From *Anopheles* to Spatial Surveillance: A Roadmap Through a Multidisciplinary Challenge

Valérie Obsomer, Nicolas Titeux,
Christelle Vancustem, Grégory Duveiller,
Jean-François Pekel, Steve Connor,
Pietro Ceccato and Marc Coosemans

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/55622

1. Introduction

Anopheles mosquito species are diverse and vector of many pathogens. A review of the genus *Anopheles* [1] recently updated [2] listed more than 520 species, some of which including subspecies and cryptic species. Each of them presents ecologic requirements and behaviours that can influence their status as vector for specific pathogens. Pathogen transmission dynamics vary greatly from one region to another such as documented for the biodiversity of malaria in the world [3]. Acknowledging these variations at local scale within a country through detailed mapping can lead to better targeted measures and improved monitoring. Interactions between vectors, pathogens and humans in a given area can be better comprehended using a spatial framework leading to what we call here a spatial surveillance.

Part of spatial variation is explained by differences in pathogen species or by successful control in some areas. Nevertheless, *Anopheles* species play a major part in the occurrence, seasonality and spatial variation of *Anopheles*-borne diseases. The environment in a given region provides or not support for a given species to breed, thrive and live long enough to be an efficient vector. Because of these variations a species might be an efficient vector in one settlement and then only a secondary vector in another. The need to clarify *Anopheles* distribution is recognised as a crucial step towards malaria eradication [4]. A recent effort to provide detailed maps of malaria and vectors has been carried out [5–8] including a description of ecological requirements. While these distribution maps are essential for an overview, some issues [9] (described further) linked to the data and modelling impede usage in an operational world. Modelling

and surveillance are key activities for successful control but integration of all components for effective operational surveillance is not straightforward.

In this chapter we review the challenges posed by spatial surveillance of *Anopheles*-borne diseases with particular attention for malaria surveillance. This challenge will mainly reside in the difficulty of getting the appropriate raw data and the large spectrum of multidisciplinary expertise. We propose here a roadmap from *Anopheles* sampling to a spatial surveillance.

2. Problem statement and application area

When working on vector-borne-diseases, decision makers and researchers often face a lack of specific quality data required for optimal targeting the intervention and surveillance. However, the results/decisions are critical as they impact on the lives of many people. Numerous studies use suboptimal vector dataset, and proxies for environmental drivers to map vector distribution or provide basis for vector surveillance. However, the uncertainties linked to the original dataset are not always well documented, in particular regarding proxies for environmental drivers derived from satellite imagery. Analysis methods also do not always take into account the specificity of species ecological distribution and inaccuracies linked to the vector dataset.

3. Research course and method used

Spatial surveillance of vector-borne diseases should integrate specialised knowledge in entomology, ecology, parasitology, epidemiology, human health, ecological modelling and social sciences. The authors of this chapter are specialized in those different fields and teamed up to offer an overview of the challenges posed by spatial surveillance. As vector-borne-diseases are linked to the environment, a spatial analysis using geographical information or remote sensing related technologies seems then appropriate.

4. Following the road map

Reliable outputs to go from *Anopheles* to spatial surveillance first depend on the data entering any analysis or decision process, being the data on *Anopheles* or on environmental factors. Environmental factors provided by remote sensing techniques that could be used to predict distribution or occurrence of malaria and *Anopheles* have already been reviewed [10]. Based on this inventory, we analyse pros and cons of *Anopheles* sampling strategies, various types of data and modelling techniques. Finally, useful initiatives to make research efforts available and operational in the field are discussed. The general scheme is provided in figure 1.
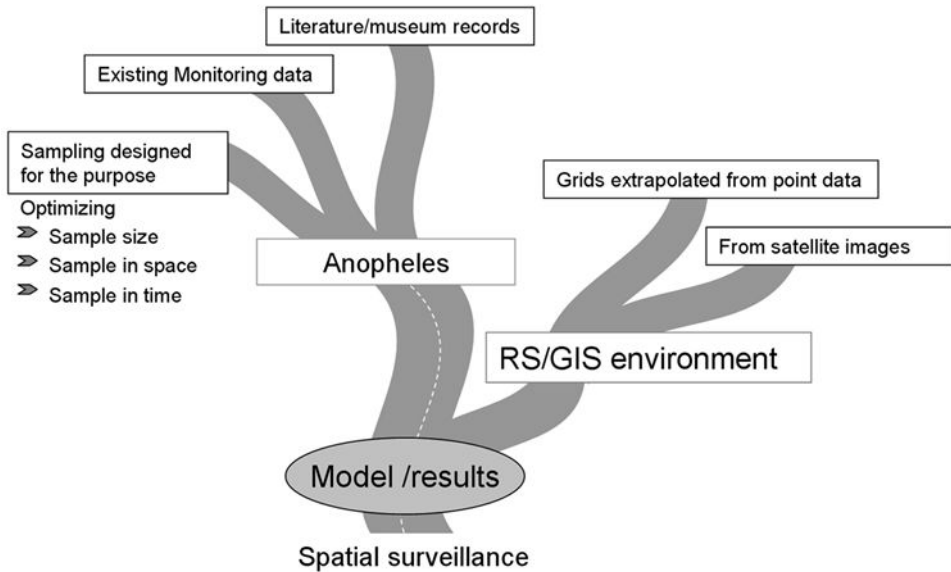
Figure 1. General scheme of the roadmap for *Anopheles* spatial surveillance

4.1. *Anopheles* and sampling strategies

Many useful attributes can be collected on *Anopheles* such as the list of species and their vector status in a given area, resistance to insecticides, behaviour influencing human vector contact, control effort avoidance (early biting, outdoor resting). Research and monitoring programmes might be based on existing entomological data whose particularities should be dealt with at the modelling step. However the most direct way is to design the collection protocol in relation to the objective, i.e. mapping the *Anopheles* species. In this last case, the quality of the dataset could be high if some rules are followed. Monitoring data are typically collected in a network of sampling locations according to a variety of standardized procedures [11] and used for mapping species distributions [12]. However, records are generally collected only in a restricted number of locations often loosely distributed across the region of interest, which is inconvenient for documenting species distribution.

Species distribution modelling techniques [13,14] provide assistance to achieve mapping based on monitoring data [15] such as detailed further in the road map. When coupling *Anopheles* monitoring and mapping efforts, defining an optimal sampling strategy becomes of highest interest. Indeed, well-designed monitoring projects have the potential to produce appropriate data to estimate changes in species attributes [16] but also document the distribution in space and time [12]. An appropriate sampling design should address key issues: what constitutes a sampling location? How many are needed? Where do they have to be located? How often to survey? When monitoring data are used to generate species distribution models, designing

the sampling strategy is a challenge because these issues are to be addressed relatively to the monitoring and the modelling objectives.

4.1.1. Optimizing sample size

Sampling locations may be sites, squares, transects or any spatial unit from which the measurements are made in the field to document attributes (e.g. presence, population density, infected/infective mosquitoes, reproductive status, insecticide resistance) that describe the *Anopheles* species. A sample is a set of sampling locations where attributes of the species are measured to estimate its characteristics over the entire study area. Hence, a sample must be representative of the whole study area and more than one sampling location is needed to account for the variation in the measurements made in the field. For instance, the population density or even the presence of a species depend on environmental conditions and this is to be taken into account to estimate the mean population density or the infection rate of the species in the study area or to document its spatial distribution with sufficient accuracy. Precision (typically measured by standard error) reflects how similar to each other are the different measurements made in the sampling locations, thereby providing a measure of sampling uncertainty. When sample measurements are similar to each other, the sample mean is likely to be estimated with an acceptable level of precision from a few sampling locations. In contrast, when the between-location variation in the measurements is high, a larger number of sampling locations is needed [11]. Achieving a sufficient level of precision is of critical importance: the higher the precision of the estimates, the better the chances to detect temporal changes using statistical hypothesis testing procedures. Sample size is also known to impact on the performance of species distribution models [17–19]: predictions based on few records are likely to be less accurate than predictions based on larger sample sizes [18]. A sufficient number of sampling locations is needed to capture in the statistical models the response of the species to the environmental conditions. A balance is, therefore, to be achieved between ensuring statistical robustness (i.e. increasing the sample size) and reducing sampling effort (i.e. decreasing the sample size) because sampling is time- and/or budget-consuming.

For monitoring purpose, a power analysis may be performed to evaluate the number of sampling locations required to detect a given level of change over time in the attributes of the species with a predetermined level of statistical certainty. First, decisions are to be made by the users on (1) the minimum level of change that is to be detected in the analysis (for instance, 10% of change between time t and t+1) and (2) the acceptable chances of making type-1 (i.e. concluding that change is taking place when it is not) and type-2 (i.e. concluding that no change is taking place when it is) errors in hypothesis testing procedures [15]. Such decisions are often based on the precautionary principle and the relative importance of type-1 and type-2 errors also depends on the objective. Then, the analysis integrates information on the precision of the estimates to calculate the optimal sample size needed to detect the desired level of change. A pilot survey is, however, required to obtain an initial approximation of the precision of the estimates linked to the variation in the field measurements. For modelling applications, modelling performance increases with sample size and impact of sample size on modelling performance may strongly depend on the modelling technique used [20]. A series of studies

have also recently shown that the performance may be sensitive to particularly small sample sizes and may reach an asymptote level beyond a sufficiently large sample size [18,19]. In order to examine how large the sample size should be to obtain sufficiently well-performing models, different alternative options are available: (1) using readily available datasets in the study area [12] or (2) creating virtual species in real landscapes [19,21]. With such data, it becomes possible to manipulate the number of sampling locations to represent a range of sample size and to examine the impact of restricted sample size on modelling performance.

4.1.2. Optimizing sampling strategy in space

An appropriate sampling design also involves positioning the sampling locations so that the full range of environmental conditions across the study area may be covered to ensure the representativeness of the sample. Several approaches are available to position the sampling locations (only some are presented below) [21] with different advantages and disadvantages (details in [11]): Those include:

Expert-based sampling – Sampling units are located based on a priori knowledge of the study area and the status of the species. This subjective strategy is to be avoided because the sample is most likely not representative of the study area and may thus not be used for statistical inference.

Random sampling – Random selection of sampling locations among a list is an easy-to-use procedure that is recommended when the aim of the sampling is to provide a picture of the situation across the study area. However, the precision of the estimates may be much lower than when using a stratified sampling (see below), especially in heterogeneous environments.

Systematic (or regular) sampling – A regular distribution of the sampling locations may prove to be appealing because the whole study area is covered with the same sampling effort. However, the sample may provide a biased picture when the fixed distance between sampling locations coincides with a particular structure in the spatial arrangement of the environmental conditions.

Stratified sampling – The study area is first divided in strata assumed to influence differently the attributes of the species measured in the field. A random sampling procedure is applied to select a number of sampling locations within the strata in ratio to their relative geographical extent. The main advantage of stratification is that the precision of the estimates based on the sample may be considerably improved compared to a simple random sampling. Stratification requires preliminary survey to be conducted to minimize the within-strata variation in the measurements. In practice, however, stratification is often applied according to environmental layers representing heterogeneity of the environment conditions that are assumed to exert an influence on the attributes of the species.

