

Chapter 7

Geospatial Environmental Data for Planetary Health Applications



Michael C. Wimberly

Abstract Planetary health research focused on vector-borne and zoonotic diseases often requires data on the environmental factors that influence vectors, hosts, and pathogens. We summarized major types of geospatial environmental data that are freely available and potentially useful for planetary health applications. There are many relevant geospatial data products that characterize weather, climate, vegetation, land surface temperature, land cover and land use, human population characteristics, and hydrology. However, these datasets differ greatly in their underlying measurement techniques and spatial and temporal resolutions. Although many datasets have global coverage, they vary considerably in their spatial accuracy and suitability for local applications. We recommend that researchers carefully consider the strengths and limitations of alternative data sources with a particular focus on the spatial and temporal scales of the data relative to the specific organisms and processes of interest. Research that addresses the sensitivities of analytical results and model predictions to alternative data sources can provide additional guidance to inform these decisions.

Keywords Climate and weather · Land cover and land use · Geographic information systems · Remote sensing

7.1 Introduction

The field of planetary health addresses myriad interconnections between global environmental change and the health of humans, animals, and the ecosystems they inhabit [1]. It shares this conceptual foundation with related interdisciplinary fields such as One Health [2], EcoHealth [3], and GeoHealth [4]. All are based on a holistic framework that emphasizes the relationships between human health, the social environment, the physical environment, and the non-human organisms that are hosts and vectors for disease-causing pathogens. Because of this breadth, there is a need for diverse sources of data to characterize multiple aspects of human and natural systems.

M. C. Wimberly (✉)

Department of Geography and Environmental Sustainability, University of Oklahoma, Norman, OK 73061, USA

e-mail: mcwimberly@ou.edu

Geospatial data that map the spatial patterns of relevant phenomena are particularly important for assessing spatial relationships and identifying hot spots with high risk of disease transmission.

Vector-borne and zoonotic diseases are particularly sensitive to features of the physical environment that influence the reproduction, growth, and survival of vectors, hosts, and pathogens. Climate, weather, water, vegetation, and land use influence transmission cycles through their effects on vector and host habitats, pathogen development and transmission, and human exposure to vectors [5–7]. To conduct research on these diseases and translate the results into applications, it is essential to measure the relevant environmental variables. Accurate and timely data are needed to test hypotheses about drivers of disease transmission, develop maps of infectious disease risk based on environmental factors, and forecast future disease risk resulting from changes in weather and climate. Even when the research is focused on other questions, such as the effectiveness of public health interventions, it is still necessary to control for background effects of environmental variation on spatial and temporal patterns of disease transmission [8].

These environmental factors are heterogeneous at multiple spatial and temporal scales. Broad climate gradients vary geographically with latitude and elevation and change gradually over decades. Within a given climate, weather fluctuates continuously and exhibits diurnal, seasonal, and interannual cycles. More localized patterns related to vegetation, topography, and human land use vary at spatial scales from hundreds of meters to hundreds of kilometers and change over time scales from years to decades. These landscape features create microclimates that differ considerably from the broader macroclimate, and these local conditions can facilitate disease transmission even when the broader macroclimate is unsuitable [9, 10]. When selecting the environmental data for a planetary health application, it is essential to understand the scales of environmental measurements and match them with the specific ecological and epidemiological processes of interest.

There are numerous geospatial data products that characterize a variety of environmental characteristics. Many of these products are updated regularly and are available at continental to global extents, providing opportunities for widespread use in planetary health. However, the underlying data are collected over a wide range of spatial and temporal scales. Measurement techniques and the resulting accuracies also vary among data products, as do the techniques used for spatial interpolation and filling of data gaps. All of these factors can affect inferences about environment-disease relationships and the accuracy of predictive models based on these relationships. The goal of this chapter is to summarize the main environmental data sources that have been used in planetary health applications related to vector-borne and zoonotic diseases. Strengths and limitations of various data products are highlighted and emerging trends are discussed.

7.2 Meteorological Data

Ground-based meteorological stations provide in situ observations of weather, and long-term summaries of these data are the basis for measuring climate and tracking climate change. Standard variables monitored at weather stations include air temperature, precipitation, humidity, atmospheric pressure, wind speed, and solar radiation. A critical objective in designing and siting weather stations is ensuring consistent observations that can be compared over time and between different locations [11]. Meteorological stations are therefore located in open areas where measurements are not influenced by buildings or tall vegetation. Instruments are enclosed to protect them from direct solar radiation, condensation, and precipitation while allowing sufficient ventilation to facilitate airflow over the sensors. Because of the expense of installation, equipment maintenance, and data curation, meteorological stations have historically been operated by government agencies [12]. However, volunteer observers are also an important part of the enterprise through programs like the National Weather Service (NWS) Cooperative Observer Program, and the availability of low-cost digital home weather stations has allowed private citizens to provide crowdsourced weather observations [13].

Station data are typically regarded as the gold standard for near-surface observations of weather and climate [14]. However, the types of instruments, frequencies of measurements, and completeness of the resulting data all vary between stations and over time. In general, high-income countries have well developed weather monitoring systems with higher densities of stations and more technologically advanced equipment and data infrastructure than lower-income countries in the Global South [15]. Even in countries with highly resourced weather monitoring infrastructures, most of the places where disease transmission occurs are located relatively far from extant weather stations. Thus, an important issue is determining the degree to which distant weather stations are representative of the environments that directly influence disease transmission cycles.

One way to obtain more spatially precise estimates of local weather and climate is to interpolate the point data collected at meteorological stations (Fig. 7.1). This approach involves predicting meteorological variables at unsampled locations based on the spatial pattern of nearby measurements. In some cases, ancillary variables that are strongly associated with climate gradients, such as elevation, are incorporated to increase the accuracy and precision of local measurements. Commonly used techniques include various types of regression, kriging, self-organizing maps, and thin-plate splines [16–18]. In most cases, data users do not need to carry out this interpolation themselves, as there are many free gridded weather and climate products produced by various institutions. Meteorological variables can also be extracted from reanalysis data sets, which are generated using data assimilation methods that combine multiple sources of historical weather data with numerical weather models [19].

