**Assessing the use of micro sensors to create a low-cost network to monitor local climate:**

Considerations for setting up a network and practical implementation in a coffee plantation in Costa Rica

*DRAFT*

*MSc Thesis by*

*Kees van Duijvendijk,*

*Reg.nr giv13kva*

*10 October 2014*

Table of Contents

1. Introduction

2. Context

*- Problem statement*

*- Climate science & precision agriculture*

*- Sensors & networks*

*- e-Agriculture & crowd-sourcing*

*- Sensors & agriculture*

*- Crop thresholds*

3. Non-spatial network aspects

*- Introduction*

*- Sensor & temporal*

*- Sensor shielding*

*- Data calibration*

*- Discussion*

4. Spatial aspects of the network

*- Sampling & spatial distribution*

*- Interpolation strategy*

*- Network accuracy*

*- Static covariates*

*- Dynamic covariates*

*- Co-kriging*

*- Discussion*

5. Practical implementation

- *Calculations (raster & fuzzy logic)*

*- Introduction to Aquiares*

*- Introduction to coffee*

*- Climate & coffee quality*

*- Climate & pests/diseases*

*- Discussion (required number of sensors)*

6. Conclusion & recommendations

1. Introduction

This MSc thesis will discuss the needs for low-cost networks of climate sensors and will deal with several issues that have to be taken into account when implementing such networks. The document will be divided into four main parts. The first section will discuss the wider context in which these networks have to be seen; this includes a short history of the existing network, the limitations, and the possible uses of a network that can be used to understand local differences in rural areas. The problem of the currently available data will be discussed, as well as projects in which these networks could be used. Several network- & sensor types will be dealt with to place the created network into a wider context. After providing the context, the second section will deal with the non-spatial aspects of the network; this will mainly relate to quality control. This section is used to calibrate the data and compare it to existing (WHO-certified) weather stations. Results of a set of experiments will be covered here, including experiments related to the sensor resolution, temporal resolution, and possibilities to use regression to make data comparable to official weather stations. This will ensure that the data that will be used in further spatial modelling will be of sufficient level, to avoid the 'garbage in, garbage out' problem for which we have all be warned during this MSc.

The third section will deal with the spatial aspects of the network; this will include the selected sampling strategy, the spatial-temporal interpolation technique, and several statistical analyses. Statistical tests will be used to find the accuracy of the network and aim to find the lowest number of sensors that will provide sufficient detail. Other tests will be conducted to find the relationship between both static (altitude, vegetation height and leaf area index) and dynamic (shading and insolation)) factors and the (min, mean & max) temperature and humidity that has been measured at different moments throughout the day (24 hours will be calculated based on a longer period). After these theoretical sections, the fourth section will focus on the practical implementation and its uses. Coffee will be selected as study crop, and the sensor-network will be used to make different analyses. The first test will relate to several crops in the EcoCrop model that could be selected as alternatives to coffee in case this crop would suffer too much from climate change and related pests and diseases. This will be a simple raster calculation project based on the absolute and optimal temperature range, rainfall, latitude and altitude in the region. The second test will be to link the data to a number of crop- and pest-thresholds that have an impact on coffee production in the region. This will be based on a fuzzy model with the different pests and diseases, which will be assigned different weights and associated with different temperature, humidity and altitudes. The fifth and final section will deal with the conclusions and recommendations of this research project. All data, graphics and (R-) scripts that have been used will be made available in an online repository, which is currently available at <https://github.com/cornelisvd/thesis>.

2. Background

**Problem statement:**

The need for better understanding of the variation in local climate is stressed from different sides; the information people need the most – local climatic changes – is also the most difficult to predict (Schiermeier 2010). Within the agricultural sector, environmental concerns emphasis the need to understand site-specific conditions in order to reduce use of external inputs (e.g. by precision agriculture) and find suitable varieties. Climate analogues are expected to move and cause shifts in crops at the global scale (Ramirez-Villegas et al. 2011), while options for adaptation will often depend on the local context (Nhemachena & Hassan 2007). At this level, there will also be different micro-climates, which are influenced by factors such as the canopy height (and density), the shelter or openness to wind, general topography (including slope aspect) and the vicinity of larger water bodies (Ashcroft & Gollan 2013). Official weather stations have to meet a lot of standards (WMO 2008) and are not available in many regions. Even if they are, data is not always digitized and thus easily accessible. Interpolation based on this data has been done for larger regions (Hijmans et al. 2005), but faces large number of uncertainties. As smallholder farmers often cultivate more complex (marginal/mountainous) terrains (Altieri 2002), information about climatic differences at this scale will be especially relevant to them. Information at this level is still lacking in most regions.