4.1.3. Optimizing sampling strategy in time

Presence-only techniques can deal with the issue of false absences in species distribution modelling studies [14,22], and failure to consider the detectability of a species (i.e. the probability of detecting it when present at a site) when designing a monitoring pro-

gramme might lead to misleading conclusions [23,24]. In order to account for detection probabilities and to provide an unbiased estimate of *Anopheles* species occupancy or infection rate, it becomes necessary to carry out repeated survey at least in some sampling locations over a single season of data collection. If the emphasis of the programme is on estimating changes in the species occupancy or infection rate over time, it is also required to repeat the surveys from one season to the other. Site occupancy modelling is a statistical framework specifically designed to jointly estimate detectability and occupancy of the species as well as changes in those parameters over time [24]. Designing effective sampling schemes to estimate *Anopheles* species dynamics in space and time requires decisions to be made about how to allocate sampling effort among spatial and temporal replicates. Power analyses may be implemented to optimize the sampling design in space and time, i.e. to achieve a compromise between the number of sampling locations and the number of repeated surveys within sampling locations in relation to (1) the acceptable level of imprecision associated with the estimates of species occupancy, (2) the occupancy and detectability of the species, (3) the available manpower and possible sampling effort.

4.2. Environmental factors

Once environmental factors of interest are identified, their importance according to the type of climate (e.g. semi-arid or humid), type of species, and altitude must be further discussed. Any place where surface water is available for breeding and emergence might lead to *Anopheles* occurrence. Vector status requires above plus (1) presence of human/animal host and their disease parasites. Then (2) suitable temperature and humidity which have then an effect on (3) vector dynamics and parasite development. A review [10] of the current state of the art in the context of remote sensing applications for malaria underlines that, temperature, humidity, surface water, climate seasonality, vegetation type and growth stage influence vector abundance irrespective of their association with rainfall. The vegetation around breeding sites may also determine abundance associated with the breeding site by providing resting sites, sugar feeding supplies for adult mosquitoes and protection from climatic conditions [25]. Furthermore, vegetation type or land use may influence mosquito abundance by affecting the presence of animal or human hosts and thus availability of blood meals [10]. Factors are of two kinds [9]: (1) abiotic slow changing factors such as long term climatic variables, soils, topography, (2) fast changing biotic factors such as vegetation, presence of predator, hosts, interactions with other *Anopheles*, seasonal temperature/ rainfall, water bodies,….

Remote sensing products provide environmental characteristics on large surfaces even in areas of limited accessibility and can provide recent information on an area compared to commonly available maps. The quality of the information provided is however dependent of the original remote sensing data quality and suitabillity. The processing required to mosaic images in order to cover a large area, to make various types of image correction, cloud screening operations and image interpretation are not straightforward for non-specialists. Derived products, such as land cover maps or composited time series of simple vegetation indices, are therefore often more adapted to the need of the users. However, the process behind the final product must be understood to a certain extent by the users, in order for them to be aware of the assumptions and simplifications done in the processing. Furthermore, different methods

are typically available to reach a given goal, and the choice of the method can strongly influence on the quality of the results.

4.2.1. Long term abiotic variables

Abiotic slow changing factors might be used to delineate a species distribution area or maximum potential extend for a species. Those factors include topography, soil types, long term climate and ecoregions (Table 1). Available source are not many but cover the world. Consistent topography is available from the USGS GTOPO 30 suite [26] including derived variables such as digital elevation model, flow accumulation, slope or aspect or from the NASA Shuttle Radar Topographic Mission (SRTM) dataset reprocessed by the CGIAR [27]. The digital soils map of the world compiled by the FAO [28] is still a reference. Long term climatic datasets of monthly temperature and rainfall are available from Worldclim [29] which provided also bioclimatic variables. A second dataset CRU CL2.0 [30] provided also monthly temperature and rainfall but also number of monthly rainy days, rainfall monthly variation and relative humidity. The datasets are based on meteorological stations data from 1950 to 1990 or 2000. The quality of the data is high in some areas and less in others due to availability of meteorological station which can be quite low, particularly in Africa. The ecoregions [31] are a useful dataset to delineate sample stratification at regional level. Those dataset are mostly not derived from remote sensing (RS) images but grids developed from point data.

Variables	Sensor/ source	Resolution	Date
Topography	USGS (not RS) (http://www1.gsi.go.jp/geowww/ globalmap-gsi/gtopo30/gtopo30.html)	1km	1996
Topography	SRTM (http://srtm.csi.cgiar.org)	100 m	February 2000
Soils	FAO (not RS) (http://www.fao.org/nr/land/soils/ digital-soil-map-of-the-world/en/	17km	1990
Climate	Worldclim (not RS) (http://www.worldclim.org/)	1km	Published 2005, (1960 to 2000)
Climate	CRU. 10 (http://www.cru.uea.ac.uk/cru/data/ hrg/)	15km	Published 2005, (1961 to 1990)
Ecoregions	OLSON		1998

Table 1. Relevant long term abiotic variables

4.2.2. Monitoring air temperature

Air temperature Ta, is commonly obtained from measurements in weather stations, which depend on the regional infrastructure. Data are collected as point samples whose distribution is rarely designed to capture the range of climate variability within a region especially in developing countries. The data is also not readily available for real time applications and need to be interpolated to obtain information everywhere in a given region. On the other hand satellite images can provide land surface temperature Ts which is different from the air temperature and corresponds to the temperature of the top of the features present on the land

surface (i.e. snow, ice, grass of a lawn, roof of a building, leaves of the canopy in a forest). Specific methods (split-windows techniques) can derive daily Ts at 1 km resolution [32,33] from two types of sensors, namely the Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) [33,34] (see description table of MODIS Ts: https://lpdaac.usgs.gov/products/modis_products_table). On the contrary, the derivation of air temperature (Ta) is far from straightforward. Recent research showed that minimum Ts retrieved from MODIS night images provide estimates of minimum Ta in different ecosystems in Africa [35]. Information on maximum Ta is also needed to study heat waves and can influence the transmission of vector-borne diseases in regions where temperature is a limiting factor. During daytime the retrieval of maximum Ta from Ts is more complex due to factors which influence (Ts-Ta): i.e. solar radiation, soil moisture and surface brightness. Methods based on Temperature Vegetation index, Normalized Difference Vegetation Index and Solar Zenith Angle to correct (Ts-Ta) are not sufficiently accurate to retrieve maximum Ta in different ecosystems [35]. Therefore, a new approach has been recently proposed [36] to estimate maximum Ta based on night AQUA-MODIS Ts data in combination with Worldclim [29] which provides long term monthly average of maximum and minimum air temperature. These inputs allow to characterize the diurnal cycle (amplitude and phase) and determine maximum Ta by extrapolating in time minimum Ta according to the determined diurnal cycle. The method is used to produce maximum Ta maps at 1km every 8 days over Africa available in real time from the International Research Institute for Climate and Society (IRI). Unfortunately Ta does represent temperature outside but no proxies are available to monitor indoor temperature or other stable microenvironment which can explain transmission in Finland when temperature in -20°c outside and is important in highlands malaria in Africa.

4.2.3. Monitoring rainfall

In some regions, the spatial distribution of weather stations is limited and the dissemination of rainfall data is variable, therefore limiting their use for real-time applications. If satellite-based data can partly compensate and help to monitor rainfall, unfortunately, no satellite yet exists which can reliably identify rainfall and accurately estimate the rainfall rate in all circumstances. Some sensors can make indirect estimates of rainfall by measuring parameters such as the thickness of clouds or the temperature of the cloud tops. Advantages and drawbacks of existing methods are summarized in [37]. Various satellite rainfall products exist at continental or global scales. The most relevant are:

- The Tropical Rainfall Measuring Mission **(TRMM)** products [38] provide better spatial (25 km) and temporal estimation (3 hours) of rainfall in Africa [39] than most products but are available only between 35° North and South latitudes.

- Products from the CPC MORPHing technique **(CMORPH)** [40] cover the world at 8 km resolution every 30 min. This technique uses precipitation estimates derived from low orbite satellite microwave observations obtained entirely from various geostationary satellite infrared (IR) data. The estimation method developed for these products is extremely flexible such that any precipitation estimates from any microwave satellite source can be incorporated.

- African Rainfall Estimation **(RFE)** products cover Africa. The current version (RFE2) uses microwave estimates in addition to the use of cloud top temperature and station rainfall data to provide daily rainfall estimation at 10 km resolution. Comparison between CMORPH and RFE over complex terrain in Africa [39] and over the Desert Locust recession regions [41] shows that no single product stands out as having the best or the worst overall performance [42,43].

- The TAMSAT African Rainfall Climatology And Time-series data version 2 **(TARCAT2)** product [44] covers Africa at 4 km resolution and is derived from the MeteoSat thermal infra-red (TIR) satellite imagery. It consists of rainfall information every 10 days.

- The Multi-sensor Precipitation Estimate **(GRIB MPE)** [45] derives instantaneous rain rate from the infrared (IR) data of the geo-stationary EUMETSAT satellites over Europe and Africa by continuous re-calibration of the algorithm with rain-rate data from polar orbiting microwave sensors. The algorithm is only suitable in convective weather. Frontal precipitation, especially at warm fronts is very often wrongly located and overestimated. Two quality indicators distributed together with the MPE product give indications where the product should be used and where it may be problematic. Temporal resolution is high (15 min) and the product available in real time.

4.2.4. Remote sensing indicators of vegetation status

Monitoring the status of green biomass from space is made possible thanks to the particular spectral properties of green vegetation. In order to drive the exothermic reaction of photosynthesis, plant pigments absorb electromagnetic radiation over different parts of the visible spectrum (400-700 nm). This is known as photosynthetically active radiation (PAR). Additionally, much of the near-infrared light (740-1100 nm) is scattered by green plant tissues to avoid overheating, and this scatter results in strong spectral reflectance at these wavelengths. These unusual spectral properties, which are directly linked to photosynthesis, stomatal resistance and evapotranspiration, facilitate the retrieval of information on plant canopies from the electromagnetic signal measured by satellite remote sensing instruments [46]. Satellites dedicated to vegetation monitoring have been equipped with sensors capable of measuring reflected electromagnetic radiations in various wavebands, with a particular emphasis on the red (Red) and near-infrared (NIR), to assess the green biomass in a canopy.

A common and simple way to resume the information content within these bands is the use of spectral vegetation indices, which is an algebraic combination of the spectral bands designed to be as sensitive to the desired factor (green biomass) and as insensitive as possible to perturbing factors affecting spectral reflectance (such as atmospheric and illumination conditions, soil properties and the viewing geometry of the imaging instrument). Indices based on red and near-infrared reflectance have been shown to be a measure of chlorophyll abundance and energy absorption [47]. Variations of across one year can help spotting vegetation types, and the quantification of the water content can help identifying areas in a similar vegetation class which retain more humidity and might thus be more favourable to mosquito breeding or survival in dry season. Dozens of vegetation indices assess the state of the vegetation qualitatively and quantitatively on the basis of reflectance values:

- The Normalized Difference Vegetation Index **(NDVI)** (NDVI = (NIR - Red) / (NIR + Red)) [48] is the most popular of such vegetation indices. NDVI is easily available because it is based only on Red and NIR bands, which are present in most satellite sensors dedicated to land surface observation. The GIMMS (Global Inventory Modelling and Mapping Studies) NDVI dataset based on NOAA-AVHRR offer the longest coherent dataset from July 1981 to December 2011 which can be useful for long term studies [49]. However the spatial resolution of 8 km limits some applications. SPOT-VEGETATION provides a regular product since 1998 at a better spatial resolution of 1 km and geo-location. Similarly, the MODIS sensor provides NDVI at 250 m resolution. NDVI can also be calculated from images with a higher spatial resolution, such as those from the Landsat or SPOT series. The NDVI is used extensively but has several disadvantages such as its sensibility to atmospheric aerosols and to soil background (particularly in sparsely vegetated areas) [50]. Additionally, NDVI also tends to saturate in forested areas and is therefore not responsive to variations in the full range of canopy vegetation content [51].