Although many of these products appear similar, there are underlying differences in the methods used to generate the data and the characteristics of the resulting

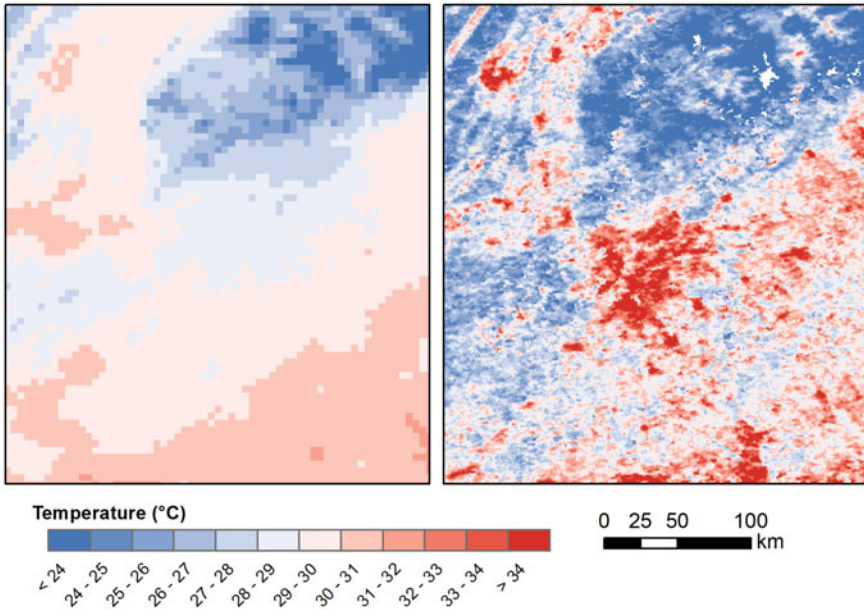


Fig. 7.1 Comparison of two temperature datasets for north Georgia, USA, in June 2020. Left: PRISM interpolated monthly maximum near-surface air temperature (4 km cell size). Right: MODIS Aqua daytime land surface temperature 8-day composite from June 10–17 (1 km cell size)

meteorological grids that can influence results when they are used for planetary health applications [20, 21]. Gridded meteorological and climate data vary considerably in their spatial and temporal scales. For example, the University of East Anglia Climate Research Unit (CRU) datasets provide global historical monthly time series and climatologies at a grid cell size of 0.5° (approximately 55 km) [22]. The Climate Hazards Group Coupled Infrared Precipitation with Stations (CHIRPS) [23] and Temperature with Stations (CHIRTS) [24] datasets combine interpolated station data with satellite estimates of precipitation and land surface temperature to produce daily and monthly estimates at a much smaller grid cell size of 0.05° (approximately 5.5 km). Other downscaled climate data products like WorldClim [25] and Chelsa [26] use high-resolution elevation data to downscale climate grids to a cell size of 30 arc seconds (approximately 1 km).

Differences in the methods used for interpolation and downscaling lead to variations in the meteorological grids that can influence results when they are used for planetary health applications [19, 20]. There are also trade-offs between dataset attributes such as grid cell size, frequency of measurement, and the time required to process the data and make them available. In the United States, the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) climate dataset provides monthly meteorological data at a spatial resolution of 800 m [27]. In contrast, the National Land Data Assimilation System (NLDAS) forcings dataset provides many of the

same meteorological variables on a 30-min time step with a latency of several days, but the grid cell size is 0.125° (approximately 14 km) [28]. The GridMET dataset combines these two data sources to provide gridded meteorological variables at a spatial resolution of 4 km and a daily time step [29].

7.3 Satellite Vegetation Indices

Earth-observing satellites are another source of geospatial environmental data that can be used to predict spatial and temporal patterns of infectious disease transmission [30–32]. Unlike the point-level data obtained from weather stations, satellite images provide spatially continuous measurements over large areas of the Earth’s surface and are repeated at intervals ranging from days to weeks. They are fundamentally different from weather station data in that they typically measure conditions on the land surface, not in the near-surface atmosphere. The most commonly used satellite remote sensing data are observations of reflected solar radiation in the visible and infrared wavelengths. These data are measured as radiance or reflectance in one or more spectral bands, where each band encompasses a specific range of wavelengths. These bands are then used to calculate spectral indices that characterize physical properties of the Earth’s surface.

The most common spectral index is the normalized difference vegetation index (NDVI, Fig. 7.2), which measures green vegetation using red and near infrared spectral bands [33]. In most cases, vegetation greenness itself is not a proximal driver of disease transmission. However, the NDVI index is highly sensitive to meteorological factors such as temperature and precipitation [34]. In temperate environments, NDVI changes in response to vegetation greenup in the spring and senescence in the fall and can provide information about timing and length of disease transmission seasons [35, 36]. In water-limited environments, NDVI is sensitive to rainfall (Fig. 7.3) and can be an indicator of water availability and drought [37]. Several variations of the NDVI have been developed to improve greenness estimates in particular situations. For example, the enhanced vegetation index (EVI) was developed to mitigate issues with index saturation in dense forests [38], and the soil-adjusted vegetation index (SAVI) was designed to correct for effects of soil brightness in areas with low vegetation cover [39].

A major advantage of NDVI is that the necessary data are widely available over long time periods for nearly every location on Earth. The NDVI can be calculated using data from a variety of satellite sensors, which provide data at different spatial and temporal scales. The earliest applications of satellite imagery for research on vector-borne diseases involved the Advanced Very High Resolution Radiometer (AVHRR), which has been operational on United States National Oceanic and Atmospheric Administration (NOAA) weather satellites since 1981 and provides daily data at pixel sizes of 1000–4000 m [40, 41]. The more recent MODIS sensor, on board the National Aeronautics and Space Administration (NASA) Terra and Aqua Satellites,

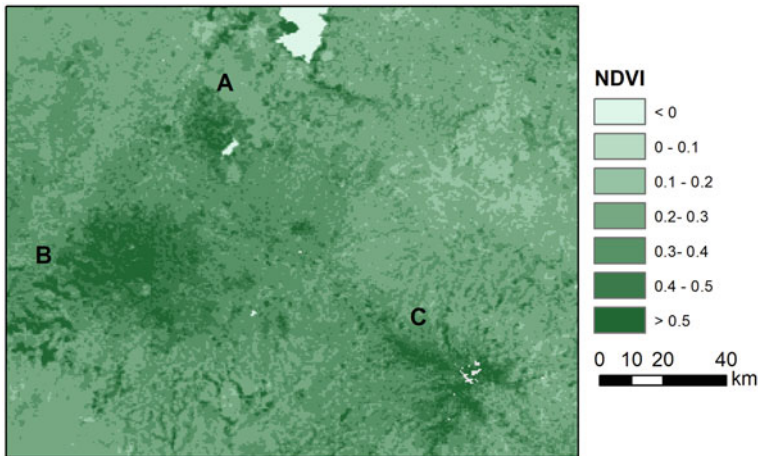


Fig. 7.2 Normalized difference vegetation index (NDVI) for part of the Amhara region of Ethiopia on May 1, 2019. The index was calculated using MODIS BRDF-Adjusted Reflectance Data (500 m cell size). Locations with high vegetation greenness include irrigated agriculture (a), areas with high densities of tree cover (b), and high elevation zones (c)

has provided daily global NDVI data since 2000 at spatial resolutions between 250–1000 m. These data have been widely used to model infectious disease outcomes over relatively large areas when frequent measurement intervals are required [36, 42]. The Visible Infrared Imaging Radiometer Suite (VIIRS) instrument, carried aboard multiple NOAA satellites, also generates daily global estimates of NDVI at spatial resolutions between 500–1000 m and will provide continuity after the end of the MODIS mission. Data from the Landsat and Sentinel missions can be used to derive NDVI at spatial resolutions from 10–30 m with weekly revisit intervals. These data can be applied when higher-resolution environmental measurements are needed for more localized predictions of vector habitats, host habitats, and disease transmission risk [43].