Downscaling of several crucial climatic factors that influence crop performance can help to better target certain crops and crop varieties to farmers. A typology of climate downscaling applications is discussed in Pielke and Wilby (2012); this project will partly aim to downscale already ‘local’ data to an even smaller scale. Existing examples will be discussed in the section on linked projects. Most of the projects that currently use site-specific and frequently updated information are focused on large-scale commercial farming systems in developed countries. While site-specific information is available on many issues in developing countries, these are generally not well-linked or regularly updated. The current models that match climate and crops are often based on maps of a scale that is not really relevant for small farmers in complex terrain; even the very high resolution climate dataset WorldClim still has a cell-size of 1km2 (Hijmans et al. 2005). Taking into account the many local factors that influence climate, and that the area that smallholders generally cultivate small areas in risk prone areas (Morton 2007), there is a need for more site-specific climate information for these farmers. Site-specific information can be provided by a range of low-cost technologies and mobile applications. Mobile applications in small-scale agriculture (hereafter referred to as *e*-Agriculture) can help in providing (a) information about markets and general farming methods (e.g. Gakuru et al. 2009), and (b) site-specific recommendations about fertilizer application based on taking images (Satyanarayana et al. 2011), but can also be used in more participatory approaches, such as crowd sourcing of crop improvement (Van Etten 2011). How to have sufficient farmer participation in projects that deal with something with less direct benefits (compared to most extension projects), will also be an issue that will be discussed in this thesis.

**Climate science & precision agriculture**

Climate science has a long history which will not be covered in detail here, as it has already been covered in much detail in numerous papers and books. The first time a network (Reseau Mondial) has been proposed for real-time data sharing (by telegram) was in 1900 (Edwards 2010), but only once the computer arrived, different formats could easily be merged to use in forecasting. The first satellites were launched in the late 1950s (Erickson 1964), after which both weather stations and satellite data could be used in climate science. When used in agriculture science, the main difference between satellites and weather station data is that weather station data give accurate but sporadic data, while for satellites it is hard to measure near-surface phenomena (e.g. precipitation), but the spatial coverage is complete. Tests have showed that satellites provide better temperature measurements than interpolated weather station data, while this is the other way around for precipitation data (Mendelsohn et al. 2007). As interpolation seems to be the main problem, increasing the number of weather stations should be able to outperform coarse satellite temperature observations. Satellite data will remain important for issues that cannot be measured at the ground level, which includes cloud data. At other levels, weather monitoring can now also be done by other – more responsive – approaches, such as by using wireless sensor networks (Popa & Iapa 2011).

Using the newest technology to improve the efficiency of resources in agriculture has been happening for decades, and options related to geographic information systems can be found in precision agriculture systems in many regions in the world with farms of sizes that make these profitable. One example of recent developments is Monsanto`s FieldScripts, which can recommend a certain seed for all farmers in the US, based on billions of soil observations and data from trillions of weather-simulation points (Beardsley 2014). Other uses of large networks of sensors include the use of a network of radio telescopes for measuring micro-climates in potato crops in Europe, and linking this to risk of phytophtora (Baggio 2005), as well as linking (Davis) weather stations to risk of pests and the dangers they pose to certain crops. The costs normally associated with these kinds of sensor-networks - *and availability of related data-sets on other environmental aspects that impact on crop-performance* - will create limitations for scaling-up these projects to regions with less commercial farming. The decreasing cost, together with the increasingly small size of sensors and advancements in technology (related to data storage & transmission), however, are making the use of micro-sensors interesting for a wide range of purposes (Ruiz-Garcia et al. 2009). While many of these sensors have initially been developed for use in other fields, such as logistics (e.g. Jedermann et al. 2006), micro-sensors have already been used in clinical studies (van Marken Lichtenbelt et al. 2006), a range of environmental studies including habitat monitoring (e.g. Mainwaring et al. 2002) and measuring ocean acidification (Rérolle et al. 2012). There are large differences between sensors when it comes to costs, accuracy, measurements, lifespan, memory, and data transmission.

While there are is thus clearly an adequate number of options to select from, the main issue in making these networks relevant is not the sensor itself, but the shield in which the sensor is placed. The actual sensor itself is small and generally cheap, while the construction in which it is placed adds important features such as storage, connectivity and sometimes robustness to shocks and water damage. This sensor housing – *in case of the cheaper sensors* – does not add active ventilation or other protection from radiation. This, together with the standards that are set by the WHO (box 1), creates the need for a sensor shield in which the sensor will be placed. Some of the recommendations are easy to address in a low-cost sensor-shield (including height, inspection, comparison, coordinates, and metadata), while others are clearly more problematic when working with a limited budget. These include the addition of adequate radiation shielding, insulating material, and protection from water. While some 'robustness' can be added to the shield at low-cost, additional correction will likely have to be done by computer models. An important note made in the WMO Guide to Meteorological Instruments and Methods of Observation is that, while it is acknowledged that it might not be economically feasible to work with sensors that directly meet the accuracy requirements, *‘it is necessary to limit the size of the corrections to keep residual errors within bounds*’. Issues related to the sensor resolution, temporal interval, and shielding, will be discussed in the section on non-spatial network aspects; a number of experiments will be analysed regarding different shield constructions.