- The Enhanced Vegetation Index **(EVI)** remains sensitive to variations in dense forests where NDVI saturates [52]. EVI calculated from MODIS imagery is provided, alongside NDVI, as standard freely available product. A disadvantage of EVI is that it requires an additional blue band, which is not available in NOAA-AVHRR, thereby blocking the possibility to exploit the long term dataset. To remediate that, a simplified 2-band EVI has also been proposed [53].

- The Normalized Difference Water Index (**NDWI** = (NIR - SWIR) / (NIR + SWIR)) [54], where SWIR is the Short wave infrared, is sensitive to vegetation water content and to the spongy mesophyll structure in vegetation canopies. Regarding vegetation water content, [55] summarized the limitations of using the NDVI: a decrease in chlorophyll content does not imply a decrease in vegetation water content and inversely. It might also help target vegetation retaining humidity in the dry season. Few studies have attempted to retrieve directly vegetation water content using operational satellite data such as provided by SPOT-VEGETATION [55], MODIS [56] and Landsat [51]. A disadvantage of NDWI is that several instruments are not equipped with detectors in the SWIR domain, and when they do they are often at lower spatial resolution than other bands.

- The **Hue index** is a qualitative index proposed recently by [57] for the monitoring of the Locust habitat. This exploits simultaneously three wavelengths (the SWIR, the NIR, and red) and has two main advantages: (i) avoiding confusions between bare soils and vegetation, and (ii) allowing the identification of green vegetation independently from the observation conditions, i.e., atmosphere and acquisition geometry, and from its intrinsic variations, i.e., the phenological stage. Potential for monitoring crops, forests and other applications still need to be assessed.

Albeit their widespread use, the use of vegetation indices over large geographic extents has its limits for describing canopy status in a fine and robust way, since both the desired information and the perturbing factors vary spatially, temporally and spectrally. Another type of information on canopy status that can be retrieved from remote sensing data is biophysical variables. The most common are the fraction of Absorbed Photosynthetically Active Radiation

(fAPAR) and the Leaf Area Index **(LAI)**, defined as half the total developed area of green leaves per unit of ground horizontal area [58]. Unlike vegetation indices, which are a convenient way to resume spectral information related to vegetation behaviour, biophysical variables such as fAPAR and LAI have a real physiological meaning. These variables govern the process of photosynthesis and the exchange of energy, water and carbon between the canopy and the atmosphere. To retrieve LAI and fAPAR from satellite remote sensing observations, the radiative transfer of photons within the canopy and through the atmosphere must be modelled. A thorough description of the physical problem, alongside caveats on its application to satellite remote sensing of vegetation, is presented in [59]. Dorigo et al. [60] provide a review of the various methods that exist to use such radiative transfer models to relate satellite observations to LAI and fAPAR. Up to recently, the two main datasets of global fAPAR and LAI are products from MODIS and CYCLOPES with different methodologies described in [61] and [62]. These datasets have been inter-compared and evaluated against ground measurements over different land cover types [63–65]. A combined product has recently been made available, GEOV1, in the framework of the Geoland2 project, in view of providing it as an operational land product service of the Global Monitoring for Environment and Security (GMES) pro-gramme [66]. This product is currently based on SPOT-Vegetation, but a compatible long term data record from 1981 to 2000 has been also constructed based on NOAA-AVHRR data (with a spatial resolution of 0.05°) [67], and in the future it is expected to be produced based on the future operational Sentinel3-OLCI mission. Such biophysical products are increasingly used but seldom in epidemiological studies.

4.2.5. Land cover

Detailed information from land cover maps is generally available in national geographical institutes but this information is often out of date due to the long process implied in developing such dataset for a whole country. Moreover, the diverse origin and scale of these datasets when considering more than one country impeded proper comparison between sites. One could thus consider producing national or regional land cover maps using satellite high-resolution data. This exercise includes the pre-processing, the interpretation of the images, and the validation through field surveys. For instance, Landsat images were used in the framework of the Food and Agriculture Organization of the United Nations (FAO) Africover program [68] to map land cover types at 30 m resolution for 11 countries in Africa. Such land cover maps present a great level of detail, but may suffer for some inconsistencies because of heterogeneity in acquisition dates, images and interpretation from one scene to another. Moreover this approach hardly takes into account the seasonal variation and phenological behaviour of different vegetation types. These datasets are also limited in their spatial coverage and cannot be regularly updated following the methodology commonly used (i.e. visual interpretation). Finally, if the whole Landsat images archive was made freely available in 2009, images from Landsat 7 present gaps since May 2003 and Landsat 5 back to activity in 2003 is now failing since November 2011.

Medium to coarse resolution imagery (250 to 1 km) can improve some major issues: the information is acquired consistently over the whole area and frequent images (every 1 or 3

days) of a same area can be combined to eliminate cloud contamination and angular effects, and characterize the vegetation phenology. These time series can be used to produce global maps such as (i) the Global Land Cover 2000 **(GLC2000)** map that is based on SPOT-VEGE-TATION data (1 km) thanks to an international partnership of research groups coordinated by the European Commission's Joint Research Centre (JRC) [69], (ii) the 500 m **MODIS global land cover** derived from collection 5 Nadir BRDF-Adjusted Reflectance (NBAR) and Land Surface Temperature (LST) products [70], (iv) the **GlobCover** map [71] at 300 m derived from a Medium Resolution Imaging Spectrometer (MERIS) time series for year 2005. These types of time series were also used to produce land cover and vegetation maps at national and regional scales such as for example [72]. These types of products have the advantage that the data preprocessing and the methodology used are adapted to the local constraints and application needs but are limited in their spatial coverage. The possibility to regularly update global land cover information has been proved recently with the second run of the GlobCover processing system [73], thus offering the potential to use such product in a monitoring program. The delineation of the vector habitat underlines the essential role of these land cover datasets which makes the necessary link between the technical remote sensing world and application requirements. Land cover dataset are one of the essential variables for the Group on Earth Observations (GEO). A major effort is to be continuously invested in the development and improvement of such dataset. The quality of this dataset can only be really tested if used for applications. Close interactions with final users remain the guarantee for the relevancy of the Earth observation product.

4.2.6. Monitoring water bodies

In order to identify the presence of water, it is also possible to use satellite-derived products that detect water bodies instead of approximate water availability using rainfall estimates. In the last 10 years, only a few operational methods applied to datasets with a spatial resolution equal or higher than 1 km, were proposed to monitor surface water at continental or global scale. Among these, two most recent offer dynamic detections in near real-time through an operational monitoring system:

- First, the Small Water Bodies **(SWB)** product based on SPOT-VEGETATION [74] available via the DevCoCast project website makes use of 10 day NDVI, the NDWI and syntheses of the SWIR band data. It is based on a contextual algorithm [75] exploiting the local contrast of the water surface with respect to the surrounding area. The product performs well in sub-humid and semi-arid regions, but limitations have been observed over dense vegetation areas. The 1 km spatial resolution is an intrinsic limitation. Nevertheless, the combination of eight years of small water body monitoring data demonstrated the value of multi-annual approaches to capture water bodies that do not replenish every year in relation with seasonal rainfall patterns [74][97].

- **The HSV WATER** product [76] based on Hue Saturation Value (HSV) transformation of SPOT-VEGETATION and MODIS time series allows consistent detection at continental scale. This pixel based approach uses SWIR, NIR and red bands and transform the RGB color space into HSV that decouples chromaticity and luminance. It presents the advantage

to have a robust and reliable image-independent discrimination between water and other land cover types. An automatic processing chain based on SPOT-VEGETATION was designed to provide a dekadal water surface product at the continental scale. The product can be ordered freely through the geoland2 web portal following the link http://www.geoland2.eu/core-mapping-services/biopar.html.

The analysis of eight years of small water body data demonstrated the capacity of such methods to capture inter-annual water bodies variability and the relation with seasonal rainfall patterns [98]. Nevertheless, the 1 km spatial resolution of products derived from SPOT-VEGETATION is still a strong intrinsic limitation. The operational production of a MODIS based product at 250 m using the second method is in progress and should be available soon.

4.2.7. Caveats on remote sensing data

Various issues have to be highlighted when looking from the application angle:

- The spatial resolution for all environmental factors necessary for a study is often not similar (table 2) and transformation to similar resolution might lead to increase geolocation imprecision when pixels limits do not correspond. The pixel size selected for the analysis is dependent of the available datasets and not the best cell size to describe the phenomena under study. Datasets not available at high resolution thus limit spatial details of results.

- Some useful dataset are not covering the world or not available at the appropriate date.

- Too detailed datasets such as rainfall products with data every 3 hours (TRMM) would require long summarizing process for non-specialists to get information per week or month.

- At high spatial resolution, geo-location accuracy can be jeopardized by the viewing angle, particularly in accidented terrain. Image distorsion needs to be corrected using a topographic information not always available at high resolution.

Spatial resolution: Remote sensing is typically characterized by a trade-off between the different types of resolutions: spatial, temporal, spectral and to a certain extent also radiometric and angular. High spatial resolution is desired to characterise the land in a detailed way. However, cloud occurrence limits its availability. Basis for land cover map might be a puzzle of images from different seasons or even years thus creating artifacts of land cover differences at the limits between the images. As it is discussed further, coarse spatial resolution imagery, with its frequent revisit and through the use of compositing can partially remediate the problem, but this can be a problem with high resolution imagery where images are costly and revisit not frequent. Having regular observations at fine spatial resolution typically limit the geographic extend that can be monitored. Even over a limited coverage, satellites providing such services are typically commercial ones for which the cost is currently high and for which there is competition for their observation capacity between different geographic sites. Such images are thus often used in studies of limited spatial extend from which the results are difficult to extrapolate to a country level needed for spatial surveillance. It is however just a matter of time before high spatial resolution (5 – 20 m) becomes available for the entire globe

Type of variable	Sensor/ source	Spatial	Timing	Area	Date
Minimum air temperature	Terra/Aqua MODIS http:// iridl.ldeo.columbia.edu/ maproom/.Health/.Regional/. Africa/.Malaria/.TMR/)	1 km		World	
Maximum air temperature	MODIS/ Worldclim	1 km		World	
Rainfall	TRMM	25 km	3 hours	35° N & 35° S	Since 1998
Rainfall	CMORPH : Mixt IR data, NOAA-CPC	8 km	30 min	World	Since December 2002
African Rainfall Estimation (RFE)	produced by NOAA-CPC	10 km	daily	Africa	Since January 2001
Rainfall	TARCAT http:// www.met.reading.ac.uk/ ~tamsat/data	4 km		Africa	Since 1983
Rainfall	EUMETSAT/ GRIB MPE				From January 2007
NDVI	GIMMS products from NOAA-AVHRR	8 km		World	Since 1981
NDVI	Spot VEGETATION	1 km	daily	World	
NDVI MODIS	Terra/Aqua MODIS	250 m		World	
EVI	Terra/Aqua MODIS	500 m		World	
NDWI	Spot VEGETATION	1 km	daily	World	
Hue Index	Spot-VEGETATION	1 km	daily	World	
fAPAR + LAI	Terra/Aqua MODIS	1 km	8 days	World	Since 2000
fAPAR+LAI	CYCLOPES products fSpot-VEGETATION	1 km	10 days	World	Since 2000
faPAR+LAI	GEOV1 products Spot-VEGETATION	1km	10 days	World	Since 2000
fAPAR+LAI	GEOV1 products from NOAA-AVHRR	5 km	10 days	World	1981 - 2000

Table 2. Some important remote sensing related products

and the European Space Agency is currently preparing its Sentinel-2 constellation (with an expected launch of its first satellite in 2014), which aims at operationally providing multispectral imagery, at spatial resolutions of 10 to 60 m for different bands, and with a 5-day revisit period. However, the challenge of collecting, processing and delivering this data may still limit its practical use for years.