Other advantages of using NDVI and related spectral indices to measure environmental variability include the global availability of satellite imagery and the relatively high spatial resolution of the data compared to the grid size of interpolated meteorological datasets [34]. However, NDVI also has important limitations as an environmental metric for planetary health. NDVI is an indirect environmental measure that is sensitive to multiple environmental factors and the ecological characteristics of the observed vegetation. Therefore, the underlying mechanisms of the relationships between NDVI and disease risk can be obscured, and it is usually not possible to generalize across multiple ecosystems with different landscapes and vegetation. Another major challenge with NDVI is that the underlying visible and infrared imagery is affected by cloud cover [44]. This results in missing data, particularly in cloudy tropical regions, which must be imputed using gap filling techniques or otherwise accounted for in subsequent analyses.

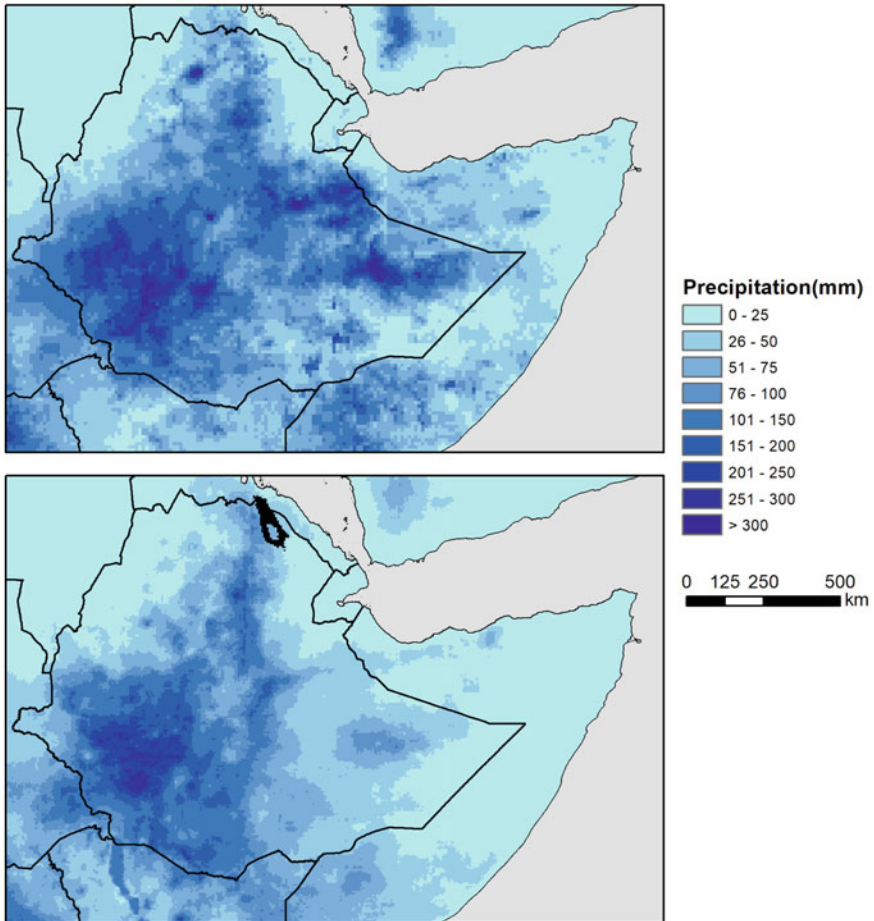


Fig. 7.3 Comparison of two gridded precipitation datasets that combine satellite estimates with ground station data for Ethiopia in March 2019. Top: IMERG (10 km cell size). Bottom: CHIRPS (5.5 km cell size)

7.4 Satellite Land Surface Temperature

Satellite sensors can also measure emitted longwave infrared radiation, which provides information about the temperature of the Earth's surface (Fig. 7.1). Land surface temperature (LST) is a characteristic of the topmost surface layer, which may be vegetation, soil, water, or human-built impervious surfaces depending on the land cover characteristics at a particular location [45]. Importantly, LST measured by satellites is not the same as the near-surface air temperatures (typically 2 m above the

land surface) measured by meteorological stations and represented in gridded meteorological data products. In meteorology, near surface air temperatures are sometimes referred to as simply “surface temperatures”, which can lead to confusion.

LST and near-surface air temperature generally exhibit similar patterns of change over time, including diurnal and seasonal cycles as well as long-term trends [46]. At global and regional scales, LST and near-surface air temperature also follow similar spatial gradients with latitude and elevation. At more localized scales, LST and air temperature usually differ because of the effects of solar radiation, wind, and soil moisture [47]. For example, during the day a paved surface will be warmer than the air above it because it absorbs and re-radiates thermal energy, whereas well-watered vegetation will be cooler because of latent heat loss due to evapotranspiration. At night when there is no incoming solar radiation, land surface temperature and air temperature are usually more similar than during the day [48]. Because of these differences, land surface temperature may not be a precise indicator of the air temperature experienced by organisms above the ground surface or underneath a forest canopy. However, LST is often a reliable proxy for relative variation in air temperature over space and time and can be particularly useful in situations where reliable in-situ measurements of localized temperature are not available.

As with greenness indices, LST data are available from multiple sensors over a range of spatial and temporal resolutions. Daily daytime and nighttime LST estimates are available from MODIS [49] and VIIRS at a grid cell size of 1000 m [50]. Biweekly daytime observations at grid cell sizes of 60–120 m are available from the Thematic Mapper (TM), Enhanced Thematic Mapper (ETM+) and Thermal Infrared Sensor (TIRS) on board Landsat 4–5, 7, and 8–9 respectively [51]. The ECOSystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) sensor provides LST measurements every 4–5 days at a grid cell size of 70 m [52]. All of these sources provide standard data products that estimate LST by combining atmospherically corrected observations of emitted radiation in the thermal wavelengths with measurements of emissivity. These methods are complex and are outside the expertise of most end users in planetary health. However, it is important to recognize that LST estimates can vary depending on the specific method used [53]. Although LST measurements are subject to missing data from cloud cover, the thermal wavelengths used to measure LST are less sensitive to clouds than the shorter-wavelength visible and near infrared bands used to compute greenness indices.

7.5 Satellite Precipitation Estimates

In addition to the interpolated meteorological data products discussed previously, gridded precipitation estimates can also be derived from satellite observations [54]. Satellite precipitation estimates are based on visible/infrared data, passive microwave data, and active microwave (radar) data. Because convecting clouds are usually bright and cold, they can be detected indirectly from their reflectance in the visible and near-infrared wavelengths combined with temperature estimates from thermal

infrared observations. Passive and active microwave observations provide more direct estimates because microwaves can penetrate clouds and are scattered by water droplets and ice particles in the atmosphere. The algorithms used to generate satellite precipitation estimates typically integrate satellite data from multiple sensors.