|  |
| --- |
| **Box 1: Summary of relevant recommendations for weather stations(WMO 2008)**   1. For general meteorological work, the observed air temperature should be representative of the free air conditions surrounding the station at a height of between 1.2 and 2.0 m above ground level. 2. A radiation shield or screen should be designed to provide an enclosure with an internal temperature that is both uniform and the same as that of the outside air; it should completely surround the thermometers and exclude radiant heat, precipitation and other phenomena that might influence the measurement. 3. Thermally insulating plastic-based material is preferred as material for the shield over the better performing highly polished, non-oxidized metal, because of its simple maintenance requirements; thermally insulating material must be used if the system relies on natural ventilation. 4. A humidity sensor may be combined with or co-located with a temperature sensor in its radiation shield as long as the sensor thermal output (self-heating) is very low. 5. Direct contact with liquid water will seriously harm sensors using hygroscopic electrolyte as a sensor element. Great care should be taken to prevent liquid water from reaching the sensitive element sensors. 6. Desirably characteristics include reliability & stability, simplicity of design, durability and acceptable cost. 7. Agricultural meteorological stations should be inspected at interval sufficiently short to ensure the maintenance of a high standard of observations and the correct functioning of the sensor. 8. Recording instruments should be compared frequently with instruments of the direct-reading type. 9. The position of a station referred to in the World Geodetic System 1984 (WGS-84) Earth Geodetic Model 1996 (EGM96) must be accurately known and recorded (1/1000 degrees latitude & longitude). 10. It is important that records should be kept not only of the temperature data, but also of the circumstances in which the measurements are taken (metadata). |

**Sensors & networks**

Micro-sensors can ‘sense’ many different phenomena and disseminate information in different ways; a detailed taxonomy of wireless micro-sensor network models has been made in Tilak et al. (2002). Micro-sensors can be used to monitor a wide range of physical and chemical factors, such as chemical compositions and temperature. Whereas the actual manufacturing of micro-sensors is not very expensive, the development process is likely more restrictive to development for specialized uses in small-scale agriculture. The main commercial demands will be either a large storage capacity in the logistics sector (to check the conditions during the transportation after it has arrived at a certain point), or a real-time reading ability in buildings or greenhouses. A combination of real-time reading and a large storage is not necessary for many purposes, while this would be essential in small-scale agriculture. (Semi-) real-time information can help in analyzing periods of crop stress or risk of diseases, while sufficient storage will be needed to only require reading of the sensors once a month or less to limit labor requirements. As developments in this field are proceeding rapidly, there will likely be affordable sensors available with all the right characteristics in a few years. At this moment the study will focus on issues that will remain relevant for other sensors; this includes the shielding, calibration and interpolation. Actual use will change a bit with the availability of real-time data, as this can be used for early warning, while the current system can only assess which areas have been (and thus will likely also be in the next season) at highest risk to surpass certain temperature and humidity thresholds.

The sensors that will be used in this research (iButtons) have mainly been developed for logistics, but have been used in a wide range of scientific field – and have been common in similar studies in ecology. The price for one iButton (if sold per thousand) ranges from under 2 USD for a basic temperature logger to over 50 USD for the sensors that will be used in this research (Maxim Integrated, web). Important issues that have to be taken into account outside the issues that can be monitored will include (a) the interval and total number of measurements that can be made and stored; (b) ways in which the data can be transmitted (wireless transmission or manual USB-upload); and (c) the energy requirements (battery life) of the sensor. For long-term monitoring, it will be crucial to find energy-efficient paths, although this will depend on the type of transmission of the sensors (Shah & Rabaey 2002). This is an issue, however, that is not relevant for the created network, as the data will be uploaded manually.

**Size of the iButton Hygrochon DS1923 (source: Maxim Integrated)**



**e-Agriculture & crowd-sourcing**

The term e-Agriculture was first promoted at the World Summit on the Information Society (Singh 2012) in which it was identified as a key-action line to address the Millennium Development Goals. E-Agriculture is defined as *“an emerging field in the intersection of agricultural informatics, agricultural development and entrepreneurship, referring to agricultural services, technology dissemination, and information delivered or enhanced through the Internet and related technologies”*, but is going beyond technology and has as aim to improve the communication and learning processes in the agricultural sector (<http://www.e-agriculture.org/>). As ICT is easily accessible, national research institutes as well as international agricultural research centers can create their own applications to provide extension to their target audience. Although e-agriculture approaches provide interesting opportunities to generate cost-effective solutions to information delivery, most solutions so far have focused on getting information to farmers.

