area. This dataset includes min and max temperature, the evaporation rate and the precipitation. The temperature will determine the development and mortality of mosquitoes. The evaporation rate is important for determining the stay of water after it has rained. This value can be calculated using the wind, temperature and humidity or be researched to fill in directly. The precipitation is of course important to determine how much water will fall that could potentially act as new ovipositions.

Table 21: Input data for the malaria model.

Input file	**With**	**For**
Buildings (Shapefile) (Human species)	Amount of water containers (int)	Determining water created and potential ovipositions
	Number of people (int)	Creating the right amount of people per building
	Type of building (int)	Determining whether it will be a location for the day or the night
	Protected (Boolean)	Determining whether humans are safe when inside the building
Spatial environment (Shapefile) (Land species)	Area in square meters (float)	Determining how much water can be created during precipitation
	Infiltration rate in mm/d (float)	Calculating the temporary stay of water from rain
	Cattle to human ratio (float)	Calculating the chance of the mosquito stinging a human instead
	Water body (Boolean)	Whether it is a water body by default. If so, all other values are 0 except for area.
	Forest (Boolean)	Whether it is an area where wind is being blocked.
Temporal environment (csv) (global values)	Day or max temperature in ^{0}C (float)	Calculating the development and mortality of mosquitoes
	Night or min temperature in ^{0}C (float)	Calculating the development and mortality of mosquitoes
	Evaporation rate in mm/day (float)	Calculating the temporary stay of water from rain
	Precipitation in mm (float)	Determining the amount of potentially create ovipositions from rainfall

7. Sub models

The known sub models for the malaria model are the adolescent mosquito development, the mosquito mortality, rainwater depletion and the incubation rates. These equations are implemented the same as they were stated in the literature chapter, except some minor functionality differences in GAMA. As such e^x is given as exp() instead.

There is no clear established relation between cattle and humans. However, there have been measurements made on how many humans are stung with a certain amount of cattle in that area. Figure 17 shows these measurements as made by Lindsay et al. [87]. Human blood index refers to the chance a human gets stung.

This figure shows the relation between the chance to be stung more clearly and a predicted trend can be made from this. The linear predicted trend has an R^2 of 0.87, which should make it accurate to follow. However, it shows two problems: Cattle have an almost total preference over 1.625 cattle per person and close to 0 cattle per person the cattle would still have some relevance. As humans are always the preferred target, a minimum chance for humans to be stung is set. Also, when the human blood index reaches above 75%, which is at 0.25 cattle per human, every 0.01 less cattle equals 1% increased chance for a human to be bitten. This results in the following equation:

$$
\begin{aligned}
&HBI = (-40\% * R_{cth} + 85\%) && \text{if } R_{cth} <= 1.625 \\
&HBI = 20\% && \text{if } R_{cth} > 1.625 \quad (6) \\
&HBI = 100\% - R_{cth} && \text{if } R_{cth} < 0.25
\end{aligned}
$$

HBI is the human blood index and R_{cth} the cattle to human ratio.