Planetary health researchers can obtain satellite precipitation data from multiple products, each of which uses different input data sources and estimation algorithms (Fig. 7.3). These products often have relatively coarse grid cell sizes, with measurements taken hourly and made available almost immediately. The satellite data can be combined with ground data from weather stations to improve the estimates. For example, the NASA IMERGE product provides global precipitation estimates at a 10 km grid cell size at a time step of 30 min [55]. It includes “Early Run” and “Late Run” datasets that are based only on satellite data and have latencies of less than one day, and a “Final Run” dataset that incorporates station data from the Global Precipitation Climatology Centre but has a latency of >3 months. Other widely-used satellite precipitation data products include Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) [56], Climate Prediction Center Morphing Technique (CMORPH) [57], and the Global Precipitation Climatology Project (GPCP) [58].

7.6 Land Cover and Land Use Change

Land cover encompasses the biophysical characteristics of the Earth’s land surface, including natural and cultivated vegetation, bare soil, human-built impervious surfaces, and water bodies. Land use describes human activities on the land surface, which can range from development and habitation to agricultural practices to nature preservation. Land cover and land use are often related. Locations with a high coverage of impervious surfaces are likely to be residential, commercial, or industrial areas, and agriculture replaces natural vegetative cover with new vegetation consisting of crop plants. However, land cover is not always an indicator of land use. Forest cover, for example, may result from low-density human habitation, forest management for timber production, or land preservation as a park or conservation area. Because satellite remote sensing measures physical characteristics of the land surface, it can be used to generate gridded maps of land cover and monitor changes over time. In some cases, it is also possible to infer information about land use from these land cover characteristics. The resulting data products are often referred to as land cover and land use (LCLU) products.

A large number of LCLU datasets are available at extents ranging from nations or regions to the entire globe. Coarse-grained global LCLU maps with a grid cell size of 500 m and an annual time step have been developed using data from the MODerate Resolution Imaging Spectroradiometer (MODIS) sensors on board NASA’s Aqua and Terra spacecraft, which began collecting data in 2000 [59]. The grid cells are classified into relatively broad land cover types such as deciduous and evergreen forests, grasslands, shrublands, croplands, and built-up areas. Because of the relatively coarse

grid size, many cells are not homogeneous, and instead contain mixtures of multiple LCLU types. An alternative approach is to map LCLU as continuous fields, where the proportion of each grid cell containing a particular LCLU is estimated. For example, the MODIS Vegetation Continuous Fields (VCF) product provides global data on the percentage of tree cover, non-tree vegetation cover, and non-vegetated cover in 250 m grid cells [60]. The Copernicus Global Land Cover fractional cover layers (Fig. 7.4) similarly provide annual fractional cover estimates for a variety of LCLU classes such as trees, shrubs, herbaceous vegetation, crops, bare soil, and built-up areas at a grid cell size of 100 m [61].

Although these datasets provide information about general patterns of LCLU over space and time, planetary health applications often require more detailed information at finer spatial resolutions. For example, research on the habitat associations of vector and host species may require maps of land use practices such as irrigated agriculture or details about the sizes, shapes, and connectivity of habitat fragments. Satellite missions with finer grid cell sizes such as Landsat (30 m multispectral), Sentinel-1

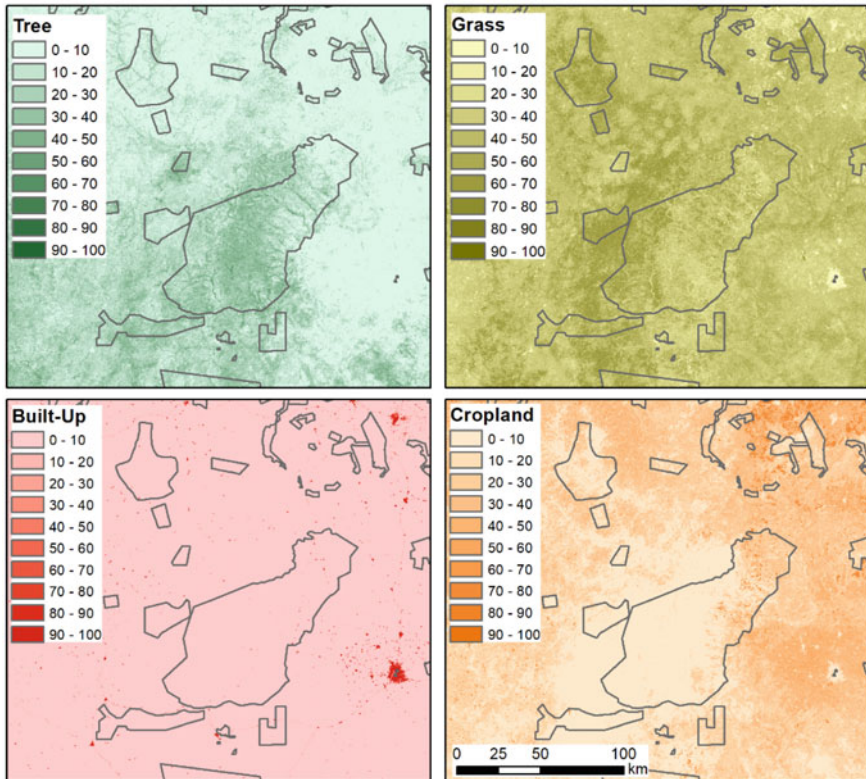


Fig. 7.4 Maps of 2019 land cover in the savanna zone of northern Ghana from the Copernicus Global Land Cover dataset. Land cover of trees, grasses, built-up areas, and croplands is represented as percent cover within 100 m grid cells

(10 m synthetic aperture radar) and Sentinel-2 (10–20 m multispectral) provide data that can be used to generate higher-resolution LCLU products. Many of these datasets are available globally, including data on forest cover and change [62], croplands [63], and cities [64]. However, global availability does not mean that a dataset is well suited for every location across the globe as local accuracy can vary considerably in different locations. In many cases, datasets developed at the regional, national, or local scales may be more accurate and include more relevant LCLU characteristics than global products [65]. These data are often more challenging to discover and access than global products.

7.7 Human Populations

Human population density is a land use characteristic that is particularly important for planetary health research and applications. Data on the human population is needed to calculate the population at risk for epidemiological rates such as incidence and prevalence, and the number of susceptible humans is an important factor influencing the transmission patterns of many infectious diseases. The most common sources of human population data are national censuses, in which people are enumerated within administrative units. Population characteristics can be summarized and mapped within polygons that outline the boundaries of these areas. Although these datasets are produced by individual countries, aggregated global populations datasets such as the Global Rural–Urban Mapping Project (GRUMP) and the Gridded Population of the World (GPW) are also available (Fig. 7.5). These products are published as grids with cell sizes from 1–110 km, but the true spatial resolution of the data is still the administrative unit within which they were aggregated.