Crowdsourcing is an approach in which tasks are outsourced to (paid or unpaid) volunteers and has a long history in ecological research (Miller-Rushing et al. 2012). Bioversity International, an international agricultural research institute member of the CGIAR Consortium has developed an approach in which large numbers of farmers test crop varieties. Each farmer tests a set of small seed samples of three different varieties. They compare the varieties for a number of characteristics and report back on their findings and can also order more seed. Innovative algorithms used in market research applications are used to extract patterns from the data which are returned to farmers in an easily interpretable form (Van Etten 2011). The contribution of farmers (land, labor, resources) and the use of mobile phone technology at different stages in the process make the process considerably cheaper than conventional approaches. Mobile phones (with limited functionality) and charging station will have to be provided to farmers in order to receive feedback. Adding climatic information to these programs can increase understanding of the crop performance and be used to recommend certain varieties that have performed well in other regions.

**Micro-sensors & Agriculture**

While a range of studies have focused on the impact of understory microclimates on crops – often related to forms of agroforestry (Brown 1993; Perfecto & Vandermeer 1996; Lin 2007), these have often only taken into account the basic vegetation differences (height, leaf area index) and not the actual temperature/humidity parameters measured regularly for a longer period. On the other hand, the impact of temperature on crops is also well studied - as can be seen in the EcoCrop database (FAO 2013), and in lesser extent also the impact of relative humidity (Easterling et al. 1992). Linking these two kinds of studies has been quite limited, as climatic parameters have not been available at the level at which the crops grow. Linking crops to suitable data can be done by matching the EcoCrop dataset with current climate maps (e.g. Hijmans et al. 2005) and future climate (e.g. Govindasamy et al. 2003). This is possibly in many of the larger desktop GIS applications, such as DIVA-GIS, but also by online models, such as CGIAR’s Climate Analogues (CGIAR/CCAFS 2011), which can take into account different models and (bio-) climatic variables. The Climate Analogues can also take into account non-climatic factors, such as socio-economic characteristics, crops, and soils (Ramirez-Villegas et al. 2011). The main limitation of both tools is that the minimum resolution in analyses will be ≥ 1 km2, although there are also other limitations to applying this approach over larger areas with different soil types, social preferences and other factor that influence crop choices.

Studies regarding the temperature thresholds and impact of relative humidity on crops are mainly conducted in controlled (greenhouse) environments. This will help to analyze the optimal and absolute temperature range, but more complex relationships, such as canopy height and density, cannot easily be studied at this level. Local climate is only one of the factors within the biophysical environment that impacts crops; and this is merely one of the reasons for the yield gaps – together with crop management, socio-economic factors, institutional environment and technology transfer (FAO 2000). Another problem with linking climate and crops in existing models is that variation at the local level is so high that models have to be validated (at the site/surrounding area) by data from a number of years (Lobell et al. 2009), resulting in smoothed datasets. While this is useful in models related to current and future crop potential (possibly linked to transfer of genetic material) by focusing on climate analogues, this will not be useful for understanding the local climatic extremes. Two approaches to analyze the link between, amongst other factors, climates and crop yield can be based on polygon (Thiessen Polygons) or grid-based methods (e.g. Inverse Distance Weighting & Ordinary Kriging) (Van Ittersum et al. 2013); polygon-based approaches uses a validated point and scales this up to a higher geographical level without data corrections, while a raster-based approach uses inter- and/or extrapolation when scaling up to a higher level. This research will use a raster-based approach to data analysis. Low-cost and robust networks of micro-sensors that measure temperature and humidity can complement satellite images and certified weather stations to give a better indication of the actual local extremes in a region.

*Drawing of differences between polygon and raster-based approach*

A network of micro-sensors that can measure temperature and humidity can have a range of possible uses. The network that will be implemented in this research will only provide information after the data has been downloaded to a computer. This will mean real-time data is not available and information can only be analyzed after the sensors have been retrieved from the field. A lot of information will become available that can be used for scientific studies that will find relations between fixed (or predictable) factors and the temperature and humidity. Basic relationships that can be studied include the relationship between topography and the climate (Whiteman 2000; Daly et al. 2002), as well as certain vegetation-related factors and climate. Another use of non-real-time information is to find local differences in the area. When taking the average over a longer period, it will become clear which areas are wetter/warmer and this can be linked to risk of certain diseases. Whether the predicted risks are linked to actual risk will require additional information of the state of the crops and pests during the growing season, but this detail of information is not available for this study region. The data will give interesting information about the climate at different altitudes, as this is often the main indicator of coffee quality. Except for studying issues that have already occurred, this data can be used to recommend other crops. If sensors have been measuring for a longer period and show a climate that is clearly not sufficiently suitable for the current crops, the dataset could be linked to information about other crops (e.g. EcoCrop) to find alternatives that could be socially acceptable. Research related to (plant and animal) diversity can also benefit from these networks, by enabling to (a) study the relationship between temperature variation and plant diversity (Letten et al. 2013), and (b) locate microrefugia (Ashcroft & Gollan 2013).