Clouds and compositing: The quality of the spatial and temporal spectral consistency of coarse resolution optical time series may be limited by processing steps of cloud-screening and compositing. The efficiency of the cloud-screening, i.e. its ability to remove clouds while keeping a maximum of useful information, depends on three factors: (i) the methodology used to identify cloud-free pixels, (ii) the type of clouds (thick clouds are easier to overcome than

veils of clouds which change surreptitiously radiation values), and (iii) the sensor character-istics. The detection of clouds is often based on specific bands, i.e. the blue, the middle infrared and the thermic infrared, and the choice of the wavelengths may vary according to the sensor. Depending on these factors, residuals clouds and haze may still remain after the cloud-screening step. Quality of time series may strongly vary according to the compositing strategy used. The most common method used for producing temporal syntheses consists of selecting the Maximum Value Composite (MVC) NDVI [77] (Figure 2) that minimizes the effect of undetected clouds since these would typically have a lower NDVI value. However, the composited reflectance bands may exhibit substantial radiometric variations, since composite radiances are generally recorded under varying atmospheric and geometric conditions. This may cause serious spatial inconsistencies in the composites and in the subsequent processing.
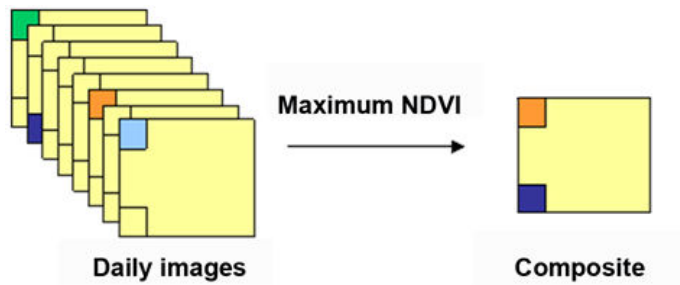

Figure 2. Maximum NDVI standard compositing

A more advanced approach consists of normalizing the bidirectional reflectance by fitting a bi-directional reflectance distribution function (BRDF) model to the available cloud free observations [78] which considerable improve the result. But operational implementation requires a large number of cloud-free observations, the BRDF retrieval has a high sensitivity to residual clouds [79] (Figure 3), the algorithm is complex and requires ancillary data. A more flexible and "user-friendly" compositing approach was recently proposed [80] where cloud free reflectance values are averaged after a quality control.
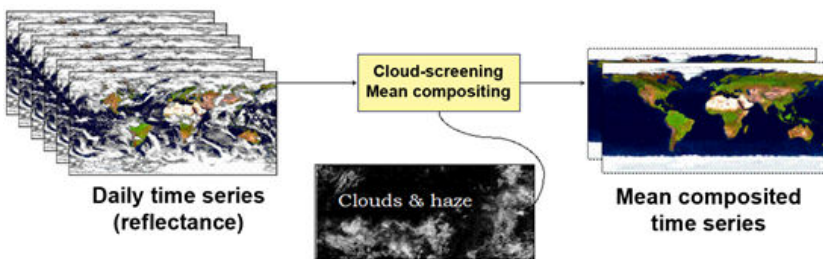

Figure 3. Mean composting method

It presents the advantages to reduce both the anisotropy effects and the possible remaining perturbation after atmospheric correction and cloud removal. Despite the benefits of compositing, for some applications it may be more interesting to avoid it altogether. Indeed, to follow vegetation changes at a finer time scale it may be better to exploit all available observations within a period (typically 10 days or more) instead of combining them together. In agriculture monitoring, considerable changes in biomass or phenology can occur within a week and exploiting all available observations should thus be preferred. Such approach has been used, to provide crop specific biophysical variable time series at regional scale by fitting a simplified model of the canopy dynamics over daily data [81] and might be of use to identify processing occurring in potential *Anopheles* habitat such as rice paddies.

What is in a pixel? Coarse spatial resolution satellite imagery has several advantages. Frequent observations enable timely detection of environmental changes that may indicate potential changes in the presence of *Anopheles*. Second, the higher frequency of available observations allows to better address the problem of lack of data due to cloud contamination and anisotropy through compositing or temporal smoothing. Third, their (relatively) long archives enable to have a picture of the past with which the actual conditions can be compared to. In the short coming future, coarse datasets may also serve as a benchmark in order to calibrate products to their signal, which could be more stable thanks to their higher revisit frequency. Finally, coarse spatial resolution data are also often the only data available and there is thus a tendency to use them at the limit of their spatial resolution by looking at individual pixels. A common misconception is that the observational footprint is the geometric projection of a rectangular pixel onto the Earth's surface [82]. The footprint rather depends on some properties of the instrument, resumed under the concept of spatial response [83], and which results in an observation footprint generally larger than the pixel delivered to the user (Figure 4).

This problem is compounded for sensors such as AVHRR, MODIS and VIIRS (the successor of MODIS), which scan the Earth with large angles, leading to an expansion of the observation footprint along the scanline (while the grid in which the data is provided keeps the same size). Furthermore, the pre-processing step of gridding, i.e. assigning an observation to a predefined system of grid, introduces a ``pixel-shift'' [84], which means that the centre of the pixel does not correspond with the centre of the observation. Such gridding artefacts have serious consequences on the quality of the MODIS signal, and more specifically on composites and band-to-band registration across various spatial resolutions [85]. Recent work [86] has further demonstrated the impact of gridding artefacts and the scan angle on the spatial purity of an observation, i.e. on the percentage of the target land cover within an observation footprint that effectively contributes to the signal encoded in the pixel.

Mosquito Land cover: Land cover provides the more understandable information to non-specialist in terms of vegetation and habitat but the classes are not always adapted to the user needs. Instead of choosing between vegetation indices which represent continuous values and land cover of more or less 20 classes, it might be useful to give access to intermediary products of land cover classification. Indeed, processing chains of a land cover such as GlobCover include a correction process, cloud screening and image compositing to improve overall quality of the data [71]. Then vegetation indices and reflectance bands linked to vegetation
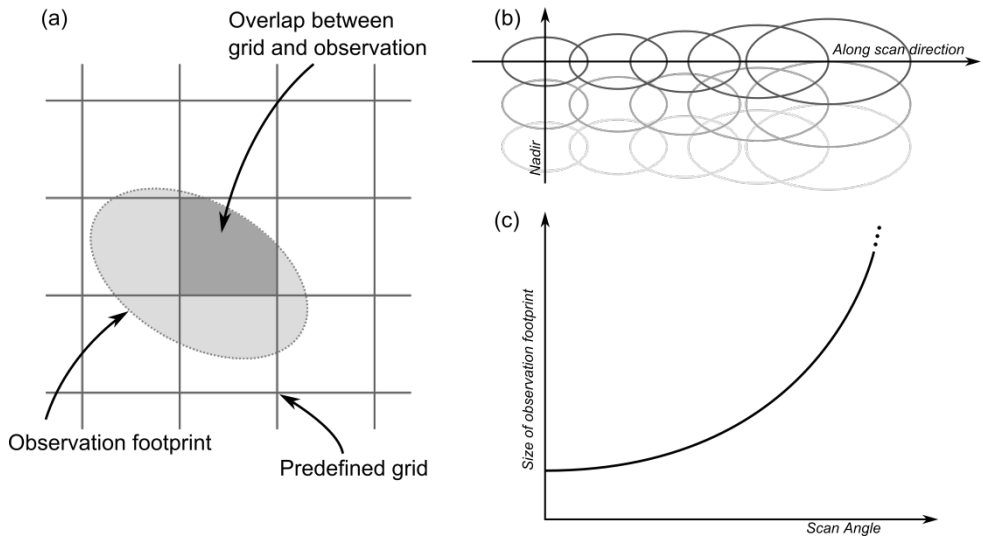
Figure 4. Some effects influencing what is really in a pixel of a satellite remote sensing image. (a) Schematic mis-alignment between the observation footprint and the arbitrary grid remote sensing observations are encoded into (i.e. the pixel). (b) Illustration of the expansion and overlap of the observation footprint along the scanline for whiskbroom sensors such as AVHRR and MODIS; and (c) representation of how this size increase as the sensor scans with larger viewing angles. Figure adapted from [87] with permission from Elsevier.

status are used to group similar adjacent pixels and assign to those a same class label through clustering method which creates a chosen number of classes. In the next step, each class is compared to classes of a reference existing dataset or other existing data. According to a set of decision rules, the classes are interpreted and grouped in definitive classes. This last step raises several issues. The transformation of continuous dataset and the separation of the continuous landscape into a set of discrete classes are bound to a loss of information and inaccuracy particularly at the border of the classes. For example, the transition from a forest to a meadow might not always present a clear cut border. Other land cover initiatives work with continuous fields to avoid this issue [88]. Moreover, at the end of the process, up to 30% of the pixels are integrated into mosaic classes used when it is impossible to attribute the group of pixels to a single class, the pixel itself being a mixture for example of forest and meadow and thus providing a signal which is neither corresponding to forest, neither corresponding to meadow. Using mosaic classes in models and analysis can create confusion, particularly if several mosaic classes are grouped together. In this context, access to intermediate products such as classes based on cluster of similar pixels produced by remote sensing specialist might allow to integrate those into ecological models integrating ecological information relevant to *Anopheles* into the building of the final land cover would allow to define a better suited product for the purpose. Integrating several sensors to build a Landover might also improve the results. Indeed, GlobCover is based on MERIS satellite images which do not contain the Short Wavelength Infrared (SWIR) useful for discrimination of the forest vegetation. A combination with spot VEGETATION could result into better discrimination power for a similar resolution.

4.3. Model development

Sampling strategies, detailed field studies and casual observations can provide data which constitute the baseline information for model development. While remote sensing products are still too coarse resolution or maybe not adapted to define microhabitats, they can however provide proxies for environmental factors influencing general habitat and might be used in two ways. (1) Environmental values can be extracted at the sampling sites or in a buffer around the sites and then related to *Anopheles* data in descriptive models. Buffer size is often a compromise between some meaningful ecological feature such as flying range and the spatial resolution of the environmental factors [89]. (2) For question regarding habitat, spatial variation in vector capacity and spatial surveillance, spatial models are needed. In these models, environmental factors are related to the species records collected in the sampling locations and this relation is then used to predict the distribution of the species beyond the sampling locations [90–93].

When working with existing data, sampling protocol cannot be influenced a posteriori but an adapted methodology can be used to take into account potential peculiarity of each dataset. Field data may be obtained as a by-product of existing operational projects. However, depending on the finality which determined sampling design, the data might not be used straightforwardly for spatial surveillance. The dataset might include non standardized data collected during different years, according to a variety of sampling strategies but might be the only data available covering many countries. Existing datasets can consist of a collection of literature records covering wide regions. However, the collection sites are seldom well geo-referenced, large areas are not covered by the studies which might use different collection techniques at different seasons. With such datasets lack of records might be linked to inefficient sampling method, wrong timing for the survey or absence of survey and according to the source of data, abundance and absence need to be treated with caution. Even certified presence might not reflect current situation if recorded years ago. These issues may partly be addressed by methods similar to the previously mentioned subsampling procedures to reduce the potential biases in readily available datasets or using adapted modelling techniques.