It is often desirable to have population data with a finer spatial grain so that urban and rural areas can be distinguished and population density can be estimated for individual settlements or neighborhoods. A common method for generating finer-grained population data is the “top down” approach, in which census data are disaggregated from their administrative units to smaller grid cells using spatial information on land cover, land use, roads, and other factors that are expected to influence population density [66]. These variables are used to calculate a layer of gridded weights that are used to distribute the population within an administrative unit to reflect differences between densely populated urban areas and more sparsely populated rural locations. An alternative is a “bottom up” approach where high-resolution imagery to census individual dwellings are combined with local survey data to estimate population density at a high resolution [67]. The WorldPop project provides a global archive of population and other demographic data products generated using both top-down and bottom-up approaches [68]. LandScan is another widely used gridded population data product that has used multiple sources of satellite imagery and other spatial data to produce annual 1km² global population grids from 2000 to the present [69].

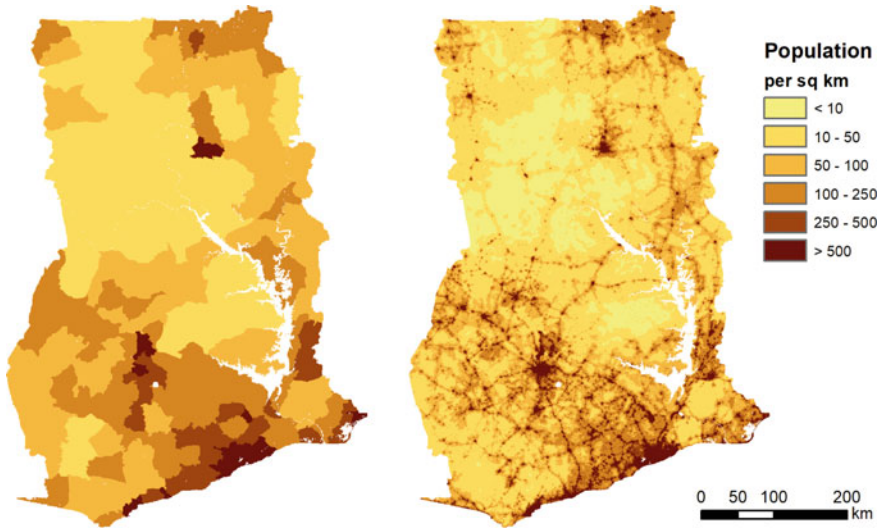


Fig. 7.5 Comparison of two population density datasets for Ghana in 2020. Left: Gridded Population of the World Version 4, which is based on administrative boundaries used for census data collection. Right: WorldPop, which downscales census data based on land cover, roads, and other localized information

7.8 Surface Water and Hydrology

Access to clean water is essential for human health, animal health, and agricultural productivity. However, water also provides habitat for vector and host species and facilitates the transmission of many disease-causing pathogens [70]. Thus, hydrological data are critical for many planetary health applications. Water bodies can be mapped with satellite remote sensing along with other LCLU features, and most LCLU data products include a classification of permanent water bodies such as lakes and large rivers. Understanding how surface water varies over time is also essential. For example, flowing water is not a suitable habitat for vector mosquitoes, but large rivers can provide suitable standing water when their flows decline and leave isolated pools on their floodplains [6]. In flood-prone areas, rising waters are often contaminated by human and animal waste, exposing local populations to a variety of water-borne pathogens [71]. Droughts can also facilitate water-borne pathogen transmission when large groups of people congregate to use the few remaining water sources [72]. More generally, hydrological events like droughts and floods often trigger large-scale human movements and resettlements that facilitate long-distance movement of pathogens and provide novel opportunities for transmission.

Several types of geospatial datasets can provide useful hydrological information for planetary health applications. Gridded elevation datasets are produced for many countries by government mapping agencies, and global elevation products derived from satellite observations are also available. At the most basic level, these data can

be used to identify topographic features such as valley bottoms that are subject to flooding and may serve as locations for water-borne disease transmission or provide larval habitats for mosquitoes [73]. The topographic index, calculated as a function of slope angle and upslope drainage area, is an important input to dynamic hydrological models such as TOPMODEL that can be coupled with mechanistic models of water-associated diseases such as fasciolosis [74]. At coarser grid cell sizes (~10 km or larger) Land Data Assimilations Systems (LDAS) combined gridded meteorological data with other environmental inputs to drive hydrological models that estimate evapotranspiration, soil moisture, and runoff. Various LDAS datasets with different spatial extents, grid cell sizes, and time steps are available, and these data have been used in a variety of disease applications [75–77].

Satellite observations, including passive sensors in the optical and infrared wavelengths and active remote sensing with synthetic aperture radar, can be used to detect and map open water. Surface water is highly variable in locations with pronounced wet and dry seasons, and individual observations are inadequate for characterizing these dynamics. Surface water data products such as the Global Surface Water Explorer [78] and the Global Surface Water Dynamics dataset [79] use time series of Landsat data to map the extent, seasonality, and long-term trends in surface water at 30 m resolution for the entire globe. Global products frequently do not capture smaller water bodies that may serve as larval habitats for mosquitoes or sources of drinking water for humans and livestock. However, they can be used to provide training data for the development of more precise, local maps that include both large and small water bodies [80]. Identifying areas with high seasonal variation in water coverage, including impoundments, wetlands, floodplains, and irrigated areas, is often particularly important in planetary health. These can be identified by analyzing satellite data over multiple seasons and by incorporating topographic variables along with spectral indices [81, 82].

7.9 Synthesis and Conclusions

Researchers and practitioners in planetary health have access to a diverse set of high-quality geospatial data products that characterize environmental factors relevant to human health. Many of these products are global in extent and available at no cost, making them ideal for planetary health assessments and applications in low- and middle-income where locally collected data are sparse. However, it is important to recognize that a global dataset is not necessarily optimal for every location on the Earth [65]. Interpolated meteorological grids and classified LCLU maps have inherent error, and their accuracies can vary considerably among locations. Similarly, the environmental sensitivities of satellite vegetation indices and land surface temperature will vary with the climatic and land surface characteristics in different areas. Before selecting a particular dataset for specific application, potential users should carefully examine the spatial and temporal patterns within their areas of interest to verify that important regional and local features are being captured. If this is not the

case, then it may be necessary to develop bespoke data products that are optimized for the particular region and application [80].

Although planetary health studies frequently incorporate geospatial datasets characterizing climate, LCLU, and human populations, they vary greatly in the specific data used and the manner in which they are applied. A recent systematic review of malaria mapping studies found that the most commonly used covariates were rainfall and temperature [83]. However, the individual studies used a variety of data sources, including ground station measurements, gridded meteorological datasets, satellite vegetation indices, land surface temperature, and satellite precipitation estimates. The degree to which different results are contingent upon differences in the underlying temperature and precipitation data are not well understood. In most cases, the rationale for using a particular source of environmental data is not stated, and decisions are presumably based at least in part on familiarity with particular datasets and ease of data access and use.