The biggest increase in uses will be made possible when adding the possibility to read the data real-time. This will make it possibly to make an early warning system out of the network. An example of a similar project, based on weather stations, is the Integrated Pest Management software that is created by Davis (<http://www.davisnet.com/weather/products/IPM/>). Another model that is related is CLIMEX, which provides information about the impact of climate change on the distribution of species ([http://www.hearne.com.au/ Software/CLIMEX/](http://www.hearne.com.au/%20Software/CLIMEX/)). Both are interesting, but very expensive (> US$ 500). The objective of this software is to take the current weather conditions and use this to calculate risk of pest (insects & diseases) and the related risk to the crop. This can be used to select the pest control method that is most suitable for the risk. In developing countries, this early warning is also done for many issues (plant protection, animal health and food safety are covered by FAO EMPRES). This early warning is mainly focused on the larger transboundary threats, and thus also larger strategies (satellite images and aerial surveys). Information based on climatic factors at the local level still have the possibility to provide large gains for pest monitoring of the smaller risks, which can be equally devastating at crops – although likely posing a smaller risk to countries' food security. In case there is an official weather station in the area that provides real-time information, this could be linked to the network. After data is available on how the climate differs over the area, a certain temperature for the official sensor can be used to predict how the climate will be over the larger area. This would provide possibilities for pest monitoring after the network has been removed from the area.

**Crop thresholds**

Many different thresholds exist for crops, linked to temperature, humidity, altitude and other factors. Of the thresholds (and optimal conditions) that are provided in the Ecocrop database, the altitude and latitude are easy to find. Light intensity and soil PH will require more expensive instruments or extensive surveys, although potential insolation (solar radiation per unit of surface) can be calculated from a DEM in combination with the sun position. Rainfall limits are requested at an annual basis, which would allow for use of satellite or interpolation weather station data. Temperature limits are provided for optimal conditions, as well as absolute limits. It is unclear at which temporal interval these absolute limits become problematic, but in this research each hour can be assessed. Linking temperature thresholds to crop production is relatively easy, as temperature is straightforward to measure (although not necessary accurate) and already has a long history in agronomic research. An extensive review of temperature thresholds for a number of crops is available in Luo (2011). As temperature is more problematic during certain crop stages (phenophases), and crops are often planted and harvested at different times in a region, accurate local temperature data can help individual farmers in determining their timing of planting and harvesting (and possible additional measures throughout the season). Due to the overall availability of some level of temperature data, most crop-related thresholds have at least some mentioning of temperature, which this might be less relevant for pests.

Humidity data is more difficult to determine than temperature data, but often has some level of (negative) correlation with temperature. Higher temperatures during the day often have a lower humidity, while humidity often is highest during the night (ref). Humidity data is often related to occurrence of pests – both during the growing season (e.g. Wilks and Shen, 1991) as during post-harvest storage (e.g. Calderon, 1981). Soil moisture is the most important water-related thresholds for crops during the season and can be linked to recommend timing of irrigation (Jensen et al. 1970). In general, crop water requirements depend on a range of weather factors (temperature, humidity, and wind speed), as well as crop factors (type, variety, phenophase) (Allen et al. 1998), which means the network can also be linked to irrigation scheduling – although soil moisture sensors would provide a more practical solution. In this research, humidity will mainly be linked to expected occurrence of pests and diseases, which pose the largest problem for coffee production in the study region. Finding the humidity thresholds of all crop-specific pests and diseases is a difficult process, especially because geographical distribution also plays a role. Existing software that provides real-time pest-risk information (e.g. Davis’ IPM software) is quite expensive at the moment.

A third important crop threshold that relates to both temperature and humidity is the Vapor Pressure Deficit (VPD), which is the difference between the actual and potential (saturated) air moisture. This can be estimated based on the daily minimum and maximum temperature (Wang et al. 2005), based on formulas explained in Murray (1996). A high VPD can result in closing of the stomata and reduced photosynthetic rates (Turner et al. 1984, Black and Unsworth, 1980) and can be used as an indicator in both agronomic studies as well as in studies of forest ecosystems.

3. Non-spatial network aspects

**Introduction**

With an objective to create a low-cost and robust network of climate-sensors, it will be important to find a good balance of costs and accuracy. The data has to be accurate enough to provide data about the climate in the area, and especially the extremes that are occurring in different regions and seasons. The spatial accuracy has to be at the same scale as at which the farmers make their decisions; this can range from under 1 to over 10 hectare for smallholders in different regions. For commercial farmers this will be even larger, but at this scale and with availability of more resources, satellite data or interpolated data from existing weather stations – in combination with some sort of software – will likely be sufficient. The costs of the proposed network can be reduced by having a low-cost housing of the sensors, but also by decreasing the labour required for data collection and interpretation. This means being able to use the sensors in the field for a long period without need to replace or read the sensor-data (preferably by wireless transmission). Analysing data is another step in which labour can be reduced by creating e-Agriculture applications that require limited input to provide useful information to farmers and researchers.