4.3.1. Species distribution modelling

Early development in the field of remote sensing and vector-borne diseases risk mapping used the following methodological steps: collecting human cases (or mosquito presence/absence), collecting relevant environmental gridded data (pixel), extracting data at sampling sites to build a logistic regression model explaining cases of occurrence according to the environmental conditions, then mapping the probabilities by calculation of the model output for each gridded cell of the original environmental maps [94]. Numerous methods have now been used to model vector-borne diseases spatially [95] and suggestions to improve frequent drawbacks include (1) using several models and select the best suited for prediction and (2) make a summary model from the best-fitting models. On the other hand innovative methods are constantly improved in spatial ecology. Quantifying the link between species and their environment is a central research area in quantitative ecology. When absence data are available / reliable, numerous methods now do exist, ranging from logistic regression, ordinary multiple regres-

sions and its generalized form (GLM), ordination, classification method, distance metrics such as Mahalonobis distances, neural networks, boosted regression tree, random forest and even more sophisticated support vector machines are some examples among the plethora of recently developed methods [14]. Multi-species community modelling methods have also been developed. One advantage of this kind of techniques is that it becomes possible to build species assemblage models that take into account the relationships between the different species in the community and so their relative location in the "environmental hyperspace", instead of modelling single species distribution independently from each other [96].

However, mapping elusive species such as mosquitoes is often a challenge mainly because of the impossible collection of reliable absence data such as described earlier. Discriminant approaches such as logistic regression analysis developed for specific diseases are thus not suited anymore because they compare environmental conditions in sites where the species is present and absent (not recorded). When only occurrence data are available, some niche-based modelling approaches offer adapted solution as they can use presence-only record information to build the statistical models. The concept of ecological niche has been defined [97] as follows: considering the n variables corresponding to all of the ecological factors relevant for the species, an n-dimensional hyper-volume can be defined in the environmental hyperspace between the limiting values permitting a species to survive and reproduce. This volume is called the fundamental niche of the species. This niche can be related to the two-dimensional geographical area of distribution considering that any point of the niche may represent a combination of environmental values that corresponds to some locations in the geographical space. Mechanistic approaches to ecological niche modelling [90] use direct measurements or physical modelling of response of individuals to parameters and infer from them individuals fitness values of different combinations of physical variables. On the contrary, correlative approaches to ecological niche models such as developed for species distribution models intend in a first step to define niches using the environmental variables at sampling point of occurrence, then assess for each spatial location in a study area probability to belong to the niche. Many large-scale species modelling techniques inspired by the principle of environmental envelopes were developed including BIOCLIM [98] based on a very simple classification tree, DOMAIN [99] based on a measure of multivariate distance, ENFA [100] based on the same principle of distance measure in an environmental hyperspace. Elith *et al.* [14] provide a good overview of most currently used methods including the Maxent method [22] based on presence data which seems to perform particularly well.

In this context classical presence-only modeling can also be integrated [9] into a hierarchical framework [101]. The first step is to model entomological data using environmental data relevant for the same time period. Indeed, mapping *Anopheles* information from literature records dating back several decades should be based on long term environmental factors such as climatic factors and not on factors such as land cover, or NDVI which are changing fast in some regions. The mapping of a first potential distribution based on long term slows changing information and literature records is then refined using a mask of fast changing updatable information such as land cover or current meteorological prediction. This allows producing a risk map or distribution map relevant for a specific date corresponding to the date of envi-

ronmental factors used to refine the map. The resulting map is thus ecologically meaningful and relevant for a precise date. Recent other improvements in the field of presence only models include selection of pseudo-absence with a spatial bias similar to the potential bias of presence data [102], selection of the environmental factors to enter the model based on ecological requirements, adapted method for species with low number of occurrence [103].

Some issues still need to be tackled however. Ecological model should be based on source populations. Those are sustainable populations in suitable habitat. To the contrary, sink populations are surviving in habitat not suitable for population persistence but persist thanks to immigration from nearby source population. Typical museum records include both sink and source populations [104]. Moreover, current vector-borne disease distribution may be limited by a number of factors both environmental and socio-economic. For example, during the past 100 years, malaria risk zone has reduced from around a half down to a quarter of the Earth's land surface. However malaria remains prevalent in 106 countries of the tropical and semitropical world, with 35 countries in central Africa bearing the highest burden of cases and deaths [105,106]. The latitudinal limits apparent today are in effect 'control frontiers' reflecting the interplay of control interventions combined with changes in environmental management and socio-economic developments that reduce community vulnerability to the disease [107]. Altitudinal limits to malaria transmission have been the subject of much discussion regarding shifting of malaria risk into highland regions, such as East Africa. If documented climate change [108] might have add a small impact, major factors for extension to new areas seem to be changes in land use and landscape leading to changes in local ecology for human and vector [109].

4.3.2. Time or space prediction — Evolution in time — Forecast

While delineation of potential habitat for a species is a first step in risk mapping for *Anopheles*-borne species, forecasting seasonal events and variation in (micro-) habitat suitability and mosquito population is essential. Remote sensing and Geographical Information Systems (GIS) contributed to the development of environmental systems to support vector control or more sophisticated early warning systems. Those systems usually target situations of epidemic malaria which occurs in regions where malaria is not present continuously but associated to climatic events such as a particularly wet season in near desert areas [110] or a hot season in African highlands [111]. Epidemic situation are predicted to increase preparedness in public health [112]. These first experiences are reviewed in [10]. Several trends are observed in current research, but a major effort is targeted towards the prediction of malaria epidemic season based on climatic/meteorological variables, particularly in the context of climate changes and availability of new meteorological data sources [35]. The disease risk is forecasted using seasonal climate prediction and in particular rainfall and sea-surface temperature [110], and influence of climate change analysed [113]. Following the development of the European ENSEMBLE System for seasonal to inter - annual prediction [114], challenging researches are now proposing to integrate the seasonal climate forecasts from climate model into malaria early warnings systems [115]. Regional specificity still needs to be integrated in such models as for example the fact that low rainfall may trigger epidemics in the highlands [116].

4.3.3. Anopheles vector capacity

When trying to assess disease occurrence risk, not only vector presence is necessary but the capacity and eagerness to transmit the diseases is essential. This capacity is well summarized in the vectorial capacity (VC) concept [117] derived from the Basic reproduction rate of MacDonald [118]. Vectorial capacity is a series of biological features that determine the ability of mosquitoes to transmit *Plasmodium.* It is defined as the daily rate at which future inoculations could arise from a currently infected case [119] and it is generally used as a convenient way to express malaria transmission risk. Interestingly, a spatial version of the VC called VCAP has been developed to propose a spatial version of the formula, allowing assessment of vectorial capacity for each pixel in a given area [120]. To be able to do so, the VCAP is VC only driven by minimum *Ta* and rainfall. Rainfall and temperature are used as inputs to the model because they have an impact on vectorial capacity. Temperature has an effect on both the vector and the parasite. For the vector, it affects the juvenile development rates, the length of the gonotrophic cycle and survivorship of larvae and adults with an optimal temperature and upper and lower lethal boundaries. For the parasite, it effects the extrinsic incubation period [121]. *Plasmodium falciparum* (the dominant parasite in Africa) requires warmer minimum temperature than *Plasmodium vivax*. This can account for the geographic limits of malaria transmission for this species in Africa [122]. At 26ºC the extrinsic incubation period of this species is about 9-10 days whereas at 20-22ºC it may take as long as 15-20 days. In highlands, where cold temperatures preclude vector and/or parasite development during part/or all of the year, increased prevalence rates may be associated with higher than average minimum temperatures [123] which might be led by period of low rainfall [116]. It is possible to use minimum *Ta* derived from MODIS for monitoring risks of malaria transmission in highlands regions including Eritrea and Ethiopia where a high proportion of the population lives at risk of epidemic malaria. Currently, the USGS EROS Center uses this temperature derived from MODIS night *Ts* on an 8-day basis jointly with rainfall data derived from the Tropical Rainfall Measuring Mission (TRMM) downscaled to 1 km spatial resolution to produce a 1 km VCAP map every 8-days specifically for the epidemic regions of sub-Saharan Africa [118]. In Eq. 1, the two raster images MODIS night time (T_s) and rainfall (TRMM) are integrated as follows:

$$VCAP = \frac{-\left(m\, a^2 \right)\, p^n}{\ln \left(p \right)} \tag{1}$$

Where:

m = 10.0 * TRMM

a = 0.7/gonotrophic

gonotrophic =[36.5/(T_s+2.0-9.9)]+0.5

p=$0.5^{(1.0/gonotrophic))}$

n=111.0/{[2.0*(36.5/ T_s+2.0-9.9)/gonotrophic]+T_s-18.0}
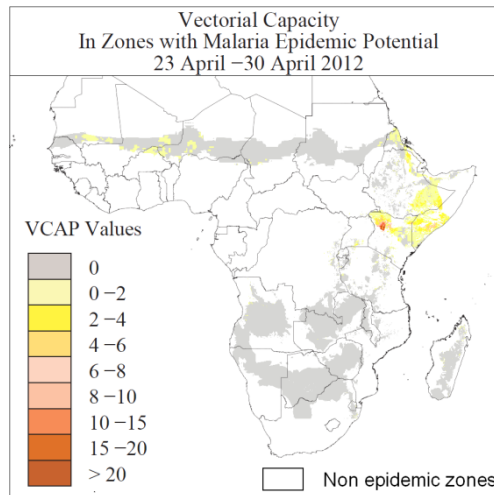
Figure 5. Vectorial capacity map VCAP provided for the epidemic zones of Africa at 1 km spatial resolution.

Parameter, m is the density of vectors (per human), a is the frequency of daily vector-man contact, p is the probability of a mosquito surviving through one whole day, and n is the extrinsic incubation period of malaria parasites or 'the time taken for completion of the extrinsic cycle'. Here the density m is estimated as a function of rainfall while the duration of the gonotrophic cycle and the extrinsic incubation period n are function of the temperature. The coefficients used in the VCAP equation are at this stage not optimized to specific regions. The variability in VCAP is only driven by the T_s and rainfall. This is a first attempt to spatially map risk of malaria transmission based on a vectorial capacity model. The product (Figure 5) is made available on a regular basis for the period Jan 2004 to present on the FEWS NET Africa Data Portal: http://earlywarning.usgs.gov/fews/africa/web and IRI data library: (http://iridl.ldeo.columbia.edu/maproom/.Health/.Regional/.Africa/.Malaria/.VCAP/)

The analysis of VCAP in relation to rainfall, temperature, and malaria incidence data in Eritrea and Madagascar shows that the VCAP correctly tracks the risk of malaria both in regions where rainfall is the limiting factor and in regions where temperature is the limiting factor [118]. However, in Burundi highlands, low rainfall triggered higher temperature and increased the risk of epidemics [116] and thus lower rainfall might be the trigger particularly because houses provide microenvironment with stable temperature 5°c higher than outside temperature and reduce influence of temperature on epidemic risk. The VCAP could also be further detailed by carrying analysis per vector species.