This chapter has focused on geospatial environmental data for planetary health applications related to vector-borne and zoonotic infectious diseases. However, geospatial information is also essential for other aspects of planetary health, including natural disasters, food systems and nutrition, and exposure to toxins and pollutants. Timely geospatial data for monitoring meteorological and hydrological variables is essential for monitoring droughts and providing early warning of the risk of food insecurity [84]. Exposure to air pollution is one of the most important global health risks [85], and satellite remote sensing is widely used to obtain spatially explicit measurements of various pollutants [86]. For example, satellite measurements of aerosol optical depth are widely used to estimate ground level concentrations of fine particulate matter generated by combustion of fossil fuels, dust storms and wildfires [87]. There is growing evidence that exposure to greenspace has a variety of health benefits for urban and suburban populations [88], and satellite-based measurements of greenness are commonly used to study these relationships [89]. Although this chapter does not address these topics in detail, many of the data sources that were highlighted in the context of vector-borne and zoonotic diseases are also relevant to these other aspects of planetary health.

Looking forward, more studies on the local accuracies of commonly used geospatial data products would provide evidence to support the choice of geospatial datasets. For example, an accuracy assessment of 20 global precipitation products in Ethiopia found that only three could adequately characterize the spatial extent and severity of historical drought events [90]. An accuracy assessment of multiple gridded climate datasets within the United States found that the most accurate dataset varied by ecoregion [20]. Additionally, comparative analyses of health outcomes based on environmental data from multiple products can help identify the data that are most suitable for specific applications. A West Nile virus risk mapping study compared three predictive models based on land cover and topography data, gridded climate data, and remotely sensed vegetation and moisture indices [91]. Overall accuracy was similar, but the resulting maps based on each dataset exhibited different spatial patterns. A combined model that incorporated variables from all three datasets had the highest overall accuracy. For species range predictions of European tick species, climate

niche models based on an interpolated meteorological dataset had higher accuracies than models based on satellite observations of LST and NDVI from MODIS [92]. Further studies like these will contribute to a broader body of evidence to inform the selection of geospatial environmental datasets for planetary health research and applications.

References

1. Whitmee S, Haines A, Beyrer C, Boltz F, Capon AG, de Souza Dias BF, Ezeh A, Frumkin H, Gong P, Head P (2015) Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation-Lancet Commission on planetary health. *Lancet* 386(10007):1973–2028
2. Zinsstag J, Schelling E, Waltner-Toews D, Tanner M (2011) From “one medicine” to “one health” and systemic approaches to health and well-being. *Prev Vet Med* 101(3–4):148–156
3. Charron DF (2012) Ecohealth: origins and approach. In: Charron DF (ed) *Ecohealth research in practice*. Springer, New York, pp 1–30
4. Gorris ME, Anenberg SC, Goldberg DL, Kerr GH, Stowell JD, Tong D, Zaitchik BF (2021) Shaping the future of science: COVID-19 highlighting the importance of GeoHealth. *GeoHealth*:e2021GH000412
5. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, Rohr JR, Ryan SJ, Savage V, Shocket MS (2019) Thermal biology of mosquito-borne disease. *Ecol Lett* 22(10):1690–1708
6. Smith M, Willis T, Alfieri L, James W, Trigg M, Yamazaki D, Hardy A, Bisselink B, De Roo A, Macklin M (2020) Incorporating hydrology into climate suitability models changes projections of malaria transmission in Africa. *Nat Commun* 11(1):1–9
7. Parham PE, Waldock J, Christophides GK, Hemming D, Augusto F, Evans KJ, Fefferman N, Gaff H, Gumel A, LaDeau S (2015) Climate, environmental and socio-economic change: weighing up the balance in vector-borne disease transmission. *Philos T R Soc B* 370(1665):20130551
8. Thomson MC, Ukawuba I, Hershey CL, Bennett A, Ceccato P, Lyon B, Dinku T (2017) Using rainfall and temperature data in the evaluation of national malaria control programs in Africa. *Am J Trop Med Hyg* 97(Suppl 3):32–45
9. Murdock CC, Evans MV, McClanahan TD, Miazgowicz KL, Tesla B (2017) Fine-scale variation in microclimate across an urban landscape shapes variation in mosquito population dynamics and the potential of *Aedes albopictus* to transmit arboviral disease. *PLoS Negl Trop Dis* 11(5):e0005640
10. Wimberly MC, Davis JK, Evans MV, Hess A, Newberry PM, Solano-Asamoah N, Murdock CC (2020) Land cover affects microclimate and temperature suitability for arbovirus transmission in an urban landscape. *PLoS Negl Trop Dis* 14(9):e0008614
11. Snyder RL, Spano D, Duce P (2013) Weather station siting: effects on phenological models. In: Schwartz MD (ed) *Phenology: an integrative environmental science*. Springer, New York, pp 367–382
12. Fiebrich CA (2009) History of surface weather observations in the United States. *Earth-Sci Rev* 93(3–4):77–84
13. Venter ZS, Brousse O, Esau I, Meier F (2020) Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. *Remote Sens Environ* 242:111791
14. Colston JM, Ahmed T, Mahopo C, Kang G, Kosek M, de Sousa JF, Shrestha PS, Svensen E, Turab A, Zaitchik B (2018) Evaluating meteorological data from weather stations, and from satellites and global models for a multi-site epidemiological study. *Environ Res* 165:91–109
15. Dinku T (2019) Challenges with availability and quality of climate data in Africa. In: Melesse AM, Wossenu A, Senay G (eds) *Extreme hydrology and climate variability*. Elsevier, pp 71–80