A straightforward option to increase the period the sensors can be used in the field is to adjust the resolution at which the sensors make their observations. The iButton Hygrochron (DS1923) can make temperature measurements at 8- or 11-bit resolution (0.5°C or 0.0625°C), and humidity measurements at 8- or 12-bit resolution, which is either 0.6% or 0.04% RH. The readings, however, are either stored in 8- or 16-bit, with the total memory for this sensor being 8192 bytes. According to the specifications, the temperature accuracy is better than ±0.5°C for most of the range (-10°C to +65°C), while the accuracy of the humidity measurements is ±5% RH. By using a lower resolution sensor for both temperature and humidity, the sensor can store double the number of observation at the same time interval (*see table 1*). The three options that are the most likely are highlighted; the options can cover a period of almost one year, but either require separate sensors to measure humidity and temperature, or measure at two-hour intervals. The other option measures both temperature and humidity (low resolution) at 1-hour intervals, but only has sufficient memory to store data over a period of 171 days, which would be sufficient for many cereal crops.

**Table 1: Number of observations and possible observation period at different resolutions**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of observations | Time covered at 2-hour interval | Time covered at 1-hour interval | Time covered at 30-min interval | Time covered at 15-min interval |
| *1 unit at low-res* | 8192 | 683 days | 341 days | 171 days | 85 days |
| *1 unit at high-res or* |  |  |  |  |  |
| *2 units at low-res* | 4096 | 341 days | 171 days | 85 days | 43 days |
| *1 unit at low-res & 1 unit at high-res* | 2560 | 213 days | 107 days | 53 days | 27 days |
| *2 units at high-res* | 2048 | 171 days | 85 days | 43 days | 21 days |

While changing the sensor resolution has the potential to provide easy gains, a more challenging approach to reduce costs relates to creating low-cost shielding. A large number of recommendations exists (see box 1), which relate to issues such as sensor-height, reflectivity and aeration of the sensor-housing. Literature provides data of related experiments that have been conducted to find low-cost sensor shielding (Tarara and Hoheisel 2007, Thomas and Smoot 2013, Holden 2013), in which differences have been found of up to 7.40C. Most of these studies have focused on natural (forest) areas – where additional shading can easily be found and shields can hung from trees. As the network that will be created will ideally be easy to use in different regions, the material that will be used for the shields will have to be very basic. This means only standard PVC tubes, basic materials for shielding (e.g. insulating foil) and basic approaches to increase aeration (drilling) will be tested. The shielding in which the reference sensors have been placed is a Stevenson screen placed at approximately 1.50m. The screen has no active aeration, but provided adequate protection against precipitation and radiation, and has been used at the station for over 50 years.

**Figure X: Stevenson screen and experimental shields at the CATIE weather station**



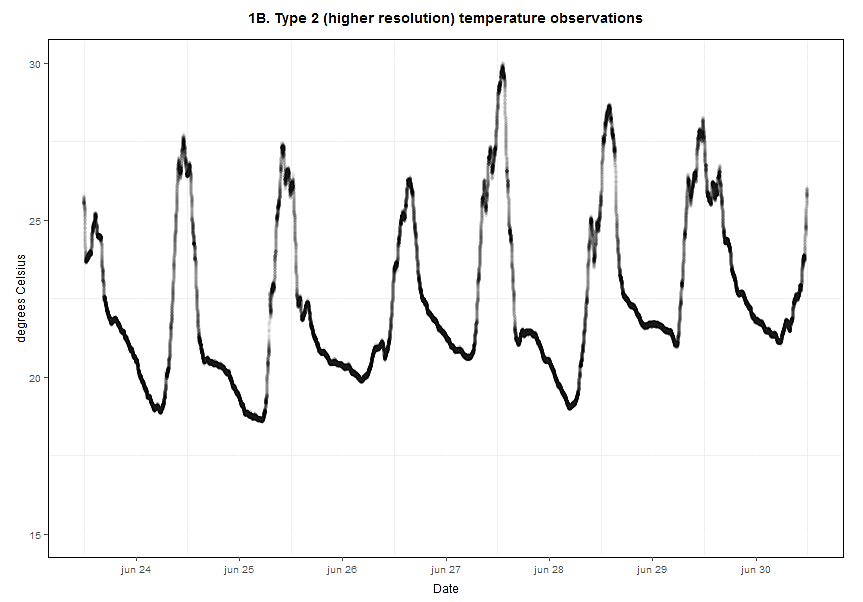
It is unlikely that these shielding constructions can provide the same accuracy as certified weather stations, but preferably the maximum and minimum will be in a small range that can be further reduced by regression of the dataset. The steps that will be taken to make the basic sensor shields and resolution accurate will be: 1) analysing the impact that different sensor resolution has on the provided data – and especially the extremes; 2) studying the impact that different temporal resolution have on the extremes; 3) testing different sensor shields and analysing how these impact on the provided temperature data ; and 4) combining the findings from step 1, 2 & 3 with regression methods and find a balance between the costs of the network and the accuracy of the temperature and humidity data. When assessing trade-offs, the most important factor will be the data robustness and accuracy.