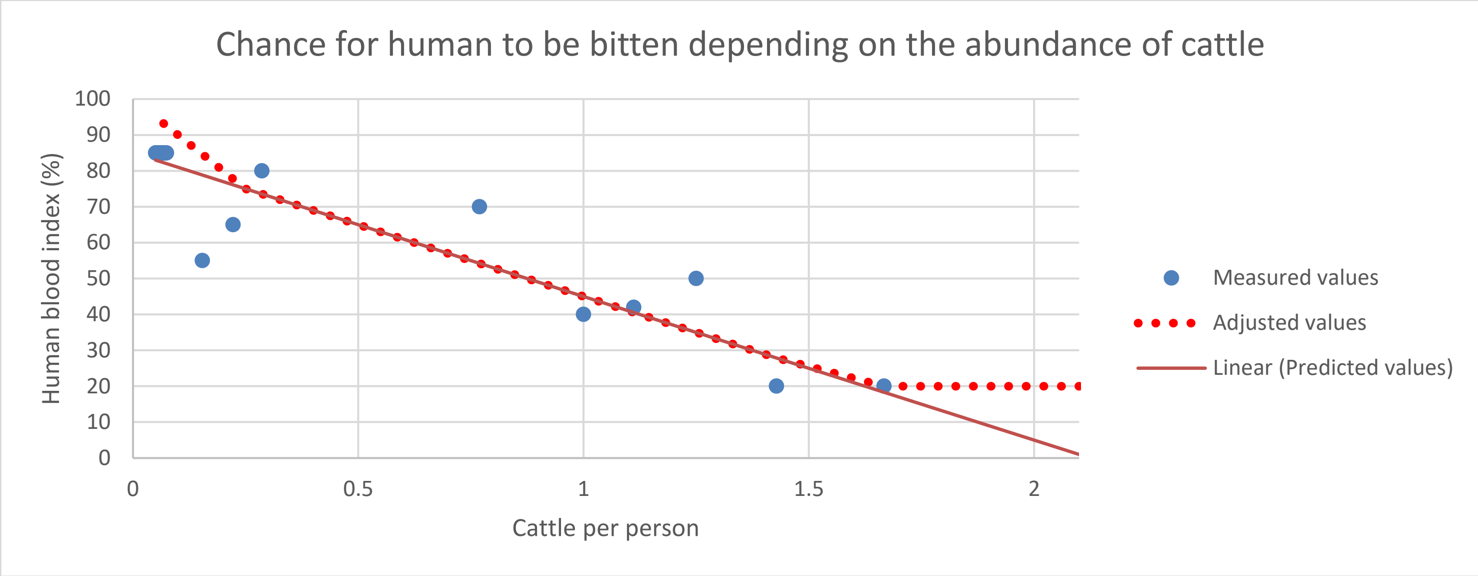

Figure 17: Chance for human to be bitten against cattle abundance. Including measured, predicted and adjusted values [87].

Annex 2: Attributes found in spatial ABM’s, their occurrence and their values

Red colors indicate the values are excluded from the model, green indicates that these are testing input variables and orange means that it is included but not stated. t% stands for the percent of the total whereas p% stands for percent probability. [S] stands for self, meaning that the value is determined by the model developers themselves most often as an assumption.