4.4. Transferring spatial information to health professionals

Roberts *et al.* [124] demonstrated many potential uses of remotely sensed data in managing and targeting vector and disease control measures. Just mapping the existing *Anopheles* species attributes can already bring information. Recently a map of all existing records for the *Anopheles*

dirus complex was proposed [125] to document ecological settings, but also to demonstrate that detailed mapping could bring much more information and could lead to more sophisticated models [9] from those datasets such as developed [6] for *Anopheles* vectors. In this context, major effort were made in the past to provide mapping expertise through customized GIS application to malaria control staff and help them to map their entomological and diseases cases records. Simply overlaying this information with existing environmental information can lead to new working hypotheses better defined by people with experience in the field. Current availability of easy to use packages such as Google earth and Google map offer new opportunities particularly in areas covered by detailed imagery. Studies carried out by scientist devoted to research provide outputs in scientific publications, in pdf format or might target small study areas not representative of the whole country. While this type of output is useful for advances in sciences it is often of little use to the health worker in the field. Two types of approaches are more adapted to the field and complementary. One is to provide ready-to-use product to integrate into operating systems, updated regularly to feed into early warning systems, or informative enough to provide the necessary clues for control and forecast. Those include vector capacity maps. The other approach is to bring the most expertise possible into the hands of the health worker.

However, to be fully operational the development of new products and early warning systems presented above must be integrated into a decision/action framework. There is currently a good deal of policy congruence through international, regional and local levels to support this effort (e.g. the Global Framework for Climate Services whose aims are to develop more effective services to meet the increasing demand coming from climate sensitive sectors including health). The remaining challenge is to get the knowledge into practice and sustaining it where it is needed. It is crucial that appropriate policies are developed and implemented to improve health system performance [126]. This may be helped by enhancing the workforces' ability to detect and treat diseases, monitor and predict spatio-temporal patterns and implement intervention and control strategies in a timely and cost-effective manner through the use of tools and analysis informed by climate data.

In order to get research outcomes into policy and practice it is important to understand the context in which policies are adopted and supported in a practical manner. Below is an example of how policies developed at the district and national level connect to the larger political agenda of international policy makers. At the global scale improved early warning, prevention and control of epidemics is one of the key technical elements of the current Global Strategy for Malaria Control [127] the RBM Partnership referenced earlier in this section. In Africa, Heads-of-State declared their support for the Roll Back Malaria initiative in April 2000 with the Abuja Targets [128]. In these targets, national malaria control services are expected to detect sixty per cent of malaria epidemics within two weeks of onset, and respond to sixty per cent of epidemics within two weeks of their detection. With the support of the WHO Regional Office for Africa, the WHO Inter-Country Programme Teams engage in the development of recommendations, guidelines and technical support to improve prevention and control of epidemics and transboundary/cross border within their various sub-regions (e.g. Regional Economic Communities (RECS) ECOWAS, IGAD and SADC) including collaborative activities with the

African Development Bank. As a consequence of these policy developments, nations epidemic prone have enhanced capabilities for delimiting epidemic/endemic prone areas; established epidemic malaria surveillance systems; and strengthening their epidemic response capacities with the help of the Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM) and other donor support.

In many national malaria control policy documents, countries now recognize that to achieve the Roll Back Malaria (RBM) and Millennium Development Goals (MDG) targets they need better information on where epidemics are most likely to occur, and some indication of when they are likely to happen. As a consequence, they have begun to explore the use of climate information in the development of integrated early warning systems. Thus, there is increasing congruence in policy initiatives from multilateral, bilateral, national and non-governmental agencies in relation to epidemic disease control and a growing demand for climate information and robust early warning systems to support these efforts. This is reflected in the newly emerging Global Framework for Climate Services. This policy congruence extends to the current discussions on adaptation to climate change. Strengthened health systems are also seen as vital to improving the management of climate-sensitive disease in the context of climate change. The IPCC identified building public health infrastructure as: *The most important, cost effective and urgently needed adaptation strategy*. Other measures endorsed by the IPCC include public health training programs, more effective surveillance and emergency response systems, and sustainable prevention and control programs. These measures are familiar to the public health community and are needed regardless of climate change and constitute what is the basis of a *no regrets* adaptation strategy [129,130].

5. Further research

In terms of data, interactions between *Anopheles* species should be investigated, those being sympatric on the same habitat or even breeding site or one dominant species deterring another species. Adapted methodology based on asymmetrical similarity coefficients, indirect clustering and the search of indicative species [131] have been proposed [132] to identify species association to help assess the risk of presence of elusive species, if another often associated species is present. Caveats and potential improvements to environmental factors have already been discussed. Remote sensing offers already a wide range of useful products but improvements could target easier delivery of products such as proposed by the IRI data library (http://iridl.ldeo.columbia.edu/) in similar standardised format and resolution and availability of all useful derived products over the world.

In terms of modelling, various issues have also already been discussed such as the necessity to better integrate ecological issues such as sink and source population [104]. Regarding the outputs, quality assessment could be attached to the resulting maps. Bayesian inference can be used [133] to quantify the uncertainty in the predictions. Rather than mapping the prevalence, what is mapped is the probability, given the data, that a particular location exceeded the predetermined high-risk prevalence threshold for which a change in strategy for control or the delivery of the drug is required. A level of uncertainty attached to each location help the decision maker choose which areas are at risk or not.

There is a necessity to document in details the data entered in models and choices of the modellers particularly when dealing with results which might trigger decision in public health [134]. Indeed, the final results do not only depend on input data but on pre-processing of those data, selection of useful variables, selection of a best model between various potential models, a whole process of model building which leads to one final result dependant on choices of the modeller. More details on dates of satellite images used to derived RS product, or even detailing quality spatially could also improve the final results and potential interpretation. Providing maps of the dataset entered in the model could help spot good spatial consistency or mismatch between adjacent raw images.

While disease occurrence prediction is generally the objective of forecasting, targeting the vector instead of the disease cases might provide several advantages. Indeed, some diseases might be present in a high number of asymptomatic carriers (lymphatic filariasis), or might not be accurately reported because the disease is not notifiable or misdiagnosis is frequent such as confusion between malaria and *Borrelia duttoni* in parts of Senegal and Togo [135]. Targeting the vector can help identify areas where asymptomatic cases might occur, target several diseases at once and predict epidemics or seasonal occurrence of diseases in advance based on fluctuations in mosquito populations.

6. Conclusions

In conclusion, providing relevant information to help disease spatial surveillance is not straighforwards and resemble more to a multidisciplinary challenge. In order to improve the current situation, increased sharing of existing data and increase transparency and documentation in the building of models could help target low quality areas such as places with low information or part of modelling process which could be improved. The quality of the entomological and environmental dataset as well as documentation of the relevant dates of each parameter such as original satellite images included in land cover maps and potential issues such as source-sink population sample could help identify new questions. Meanwhile, the information is still needed for the support of essential activities such as malaria control or for scientific research. A better interaction between research and operational work also seems to be necessary. Research product and results can only be useful if validated in the field and the best research questions are defined by people working in the field. Constant interactions can improve quality of research products and finally improve surveillance. Reinforcing the research capabilities in the region and in the malaria centres is of up-most importance. Indeed malaria workers in-countries have an extended experience of the field. They are in a better position to analyze the situation, identify their needs and find the answers. This would help bringing the data and the expertise where it is mostly needed: in the malaria centres.

Abbreviations

AVHRR Advanced Very High Resolution Radiometer

BRDF bi-directional reflectance distribution function

CGIAR Consultative Group on International Agricultural Research

CMORPH Products from the CPC MORPHing technique

CRU Climate Research Unit, University of East Anglia, UK

ECOWAS Economic Community Of West African States

EUMETSAT European Organisation for the Exploitation of Meteorological Satellites

EVI The Enhanced Vegetation Index

FAO Food and Agriculture Organization of the United Nations

fAPAR fraction of Absorbed Photosynthetically Active Radiation

FEWS NET Famines Early Warning Systems Network

GFATM Global Fund to Fight AIDS, Tuberculosis and Malaria

GIMMS Global Inventory Modelling and Mapping Studies

GIS Geographical Information Systems

GMES Global Monitoring for Environment and Security

GRIB MPE The Multi-sensor Precipitation Estimate

HSV Hue saturation value

IGAD The Intergovernmental Authority on Development, East Africa

IPCC Intergovernmental Panel on Climate Change

IR Infrared

IRI International Research Institute for Climate and Society

LAI Leaf Area Index

LST Land Surface Temperature

MDG Millennium Development Goals

MERIS Medium Resolution Imaging Spectrometer

MODIS Moderate Resolution Imaging Spectroradiometer

MPE Multisensor Precipitation Estimator

MVC Maximum Value Composite

NASA National Aeronautic and Space Administration

NBAR Nadir BRDF-Adjusted Reflectance

NDVI The Normalized Difference Vegetation Index

NDWI The Normalized Difference Water Index

NIR Near-Infrared

NOAA-CPC National Oceanic and Atmospheric Administration – Climate Prediction Centre

RBM Roll Back Malaria

RECS Regional Economic Communities

RFE African Rainfall Estimation

SADC Southern African Development Community

SRTM Shuttle Radar Topographic Mission

SWIR Short Wavelength Infrared

TARCAT The TAMSAT African Rainfall Climatology And Time-series

TIR Thermal infra-red

TRMM The Tropical Rainfall Measuring Mission

USGS United states Geological Survey Agency

VC (VCAP) Vectorial Capacity (spatial version)

WHO World Health Organisation

Author details

Valérie Obsomer[1*], Nicolas Titeux[2], Christelle Vancustem[3], Grégory Duveiller[3], Jean-François Pekel[3], Steve Connor[4], Pietro Ceccato[5] and Marc Coosemans[6]

*Address all correspondence to: valerie.obsomer@uclouvain.be

1 Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

2 Department Environment & Agro-biotechnologies, Public Research Centre - Gabriel Lippmann, Belvaux, Luxembourg

3 Joint Research Centre of the European Commission (JRC), Ispra, Italy

4 School of Environmental Sciences, University of Liverpool, Liverpool, United Kingdom

5 International Research Institute for Climate and Society, The Earth Institute, Columbia University, New York, USA

6 Institute for Tropical Medicine, Antwerp, Belgium

References

[1] Harbach RE. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. Bulletin of Entomological Research 2004; 94(6) 537-553.

[2] Harbach R. Mosquito Taxonomic Inventory. 2012. http://mosquito-taxonomic-inventory.info /sites/mosquito-taxonomic-inventory.info/files/Anopheles%20classification_25.pdf

[3] Manguin S, Carnevale P, Mouchet J, Coosemans M, Julvez J, Richard-Lenoble D, Sircoulon J. Biodiversity of malaria in the world. John Libbey; 2008.

[4] The malERA Consultative Group on Vector Control. A research agenda for malaria eradication: Vector control. PLoS Medicine 2011; 8(1):e1000401.

[5] Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, et al. A World malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Medicine 2009; 6(3):e1000048.

[6] Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, Gething PW, Elyazar IRF, Kabaria CW, Harbach RE, Hay SI. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. Parasites & Vectors 2011; 4:89.

[7] Sinka M, Bangs M, Manguin S, Coetzee M, Mbogo C, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW, Okara RM, Van Boeckel T, Godfray HCJ, Harbach RE, I Hay S. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. Parasites & Vectors 2010; 3 (1) 117.

[8] Sinka M, Rubio-Palis Y, Manguin S, Patil A, Temperley W, Gething P, Van Boeckel T, Kabaria CW, Harbach RE, Hay SI. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. Parasites & Vectors 2010; 3 (1) 72.

[9] Obsomer V, Defourny P, Coosemans M. Predicted distribution of major malaria vectors belonging to the *Anopheles dirus* Complex in Asia: ecological niche and environmental influences. PLoS ONE 2012; 7(11):e50475.

[10] Ceccato P, Connor SJ, Jeanne I, Thomson MC. Application of geographical information systems and remote sensing technologies for assessing and monitoring malaria risk. Parassitologia 2005;47 (1) 81-96.