Having similar data in the created shields as in a certified weather station is the best (physical calibration), but will be nearly impossible. The difference between the shields and a certified weather station will have to be the same, so that a computer-based correction can take place based on the expected differences at certain temperatures. It can also be expected that a lower temporal resolution (e.g. 2 hours) will result in smoothing of the data. This can be corrected by different regression techniques, but it will be hard to do this for the extremes, as wind and clouds will play a role at the local level which will be hard to model and hence correct. On the other hand, it can be expected that it will be very hard to reduce the impact of midday radiation on temperature. As the shields are projected to result in an exaggeration of temperature extremes, reducing the temporal interval (possibly in combination with some correction), could be used to offset these differences and will be tested.

**Sensor resolution**

An experiment has been done with a total of 30 sensors, making observations at 5-minute interval during one week. The sensors have been placed in a certified weather station, which also houses a long-term observation instrument, to make sure the conditions would be the same for all sensors. Half the sensors have been set at the highest temperature and humidity resolution, while the other half has been set at the lower resolution. The sensors have been paired throughout the weather station and linear interpolation has been used to get the value at 1-minute intervals. Statistics have been calculated over the dataset at minute-intervals, which contained 30 sensors with 10.080 observations each. Basic temperature graphs are provided in figure x, while more detailed plots from this experiment (including humidity) are provided in Annex X. The mean temperature and humidity for the low-res sensors are 22.37°C and 90.45% RH; for the high-res sensors these values are 22.35°C and 90.93% RH. Temperature data is all within a small range (< 0.50C difference) for both resolutions. Another important issue to take into account is the variability between sensors; a visual analysis shows that this is especially an issue for the low-res temperature observations. The standard deviation has been calculated as the mean of every row of the data-matrix. Correlation between temperature and humidity is -0.91 for the low- and -0.92 for the high-resolution sensors.

**Figure 1A: temperature graph at low-resolution Figure 1B: temperature graph at high-resolution**

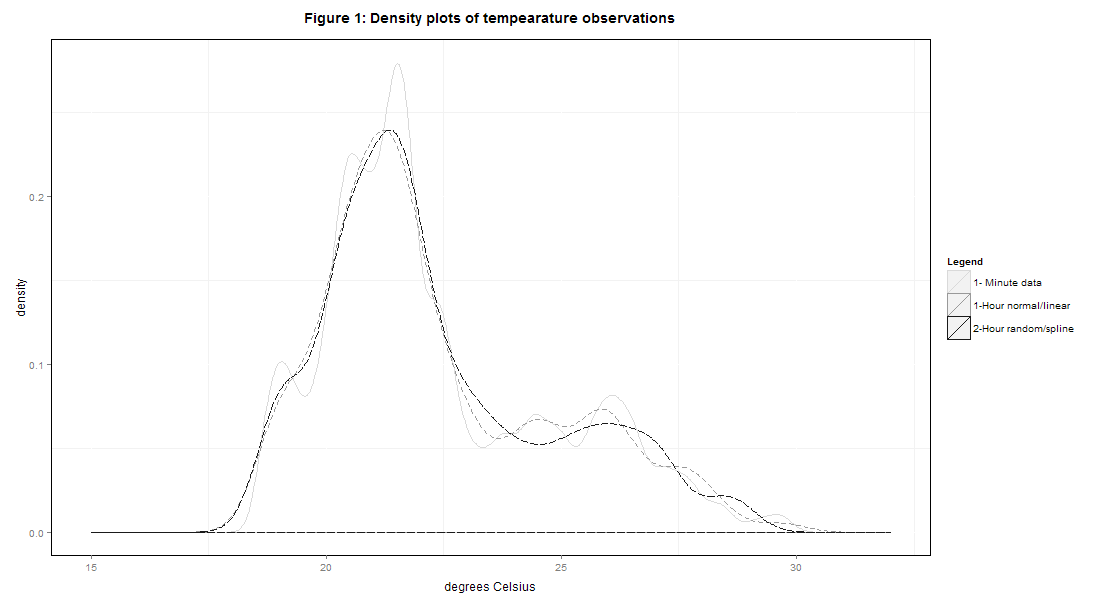


**Temporal resolution**

This experiment will initially use the high-resolution dataset that has also been used in the previous experiment, and after this the most promising regression techniques will also be applied to the low-resolution dataset to see whether it will provide sufficiently similar results to be able to work with only the low-resolution sensors. The original data-set (5-min interval) will be used, in which the start-moment will be chosen at random. This means no average will be taken, but only one value from the original data-set will be used and the rest will be ignored. Different interpolation techniques are available in the R-package {zoo} (Zeileis & Grothendieck 2005), which is developed especially for irregular time-series. The techniques to estimate values at a certain interval which are available in this package include linear (*na.approx*) and cubic spline (*na.spline*) methods, which have limited room for adjustment.