Input variable	Gu [7]	Bomblies [8, 12]	Arifin [19, 20]	Gerardin [21, 22]	Zhu [23, 24]	Pizzitutti [9]	Jindal [10]	Linard [18]
Mosquitoes life cycle	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	***Anopheles Darlingi***	***Aedes Aegypti***	***Anopheles Hyrcanus***
Development as an egg	30 p%/day [S]	~1 day [28]	hours = -0.9*T + 61 [29]		5 t% of fecundity survival and 12 days (includes other phases) [32]	*15 days (including larvae and pupae phases) [93]*	*maturation (days) = 8 + | T − 25| (includes larvae and pupae phases) [94]*	
Development as a larva	20 p%/day [S]	T ~ days (7.5 - 30 days) [28]	t% per hour = T * 0.000305–0.003285 [30]	~10 days at 25 ^{0}C using Arrhenius equation (includes eggs and pupae) [31]				*t%/d = 0.021*(EXP(0.162*(B64-10))-EXP(0.162*(35-B64)-((35-B64)/5.007))) (includes eggs and pupae) [95]*
Development as a pupa	30 p%/day [S]	~1 day [28]	As stated for egg					
Time before first blood meal	2 days [33]	10 days [28]	hours = -2.67 * T + 120 [28]	4 days [S]	2-3 days [33]			
Mortality mosquito	20 p%/day [34]	p%/day = 1 - exp(-(1/(-4.4 + 1.31 * T - 0.03 * T^2))) [35]	10 p%/day [S] and dead after 10 days [36]	Dead after 10 days [37]	Max 30 days [38]	*p%/day = 1 - exp(-(1/(-4.4 + 1.31 * T - 0.03 * T^2))) [96]*	*5-20 p%/day [52]*	*21 p%/day [97]*
Eggs per oviposition (fecundity)	80 female eggs [39]	150 female eggs [40]	Mean = 170 stdv = 30 [41]	80-100 female eggs [S]	100 [42]	*7-13.5 per OP [S]*	*100 [98]*	*100-350 [95]*
Reproduction success rate					100 p% [S]		*20 p% [S]*	
Includes male mosquitoes			Yes, but only as mate seeking [S]	Yes, to determine genetic spread [S]				
Mosquito day cycle	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	***Anopheles Darlingi***	***Aedes Aegypti***	***Anopheles Hyrcanus***
Active period		Night [43]	6pm to 6am Night [45]	Night [S]	7 pm to 5 am [S]	*6pm to 6am [44]*	*6am to 7pm [99]*	*Night [S]*
Resting		24 hours [S]		3 days [46]	During daytime [47]	*4 hours [48]*		
Flight speed	400 m / day [49]	15 m / hour [51]					*0-1 km/h [53]*	*As flight distance*
Flight distance	200-400 m [50]			90% within 370m, 97% within 500m, other up to 3km [50]	200-400m [55]	*m = 258-266m, stdv = 128-230m, max = 1944-2704m [S]*	*350m [52]*	*Max 2100m / 8 hours [54]*
Detection range (other)[1]	40m[49]	15m [56]			5m [S]		*3m [S]*	
Detection range (CO2)[1]		> 0.01 t% above background with complex formula [57]						*Max 20m; same cell [58]*
Max meals per day				Rest happens after blood meal	Minimum of 2 [100]		*1 [S]*	*PBRx=adultsMovex; ABR = PBR/H; ABRt = sum ABR [S]*
Successful bite		80 p%/bite [S]	25% to find blood meal [S]		100 p% [S]	*random p% [S]*		*As max meals*
Malaria	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	**Anopheles Gambiae**	***Anopheles Darlingi***	**Aedes Aegypti**	**Anopheles Hyrcanus**
Transmission (mosquito to human)	50 p%/bite [60]			50 p%/bite [61]		*100 p%/bite [S]*	*60 p%/bite [59]*	
Transmission (human to mosquito)	15 p%/bite [60]			20 t% mosquitoes assumed infected by default [S]	20 p%/bite [62]	*40 p%/bite [63]*	*27.5 p%/bite [59]*	
External incubation	10 days [62]	111 degree days above 18^{0}C [64]		Using Arrhenius [96]	10 days [66]	*111 degree days above 16 ^{0}C [64]*		
Intrinsic incubation	15 days [S]	Random days [S]		Using Arrhenius [96]		*9 - 14 days [101]*		
Susceptible	Any	Any	Any	Any	Any	Any	Non-recovered [67]	Any
Exposed	As incubation	As incubation				As incubation	2-6 days [69]	
Infectious	Until recovered	336 h [68]				After incubation	4-7 days [69]	
Recovered	1 t%/day [37]	0.05 t% / h [68]					5 p%/agent [69]	
Includes cattle and/or domestic animals		Mostly human, but cows and others can attract if in high abundance [87]				TC: 42 p%, PC: 88 p% [S]		Yes
Prevention methods								
Insecticide treated nets (ITN)		20 t% safe [70]	Testing different t% coverages	Testing different t% coverages	Testing different days of coverage	"A high fraction" [71]	Random distribution based on cost [72]	70p% local, 50p% hotel, 30p% camping (for all methods) [S]
Indoor Residual Spraying (IRS)			Testing different t% coverages	Testing different t% coverages	Testing different days of coverage	"A high fraction" [71]	Random distribution based on cost [73]	As ITN values
Larval Source Management (LSM)	100, 200, 300m		Testing different t% coverages					As ITN values
Environment	Hypothetical	Niger	Hypothetical	Lake Kariba, Zambia	Hypothetical	Amazon	Caribbean	Camargue, France
Location water bodies	Randomized [102]	Multispectral images (Quickbird)	HC=HCBaseline*r [S]	Field study	Randomized	High res multispectral images	Sentinel satellite	Landsat 2001 and field data
Max eggs per habitat	5000 [S]	300 mg / m2 [28]	Biomass=NEggs+Ne+NPupae		400 [S]	0.30 MOSQ / m2 / 12h [S]		400.000 total/moment
Location bloodmeals / housing	Randomized [102]	Field data	Randomized	Field study	Randomized	Field and satellite	Sentinel satellite	Landsat 2001 and field data for housing and cattle
Risk areas				Notifies increased risk with increased population density [103]		Farming during day 100 p%	Going to the park as attribute	Activities carry different risks
Weather	Hypothetical	Niger	Hypothetical	Lake Kariba, Zambia	Hypothetical	Amazon	Caribbean	Camargue, France
Temperature		Meteo	Hourly	Meteo		Meteo	Meteo	14-28 Cº [Meteo]
Precipitation		Meteo	Seasonal	Meteo		Meteo	Meteo	
Wind		Meteo						
Model input variables								
Number of humans	By housing	10 per 100m2 village		8000, 12000, 25000	25 houses	1400	1000	200
Number of mosquitoes	20	Field data	25000 + 1*10^6 immature		400	7 per OP	2000 (+500 eggs)	0-40
Timestep		Hourly	Hourly		One second	Hourly	10 minutes	8 hours
Total time simulation run	200 days	June until November (182 days, "mosquito season")	1 year	2014 until 2021 (2921 days)	60 days	2 years	Jan until March (89 days, epidemic)	1 year
Size study area	> 2 x 2 km	2.5 x 2.5 km		112 km2	600 x 600m	2370 x 2520 m and 10252 x 4200 m		
Pixelsize	50 x 50 m	25 x 25 m			1 x 1m	10 x 10 m		30 x 30m
Amount of simulations	63	4	30 per experiment	1000	50	150		890
Programming language or system	C++		Java and C++	EMOD (JSON)	Java 7 (MASON)	MASON	GAMA	NetLogo and UML
Other attributes		Soil, weights of larvae, cannibalism, mosquito mortality in earlier stages, water utilization	Enthalpy and chance of failure eggs. Egg mortality etc			Death due to rain, human ages and demographics		Movement determined per time unit, gonotrophic cycle