[11] Hill D, Fasham M, Tucker G, Shewry M, Shaw P. Handbook of Biodiversity Methods: Survey, Evaluation and Monitoring. Cambridge University Press 2005.

[12] Brotons L, Herrando S, Pla M. Updating bird species distribution at large spatial scales: applications of habitat modelling to data from long-term monitoring programs. Diversity and Distributions 2007; 13 (3) 276-88.

[13] Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. Ecology Letters 2005; 8 (9) 993-1009.

[14] Elith J, Graham C, Anderson R, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JMM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. Novel methods improve prediction of species' distributions from occurrence data. Ecography 2006; 29 (2) 129-151.

[15] Brotons L, Sierdsemama H, Newson S, Jiguet F, Gregory R. Spatial modeling of large-scale bird monitoring data: Towards pan-European quantitative distribution maps. Journal of Ornithology. 2006; 147:29.

[16] Kéry M, Dorazio RM, Soldaat L, Van Strien A, Zuiderwijk A, Royle JA. Trend estimation in populations with imperfect detection. Journal of Applied Ecology 2009; 46 (6) 1163-72.

[17] Pearce J, Ferrier S. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modelling 2000; 128 (2–3) 127-147.

[18] Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, et al. Effects of sample size on the performance of species distribution models. Diversity and Distributions 2008; 14 (5) 763-773.

[19] iménez-Valverde A, Lobo JM, Hortal J. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Community Ecology 2009; 10(2) 196-205.

[20] Feeley KJ, Silman MR. Keep collecting: accurate species distribution modelling requires more collections than previously thought. Diversity and Distributions 2011; 17(6) 1132-1140.

[21] Hirzel A, Guisan A. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling 2002; 157 (2–3) 331-341.

[22] Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecological Modelling 2006; 190 (3–4) 231-259.

[23] Mackenzie DI, Royle JA. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 2005; 42 (6) 1105-1114.

[24] MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. 1er éd. Academic Press 2005.

[25]   Beck LR, Rodriguez MH, Dister SW, Rodriguez AD, Rejmankova E, Ulloa A, et al. Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. American Journal of Tropical Medicine and Hygiene 1994; 51 (3) 271-280.

[26]   USGS United States Geological Survey. Hydro 1km, elevation derivative database. Sioux Falls, South Dakota, USA. 2001

[27]   CGIAR CIAT International Center for Tropical Agriculture. Void-filled seamless SRTM data VI. 2004.

[28]   FAO United Nation Food and Agriculture Organisation. Digital soil map of the world 1990.

[29]   Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpo-lated climate surfaces for global land areas. International Journal of Climatology 2005; 25 (15) 1965-1978.

[30]   New M, Lister D, Hulme M, Makin I. A high-resolution dataset of surface climate over global land areas. Climate research 2002; 21:1-25.

[31]   Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Under-wood EC, et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth. BioScience 2001; 51(11) 933-938.

[32]   Price JC. Land surface temperature measurements from the split window channels of the NOAA 7 Advanced Very High Resolution Radiometer. Journal of Geophysical Research 1984; 89 (D5): 7231-7237.

[33]   Wan Z, Zhang Y, Zhang Q, Li Z. Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote Sensing of Environment 2002; 83 (1–2) 163-80.

[34]   Pinheiro ACT, Mahoney R, Privette JL, Tucker CJ. Development of a daily long term record of NOAA-14 AVHRR land surface temperature over Africa. Remote Sensing of Environment 2006; 103 (2) 153-164.

[35]   Vancutsem C, Ceccato P, Dinku T, Connor SJ. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. Re-mote Sensing of Environment 2010; 114 (2) 449-465.

[36]   Ceccato P, Vancutsem C, Temimi M. Monitoring air and Land Surface Temperatures from remotely sensed data for climate-human health applications. Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International. 2010. page 178 -180.

[37]   Ceccato P, Dinku T. Introduction to remote sensing for monitoring rainfall, tempera-ture, vegetation and water bodies. Palisades, New York: International Research Insti-tute for Climate and Society; 2010 page 15 pp. Report No.: 10-04.

[38]   Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, Hong Y, Bowman KP, Stocker EF. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global,

Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. Journal of Hydrometeorology 2007; 8 (1) 38-55.

[39] Dinku T, Chidzambwa S, Ceccato P, Connor SJ, Ropelewski CF. Validation of high-resolution satellite rainfall products over complex terrain. International Journal of Remote Sensing 2008; 29(14) 4097-4110.

[40] Joyce RJ, Janowiak JE, Arkin PA, Xie P. CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. Journal of Hydrometeorology 2004; 5:487.

[41] Dinku T, Ceccato P, Cressman K, Connor SJ. Evaluating Detection Skills of Satellite Rainfall Estimates over Desert Locust Recession Regions. Journal of Applied Meteorology and Climatology 2010; 49 (6) 1322-1332.

[42] Dinku T, Ceccato P, Connor SJ. Challenges of satellite rainfall estimation over mountainous and arid parts of east Africa. International Journal of Remote Sensing 2011; 32 (21) 5965-5979.

[43] Dinku T, Connor S, Ceccato P. Evaluation of Satellite Rainfall Estimates and Gridded Gauge products over the Upper Blue Nile Region. In: Melesse AM, (ed.) Nile River Basin. Part II.; 2011. p109-127.

[44] Maidment RI, Grimes DIF, Allan RP, Greatrex H, Rojas O, Leo O. Evaluation of satellite-based and model re-analysis rainfall estimates for Uganda. Meteorological Applications 2012; ISSN 1469-8080 (in press)

[45] Heinemann T, Kenényi J. The Eumestsat multi sensor precipitation estimate (MPE): concept et validation. Proceedings of the EUMETSAT users conference. Weimar, Germany; 2003.

[46] Tucker CJ, Sellers PJ. Satellite remote sensing of primary production. International Journal of Remote Sensing 1986; 7 (11) 1395-1416.

[47] Myneni RB, Hall FG, Sellers PJ, Marshak AL. The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience and Remote Sensing 1995; 33 (2)481-486.

[48] Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring Vegetation Systems in the Great Plains with Erts. NASA Special Publication 1974; 351:309.

[49] Tucker CJ, Pinzon JE, Brown ME, Slayback DA, Pak EW, Mahoney R, Vermote, EF, El Saleous, N. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing 2005; 26(20) 4485-4498

[50] Huete R. Soil influence in remote sensed vegetation-canopy spectra. Introduction to the physics and techniques of remote sensing. Wiley-interscience. New York; 1987. page 107-41.

[51] Jackson TJ, Chen D, Cosh M, Li F, Anderson M, Walthall C, Doriaswamy P, Hunt ER. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. Remote Sensing of Environment 2004; 92(4) 475-82.

[52] Huete A, Didan K, Miura T, Rodriguez E., Gao X, Ferreira L. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 2002; 83(1–2) 195-213.

[53] Jiang Z, Huete AR, Didan K, Miura T. Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment 2008; 112(10) 3833-3845.

[54] Gao B. NDWI: A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 1996; 58 (3) 257-266.

[55] Ceccato P, Gobron N, Flasse S, Pinty B, Tarantola S. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1: Theoretical approach. Remote Sensing of Environment. 2002; 82 (2–3) 188-97.

[56] Zarco-Tejada P., Rueda C., Ustin S. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. Remote Sensing of Environment 2003; 85 (1) 109-24.

[57] Pekel J-F, Ceccato P, Vancutsem C, Cressman K, Vanbogaert E, Defourny P. Development and Application of Multi-Temporal Colorimetric Transformation to Monitor Vegetation in the Desert Locust Habitat. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2011; 4(2) 318 -326.

[58] Chen JM, Black TA. Defining leaf area index for non-flat leaves. Plant, Cell & Environment 1992; 15(4): 421-429.

[59] Myneni RB, Maggion S, Iaquinta J, Privette JL, Gobron N, Pinty B, Kimes DS, Verstraete MM, Williams DL. Optical remote sensing of vegetation: Modeling, caveats, and algorithms. Remote Sensing of Environment 1995; 51(1) 169-188.

[60] Dorigo WA, Zurita-Milla R, de Wit AJW, Brazile J, Singh R, Schaepman ME. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation and Geoinformation 2007; 9(2) 165-193.

[61] Myneni R., Hoffman S, Knyazikhin Y, Privette J., Glassy J, Tian Y, Wang Y, Song X, Zhang Y, Smith GR, Lotsch A, Friedl M, Morisette JT, Votava P, Nemani RR, Running SW. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sensing of Environment 2002; 83 (1–2) 214-31.

[62] Baret F, Hagolle O, Geiger B, Bicheron P, Miras B, Huc M, Berthelot B, Niño F, Weiss M, Samain O, Roujean J-L, Leroy M. LAI, fAPAR and fCover CYCLOPES global

products derived from VEGETATION: Part 1: Principles of the algorithm. Remote Sensing of Environment 2007; 110 (3) 275-86.

[63] Weiss M, Baret F, Garrigues S, Lacaze R. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. Remote Sensing of Environment 2007; 110 (3) 317-331.

[64] Garrigues S, Lacaze R, Baret F, Morisette JT, Weiss M, Nickeson JE, Fernandes R, Plummer S, Shabanov NV, Myneni RB, Knyazikhin Y, Yang W. Validation and inter-comparison of global Leaf Area Index products derived from remote sensing data. Journal of Geophysical Research 2008; 113(G2):G02028.

[65] Fang H, Wei S, Liang S. Validation of MODIS and CYCLOPES LAI products using global field measurement data. Remote Sensing of Environment 2012; 119(0) 43-54.

[66] Baret F, Weiss M. Biopar Methods Compendium LAI, FAPAR, FCOVER, NDVI. Geo-land2 rep. BP-RP038. 2010. 50pp.

[67] Baret F, Pacholczyk P, Lacaze R. BioPar Product User Manual: LAI, FAPAR, FCover from AVHRR-LTDR data. Geoland 2 rep BR-RP053.2012. Geoland 2 rep BR-RP053.2012; 2012.

[68] Latham J. Africover eastern Africa. LUUC NewsLetter 2001; (7) 15-6.

[69] Bartholomé E, Belward AS. GLC2000: a new approach to global land cover mapping from Earth observation data. International Journal of Remote Sensing 2005; 26 (9)1959-1977.

[70] Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, et al. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment 2010; 114 (1) 168-182.

[71] Defourny P, Vancutsem C, Pekel J-F, Bicheron P, Brockmann C, Nino F, Schouten L, Leroy M. Towards a 300 m global land cover product - the Globcover initiative. Snd workshop of the EARSeL Special Interest Group on land use and land cover: application & development. Bonn (Germany): Braun M; 2008.

[72] Giri C, Jenkins C. Land cover mapping of Greater Mesoamerica using MODIS data. Canadian Journal of Remote Sensing 2005; 31(4) 274-282.

[73] Bontemps S, Arino O, Bicheron P, Carsten B, Leroy M, Vancutsem C, Defourny P. Operational service demonstration for global land cover mapping: the GlobCover and GlobCorine experience for 2005 and 2009. Remote sensing of land use and land cover: principle and applications. 2012.

[74] Haas EM, Bartholomé E, Combal B. Time series analysis of optical remote sensing data for the mapping of temporary surface water bodies in sub-Saharan western Africa. Journal of Hydrology 2009; 370(1–4) 52-63.

[75] Gond V, Bartholomé E, Ouattara F, Nonguierma A, Bado L. Surveillance et cartographie des plans d'eau et des zones humides et inondables en régions arides avec l'in-

strument VEGETATION embarqué sur SPOT-4. International Journal of Remote Sensing 2004; 25(5) 987-1004.