16. Hofstra N, New M, McSweeney C (2010) The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Clim Dyn* 35(5):841–858
17. Bostan P, Heuvelink GB, Akyurek S (2012) Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *Int J Appl Earth Obs Geoinf* 19:115–126
18. Herrera S, Fernández J, Gutiérrez JM (2016) Update of the Spain02 gridded observational dataset for EURO-CORDEX evaluation: assessing the effect of the interpolation methodology. *Int J Climatol* 36(2):900–908
19. Fuka DR, Walter MT, MacAlister C, Degaetano AT, Steenhuis TS, Easton ZM (2014) Using the Climate Forecast System Reanalysis as weather input data for watershed models. *Hydrol Process* 28(22):5613–5623
20. Behnke R, Vavrus S, Allstadt A, Albright T, Thogmartin WE, Radeloff VC (2016) Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecol Appl* 26(5):1338–1351
21. Zandler H, Haag I, Samimi C (2019) Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions. *Sci Rep* 9(1):1–15
22. Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset. *Int J Climatol* 34(3):623–642
23. Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci Data* 2:150066
24. Verdin A, Funk C, Peterson P, Landsfeld M, Tuholske C, Grace K (2020) Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. *Sci Data* 7(1):1–14
25. Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37(12):4302–4315
26. Karger DN, Conrad O, Böhrer J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M (2017) Climatologies at high resolution for the earth’s land surface areas. *Sci Data* 4(1):1–20
27. Daly C, Neilson RP, Phillips DL (1994) A statistical topographic model for mapping climatological precipitation over mountainous terrain. *J Appl Meteorol* 33(2):140–158
28. Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, Luo L, Alonge C, Wei H, Meng J, Livneh B, Lettenmaier D, Koren V, Duan Q, Mo K, Fan Y, Mocko D (2012) Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J Geophys Res* 117(D3)
29. Abatzoglou JT (2013) Development of gridded surface meteorological data for ecological applications and modelling. *Int J Climatol* 33(1):121–131
30. Ford TE, Colwell RR, Rose JB, Morse SS, Rogers DJ, Yates TL (2009) Using satellite images of environmental changes to predict infectious disease outbreaks. *Emerg Infect Dis* 15(9):1341–1346
31. Parselia E, Kontoes C, Tsouni A, Hadjichristodoulou C, Kioutsioukis I, Magiorkinis G, Stilianakis NI (2019) Satellite earth observation data in epidemiological modeling of malaria, dengue and West Nile virus: a scoping review. *Remote Sens* 11(16):1862
32. Wimberly MC, de Beurs KM, Loboda TV, Pan WK (2021) Satellite observations and malaria: new opportunities for research and applications. *Trends Parasitol* 37(6):525–537
33. Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens Environ* 8:127–150
34. Pettorelli N, Vik JO, Mysterud A, Gaillard J-M, Tucker CJ, Stenseth NC (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol Evol* 20(9):503–510
35. Merkle JA, Cross PC, Scurlock BM, Cole EK, Courtemanch AB, Dewey SR, Kauffman MJ (2018) Linking spring phenology with mechanistic models of host movement to predict disease transmission risk. *J Appl Ecol* 55(2):810–819

36. Chuang TW, Wimberly MC (2012) Remote sensing of climatic anomalies and West Nile virus incidence in the northern Great Plains of the United States. *PLoS ONE* 7(10):e46882
37. Klisch A, Atzberger C (2016) Operational drought monitoring in Kenya using MODIS NDVI time series. *Remote Sens* 8(4):267
38. Jiang Z, Huete A, Didan K, Miura T (2008) Development of a two-band enhanced vegetation index without a blue band. *Remote Sens Environ* 112(10):3833–3845
39. Huete AR (1988) A soil-adjusted vegetation index (SAVI). *Remote Sens Environ* 25(3):295–309
40. Thomson M, Connor S, Milligan P, Flasse S (1997) Mapping malaria risk in Africa: What can satellite data contribute? *Parasitol Today* 13(8):313–318
41. Rogers D, Randolph S (1991) Mortality rates and population density of tsetse flies correlated with satellite imagery. *Nature* 351(6329):739–741
42. Midekisa A, Senay G, Henebry GM, Semuniguse P, Wimberly MC (2012) Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. *Malar J* 11:165
43. Brown HE, Diuk-Wasser MA, Guan Y, Caskey S, Fish D (2008) Comparison of three satellite sensors at three spatial scales to predict larval mosquito presence in Connecticut wetlands. *Remote Sens Environ* 112(5):2301–2308
44. Hilker T, Lyapustin AI, Tucker CJ, Sellers PJ, Hall FG, Wang Y (2012) Remote sensing of tropical ecosystems: atmospheric correction and cloud masking matter. *Remote Sens Environ* 127:370–384
45. Li Z-L, Tang B-H, Wu H, Ren H, Yan G, Wan Z, Trigo IF, Sobrino JA (2013) Satellite-derived land surface temperature: Current status and perspectives. *Remote Sens Environ* 131:14–37
46. Liu J, Hagan DFT, Liu Y (2020) Global land surface temperature change (2003–2017) and its relationship with climate drivers: airs, modis, and era5-land based analysis. *Remote Sens* 13(1):44
47. Cao J, Zhou W, Zheng Z, Ren T, Wang W (2021) Within-city spatial and temporal heterogeneity of air temperature and its relationship with land surface temperature. *Landsc Urban Plan* 206:103979
48. Vancutsem C, Ceccato P, Dinku T, Connor SJ (2010) Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sens Environ* 114(2):449–465
49. Wan Z (2014) New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sens Environ* 140:36–45
50. Hulley GC, Malakar NK, Islam T, Freepartner RJ (2017) NASA's MODIS and VIIRS land surface temperature and emissivity products: a long-term and consistent earth system data record. *IEEE J Sel Top Appl Earth Obs Remote Sens* 11(2):522–535
51. Malakar NK, Hulley GC, Hook SJ, Laraby K, Cook M, Schott JR (2018) An operational land surface temperature product for Landsat thermal data: methodology and validation. *IEEE Trans Geosci Remote Sens* 56(10):5717–5735
52. Hulley GC, Göttsche FM, Rivera G, Hook SJ, Freepartner RJ, Martin MA, Cawse-Nicholson K, Johnson WR (2021) Validation and quality assessment of the ECOSTRESS level-2 land surface temperature and emissivity product. *IEEE Trans Geosci Remote Sens* 60:1–23
53. Chakraborty T, Lee X, Ermida S, Zhan W (2021) On the land emissivity assumption and Landsat-derived surface urban heat islands: a global analysis. *Remote Sens Environ* 265:112682
54. Sun Q, Miao C, Duan Q, Ashouri H, Sorooshian S, Hsu KL (2018) A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev Geophys* 56(1):79–107
55. Kirschbaum DB, Huffman GJ, Adler RF, Braun S, Garrett K, Jones E, McNally A, Skofronick-Jackson G, Stocker E, Wu H (2017) NASA's remotely sensed precipitation: a reservoir for applications users. *Bull Am Meteorol Soc* 98(6):1169–1184
56. Nguyen P, Ombadi M, Sorooshian S, Hsu K, AghaKouchak A, Braithwaite D, Ashouri H, Thorstensen AR (2018) The PERSIANN family of global satellite precipitation data: A review and evaluation of products. *Hydrol Earth Syst Sci* 22(11):5801–5816