[1] if this is empty it is assumed that feeding only happens when both agents are in the same cell.

t% means it’s a percentage of the total whereas p% means there is a probability something might occur.

Annex 3: Table of contents for the research folder

Table of Content of the DVD that accompanies the thesis report

- Report (both in word and pdf format)
- Presentations
  - Midterms
  - Final presentation
- Sources
  - My EndNote Library.Data (includes all endnote citations)
  - Non-imported sources (referenced documents that are not academic, for example WHO and UN documents)
- Datasets
  - Buildings (contains imported data from online and clipped to the study area)
  - Results (contains all resulting shapefiles from the result producing R script)
  - Spatial environment adaptations (mid and corner versions of the environment)
  - Study area (the area to which all other input shapefiles are clipped to)
  - Water (contained imported data from online and clipped to the study area)
- Scripts
  - GAMA
    - Includes (contains all inputs created in input_generation)
    - Models (contains script files and the output shapefiles of the model)
    - Weather (a year long of weather data on Rusinga)
  - R
    - Input_generation (which clips downloaded inputs and grants them random values)
    - Result_production (which takes all the shapefiles from the Monte Carlo runs and calculates the mean and SD as well as preparing other overviews)
    - Result_visualization (takes the data from result_production and puts this in maps that are pleasant to look at)
- Schematics
  - Contains all schemes that represent model processes in this report