[76] Pekel J-F, Vancutsem C, Bastin L, Clerici M, Vanbogaert E, Bartholomé E, et al. A water surface detection method based on hSV transformation of MODIS multi-spetral time series suitable for near real-time monitoring in Africa. Remote Sensing of Environment (submitted). 2013

[77] Holben BN. Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing 1986; 7(11) 1417-1434.

[78] Schaaf CB, Gao F, Strahler AH, Lucht W, Li X, Tsang T, Strugnella NC, Zhanga X, Jina Y, Mullerc J-P, Lewis P, Barnsley M, Hobson P, Disney M, Roberts G, Dunderdale M, Doll C, d'Entremont RP, Hug B, Liangh S, Privettei JL, Royh D. First operational BRDF, albedo nadir reflectance products from MODIS. Remote Sensing of Environment 2002; 83 (1-2) 135-148.

[79] Hagolle O, Lobo A, Maisongrande P, Cabot F, Duchemin B, De Pereyra A. Quality assessment and improvement of temporally composited products of remotely sensed imagery by combination of VEGETATION 1 and 2 images. Remote Sensing of Environment 2005; 94 (2) 172-86.

[80] Vancutsem C, Pekel J -F., Bogaert P, Defourny P. Mean Compositing, an alternative strategy for producing temporal syntheses. Concepts and performance assessment for SPOT VEGETATION time series. International Journal of Remote Sensing 2007; 28(22) 5123-5141.

[81] Duveiller G, Baret F, Defourny P. Remotely sensed green area index for winter wheat crop monitoring: 10-Year assessment at regional scale over a fragmented landscape. Agricultural and Forest Meteorology 2012; 166–167 (0) 156-168.

[82] Cracknell AP. Review article Synergy in remote sensing-what's in a pixel? International Journal of Remote Sensing 1998; 19(11) 2025-47.

[83] Schowengerdt. Remote sensing models and methods for image processing. San Diego, USA: San Diego Academic Press; 2007.

[84] Wolfe RE, Roy DP, Vermote E. MODIS land data storage, gridding, and compositing methodology: Level 2 grid. IEEE Transactions on Geoscience and Remote Sensing 1998; 36(4) 1324 -1338.

[85] Tan B, Woodcock CE, Hu J, Zhang P, Ozdogan M, Huang D, Yang W, Knyazikhin Y, Myneni RB. The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. Remote Sensing of Environment 2006; 105(2) 98-114.

[86] Duveiller G. Caveats in calculating crop specific pixel purity for agricultural monitoring using MODIS time series. Proc. SPIE 8531, Remote Sensing for Agriculture, Ecosystems, and Hydrology XIV, 85310J (October 19, 2012).

[87] Duveiller G, Baret F, Defourny P. Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. Remote Sensing of Environment 2011; 115 (10) 2686-2701.

[88] Hansen MC, DeFries RS, Townshend JRG, Carroll M, Dimiceli C, Sohlberg RA. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm. Earth Interactions 2003; 7 (10) 1-15.

[89] Overgaard HJ, Ekbom B, Suwonkerd W, Takagi M. Effect of landscape structure on anopheline mosquito density and diversity in northern Thailand: Implications for malaria transmission and control. Landscape ecology 2003; 18 (6) 605-619.

[90] Guisan A, Zimmermann NE. Predictive habitat distribution models in ecology. Ecological Modelling 2000; 135 (2–3) 147-186.

[91] Brotons L, Thuiller W, Araújo MB, Hirzel AH. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 2004; 27(4) 437-448.

[92] Araújo MB, Guisan A. Five (or so) challenges for species distribution modelling. Journal of Biogeography 2006; 33 (10) 1677-1688.

[93] Franklin J. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press; 2010.

[94] Thomson M, Obsomer V, Dunne M, Connor S, Molyneux D. Satellite mapping of *Loa loa* prevalence in relation to ivermectin use in west and central Africa. The Lancet 2000; 356 (9235) 1077-1078.

[95] Rogers DJ. Models for vectors and vector-borne diseases. Advances in Parasitology 2006; 62: 1-35.

[96] Titeux N, Dufrêne M, Jacob J-P, Paquay M, Defourny P. Multivariate analysis of a fine-scale breeding bird atlas using a geographical information system and partial canonical correspondence analysis: environmental and spatial effects. Journal of Biogeography 2004; 31 (11) 1841-1856.

[97] Hutchinson G. Concluding remarks 1957. Cold Spring Harbor Symposia on Quantitative Biology. p415-427.

[98] Busby J. BIOCLIM - A bioclimate analysis and prediction system. In: Margules C, Austin M,. Nature conservation: cost effective biological surveys and data analysis. Melbourne: CSIRO; 1991.

[99] Carpenter G, Gillison AN, Winter J. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation 1993; 2 (6) 667-80.

[100] Hirzel AH, Hausser J, Chessel D, Perrin N. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 2002; 83(7) 2027-2036.

[101] Soberón J. Grinnellian and Eltonian niches and geographic distributions of species. Ecology Letters 2007; 10 (12) 1115-1123.

[102] Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 2009; 19(1) 181-197.

[103] Pearson RG, Raxworthy CJ, Nakamura M, Townsend Peterson A. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 2007; 34(1) 102-117.

[104] Rascalou G, Pontier D, Menu F, Gourbière S. Emergence and Prevalence of Human Vector-Borne Diseases in Sink Vector Populations. PLoS ONE. 2012; 7(5):e36858.

[105] Mouchet J, Manguin S, Sircoulon J, Laventure S, Faye O, Onapa AW, Carnevale P, Julvez J, Fontenille D. Evolution of malaria in Africa for the past 40 years: impact of climatic and human factors. Journal of the American Mosquito Control Association 1998; 14 (2) 121-130.

[106] Bouma MJ. Methodological problems and amendments to demonstrate effects of temperature on the epidemiology of malaria. A new perspective on the highland epidemics in Madagascar, 1972–1989. Transactions of the Royal Society of Tropical Medicine and Hygiene 2003; 97 (2) 133-139.

[107] WHO/UNICEF. The World Malaria Report 2010. 2011.

[108] Omumbo JA, Lyon B, Waweru SM, Connor SJ, Thomson MC. Raised temperatures over the Kericho tea estates: revisiting the climate in the East African highlands malaria debate. Malaria Journal. 2011; 10(1)12.

[109] Reiter P. Global warming and malaria: knowing the horse before hitching the cart. Malaria Journal 2008; 7(Suppl 1): S3

[110] Thomson MC, Mason SJ, Phindela T, Connor SJ. Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. The American journal of tropical medicine and hygiene 2005; 73(1): 214-221.

[111] Abeku TA, Hay SI, Ochola S, Langi P, Beard B, de Vlas SJ, Cox J. Malaria epidemic early warning and detection in African highlands. Trends in Parasitology 2004; 20 (9):400-405.

[112] Hay SI, Were EC, Renshaw M, Noor AM, Ochola SA, Olusanmi I, Alipui N, Snow RW. Forecasting, warning, and detection of malaria epidemics: a case study. The Lancet 2003; 361(9370):1705-1706.

[113] Hay SI, Shanks GD, Stern DI, Snow RW, Randolph SE, Rogers DJ. Climate variability and malaria epidemics in the highlands of East Africa. Trends in Parasitology 2005; 21(2):52-53.

[114] Palmer TN, Doblas-Reyes FJ, Hagedorn R, Alessandri A, Gualdi S, Andersen U, Feddersen H, Cantelaube P, Terres J-M, Davey M, Graham R, Délécluse P, Lazar A, Déqué M, Guérémy J-F, Díez E, Orfila B, Hoshen M, Morse AP, Keenlyside N, Latif M, Maisonnave E, Rogel P, Marletto V, Thomson MC. Development of an European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bulletin of the American Meteorological Society 2004; 85 (6) 853-72.

[115] Thomson MC, Doblas-Reyes FJ, Mason SJ, Hagedorn R, Connor SJ, Phindela T, Morse AP, Palmer TN. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature 2006; 439(7076) 576-9.

[116] Protopopoff N, Van Bortel W, Speybroeck N, Van Geertruyden J-P, Baza D, D'Alessandro U, Coosemans M. Ranking Malaria Risk Factors to Guide Malaria Control Efforts in African Highlands. PLoS ONE 2009; 4:e8022.

[117] Garret-Jones C. Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. Nature 1964; 204: 1173–1175.

[118] Macdonald G. Theory of the eradication of malaria. Bulletin of the World Health Organisation 1956; 15: 369–387.

[119] Dye C. The Analysis of parasite transmission by bloodsucking insects. Annual Review of Entomology 1992; 37 (1) 1-19.

[120] Ceccato P, Vancutsem C, Klaver R, Rowland J, Connor SJ. A Vectorial Capacity product to monitor changing malaria transmission potential in epidemic regions of Africa. Journal of Tropical Medicine 2012; 2012 (595948) 1-6.

[121] Lactin DJ, Holliday NJ, Johnson DL, Craigen R. Improved rate model of temperature-dependent development by arthropods. Environmental entomology 1995; 24(1) 68-75.

[122] Bruce-Chwatt L. Essential malariology. London, UK: Heinemann medical books; 1991.

[123] Bouma M, Sondorp H, van der Kaay H. Health and climate change. The Lancet 1994; 343 (8892) 302-304.

[124] Roberts DR, Rodriguez MH. The environment, remote sensing, and malaria control. Annals of the New York Academy of Sciences 1994; 15 (740) 396-402.

[125] Obsomer V, Defourny P, Coosemans M. The *Anopheles dirus* complex: spatial distribution and environmental drivers. Malaria Journal 2007; 6(1):26.

[126] Travis P, Bennett S, Haines A, Pang T, Bhutta Z, Hyder AA, et al. Overcoming health-systems constraints to achieve the Millennium Development Goals. The Lancet 2004; 364 (9437) 900-906.

[127] WHO. A global strategy for malaria control. Geneva: World Health Organisation; 1993.

[128] WHO. The Abuja declaration on Roll Back Malaria. Geneva: World Health Organisation; 2000.

[129] Grambsch A, Menne B. Adaptation and adaptive capacity in the public health context. Climate change and human health. Geneva: World Health organisation; 2003. p220-236.

[130] Ceccato P, Connor S. Early Warning Systems. In: Morain S, Budge A. (ed.) Environmental tracking for public health surveillance. International society for photogrammetry and remote sensing (ISPRS) Commission VIII/WG-2; 2012.

[131] Dufrene M, Legendre P. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs 1997; 1(67)3: 345-66.

[132] Obsomer V, Dufrene M, Defourny P, Coosemans M. *Anopheles* species associations in Southeast Asia: indicator species and environmental influences. Parasites & vectors (submitted).

[133] Diggle P, Thomson M, Christensen O, Rowlingson B, Obsomer V, Gardon J, et al. Spatial modeling and prediction of *Loa loa* risk: decision making under uncertainty. Annals of Tropical Medicine and Parasitology 2007; 101(6) 499-509.

[134] Pineros Garcet J. Quality assessment of environmental simulation scenarios: Scientific, expertise, interpretive and participatory methods based on mathematical, epistemological and socio-anthropological approaches. PhD thesis. Université catholique de Louvain, Louvain-la-Neuve; 2012.

[135] Nordstrand A, Bunikis I, Larsson C, Tsogbe K, Schwan TG, Nilsson M, et al. Tick-borne relapsing fever diagnosis obscured by malaria, Togo. Emerging Infectious Diseases 2007; 13(1) 117-23.