57. Xie P, Joyce R, Wu S, Yoo S-H, Yarosh Y, Sun F, Lin R (2017) Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *J Hydrometeorol* 18(6):1617–1641
58. Adler RF, Sapiano MR, Huffman GJ, Wang J-J, Gu G, Bolvin D, Chiu L, Schneider U, Becker A, Nelkin E (2018) The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere* 9 (4):138
59. Sulla-Menashe D, Gray JM, Abercrombie SP, Friedl MA (2019) Hierarchical mapping of annual global land cover 2001 to present: the MODIS collection 6 land cover product. *Remote Sens Environ* 222:183–194
60. DiMiceli C, Townshend J, Carroll M, Sohlberg R (2021) Evolution of the representation of global vegetation by vegetation continuous fields. *Remote Sens Environ* 254:112271
61. Buchhorn M, Lesiv M, Tsendbazar N-E, Herold M, Bertels L, Smets B (2020) Copernicus global land cover layers—collection 2. *Remote Sens* 12(6):1044
62. Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D, Stehman SV, Goetz SJ, Loveland TR, Kommareddy A, Egorov A, Chini L, Justice CO, Townshend JR (2013) High-resolution global maps of 21st-century forest cover change. *Science* 342(6160):850–853
63. Potapov P, Turubanova S, Hansen MC, Tyukavina A, Zalles V, Khan A, Song X-P, Pickens A, Shen Q, Cortez J (2022) Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat Food* 3(1):19–28
64. Gong P, Li X, Wang J, Bai Y, Chen B, Hu T, Liu X, Xu B, Yang J, Zhang W (2020) Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sens Environ* 236:111510
65. Tulbure MG, Hostert P, Kuemmerle T, Broich M (2021) Regional matters: on the usefulness of regional land-cover datasets in times of global change. *Remote Sens Ecol Conserv* 8(3):272–283
66. Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* 10(2):e0107042
67. Wardrop N, Jochem W, Bird T, Chamberlain H, Clarke D, Kerr D, Bengtsson L, Juran S, Seaman V, Tatem A (2018) Spatially disaggregated population estimates in the absence of national population and housing census data. *Proc Natl Acad Sci USA* 115(14):3529–3537
68. Tatem AJ (2017) WorldPop, open data for spatial demography. *Sci Data* 4(1):1–4
69. Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA (2000) LandScan: a global population database for estimating populations at risk. *Photogramm Eng Remote Sens* 66(7):849–857
70. Yang K, LeJeune J, Alsdorf D, Lu B, Shum CK, Liang S (2012) Global distribution of outbreaks of water-associated infectious diseases. *PLoS Negl Trop Dis* 6(2):e1483
71. Cann K, Thomas DR, Salmon R, Wyn-Jones A, Kay D (2013) Extreme water-related weather events and waterborne disease. *Epidemiol Infect* 141(4):671–686
72. Stanke C, Kerac M, Prudhomme C, Medlock J, Murray V (2013) Health effects of drought: a systematic review of the evidence. *PLoS Curr* June 5
73. Cohen JM, Ernst KC, Lindblade KA, Vulule JM, John CC, Wilson ML (2008) Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands. *Malar J* 7:40
74. Beltrame L, Dunne T, Vineer HR, Walker JG, Morgan ER, Vickerman P, McCann CM, Williams DJ, Wagener T (2018) A mechanistic hydro-epidemiological model of liver fluke risk. *J R Soc Interface* 15(145):20180072
75. Colston JM, Zaitchik B, Kang G, Yori PP, Ahmed T, Lima A, Turab A, Mduma E, Shrestha PS, Bessong P (2019) Use of earth observation-derived hydrometeorological variables to model and predict rotavirus infection (MAL-ED): a multisite cohort study. *Lancet Planet Health* 3(6):e248–e258
76. Shaman J, Day JF, Komar N (2010) Hydrologic conditions describe West Nile virus risk in Colorado. *Int J Environ Res Public Health* 7(2):494–508

77. Davis JK, Vincent GP, Hildreth MB, Kightlinger L, Carlson C, Wimberly MC (2018) Improving the prediction of arbovirus outbreaks: a comparison of climate-driven models for West Nile virus in an endemic region of the United States. *Acta Trop* 185:242–250
78. Pekel J-F, Cottam A, Gorelick N, Belward AS (2016) High-resolution mapping of global surface water and its long-term changes. *Nature* 540(7633):418–422
79. Pickens AH, Hansen MC, Hancher M, Stehman SV, Tyukavina A, Potapov P, Marroquin B, Sherani Z (2020) Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. *Remote Sens Environ* 243:111792
80. Worden J, de Beurs KM (2020) Surface water detection in the Caucasus. *Int J Appl Earth Obs Geoinf* 91:102159
81. Midekisa A, Senay GB, Wimberly MC (2014) Multisensor earth observations to characterize wetlands and malaria epidemiology in Ethiopia. *Water Resour Res* 50(11):8791–8806
82. Catry T, Li Z, Roux E, Herbreteau V, Gurgel H, Mangeas M, Seyler F, Dessay N (2018) Wetlands and malaria in the Amazon: guidelines for the use of synthetic aperture radar remote-sensing. *Int J Environ Res Public Health* 15(3):468
83. Odhiambo JN, Kalinda C, Macharia PM, Snow RW, Sartorius B (2020) Spatial and spatio-temporal methods for mapping malaria risk: a systematic review. *BMJ Glob Health* 5(10):e002919
84. Senay G, Velpuri NM, Bohms S, Budde M, Young C, Rowland J, Verdin J (2015) Drought monitoring and assessment: remote sensing and modeling approaches for the famine early warning systems network. In: Parron P, Baldassarre GD (eds) *Hydro-meteorological hazards, risks and disasters*. Elsevier, pp 233–262
85. Lelieveld J, Pozzer A, Pöschl U, Fnais M, Haines A, Münzel T (2020) Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective. *Cardiovasc Res* 116(11):1910–1917
86. Anenberg SC, Bindl M, Brauer M, Castillo JJ, Cavalieri S, Duncan BN, Fiore AM, Fuller R, Goldberg DL, Henze DK (2020) Using satellites to track indicators of global air pollution and climate change impacts: lessons learned from a NASA-supported science-stakeholder collaborative. *GeoHealth* 4(7):e2020GH000270
87. Chu Y, Liu Y, Li X, Liu Z, Lu H, Lu Y, Mao Z, Chen X, Li N, Ren M (2016) A review on predicting ground PM_{2.5} concentration using satellite aerosol optical depth. *Atmosphere* 7(10):129
88. Frumkin H, Bratman GN, Breslow SJ, Cochran B, Kahn PH Jr, Lawler JJ, Levin PS, Tandon PS, Varanasi U, Wolf KL (2017) Nature contact and human health: a research agenda. *Environ Health Persp* 125(7):075001–075001
89. Labib S, Lindley S, Huck JJ (2020) Spatial dimensions of the influence of urban green-blue spaces on human health: a systematic review. *Environ Res* 180:108869
90. Degefu MA, Bewket W, Amha Y (2022) Evaluating performance of 20 global and quasi-global precipitation products in representing drought events in Ethiopia I: visual and correlation analysis. *Weather Clim Extrem* 35:100416
91. Hess A, Davis JK, Wimberly MC (2018) Identifying environmental risk factors and mapping the distribution of West Nile virus in an endemic region of North America. *GeoHealth* 2(12):395–409
92. Estrada-Peña A, Estrada-Sánchez A, Estrada-Sánchez D, de la Fuente J (2013) Assessing the effects of variables and background selection on the capture of the tick climate niche. *Int J Health Geogr* 12(1):1–13

Dr. Michael C. Wimberly is an ecologist and geographer who studies environmental change and its effects on ecosystem and human health. He is a professor at the University of Oklahoma, where he conducts research on the influences of climate and land use on malaria and West Nile virus. He applies the results to develop risk maps and early warning systems for mosquito-borne diseases.