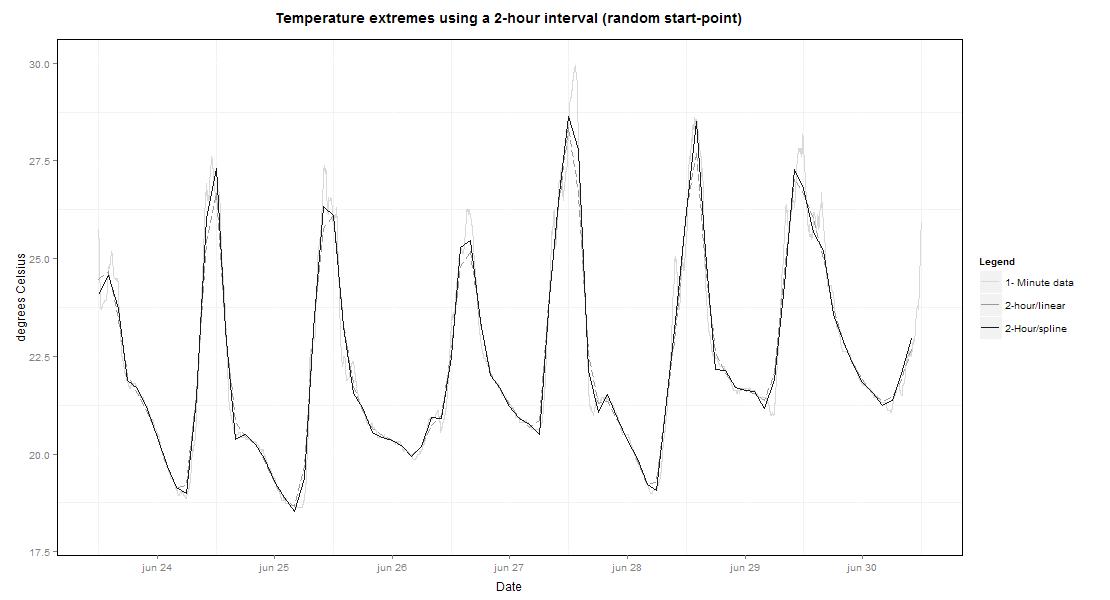
When the data is reduced to only have observations at the hour, instead of every minute, the data continues to have the same complexity and very little data will be lost (Annex X). The range for the temperature observations with the high-resolution sensor will be 18.59 – 29.85°C, while the mean will be 22.36°C. While measurements, when interpolated to 1-minute intervals, surpass 30°C (which might be a common threshold), this was only the case at two observations out of the total of 151,200, which makes this negligible. The overall density plot does not change a lot (figure X), which would result in similar conclusions that will be drawn from these datasets. Although most conclusions that will be drawn from this dataset will be similar to those drawn from the original data-set, a reduction in range could be a problem, as the absolute extremes will change. When only using one value for each hour, randomly selected from the original dataset (Annex X), and using a linear interpolation approach to get the value at the exact hour, the resulting plot shows more data smoothing. The range for this data is 18.59 – 29.66°C, while the mean is almost the same (22.35°C). The standard deviation in this experiment is still only ±0.09°C.

**Figure 2: Density plots of temperature observations selected by two different approaches**



When using the low resolution, the range and mean become 18.11 - 30.11°C and 22.38°C when using the closest observation, while it will be 18.11 – 29.85°C and 22.37°C for the random start-moment (Annex X). The maximum difference in temperature between any two sensors at a certain moment (the exact hour) is lowest when using the closest observation with the high-res sensor (0.65°C), followed by the low-res sensor with the closest value (1.02°C), the high-res sensor with a random start-moment (1.14°C) and finally the low-res sensor with random moment of start (1.69°C). The difference was 0.65/1.02°C (low/high-res) in the original dataset, which shows that the difference does not change when using the closest value. Taking a random starting point increases the maximum difference by ±0.5 degrees Celsius for both types of sensors and would thus best be avoided.

**Figure 3: Temperature extremes using a 2-hour interval (random starting-point)**



The graphs resulting from the built-in cubic spline function *na.spline* in the package {zoo} can be seen in Annex X. The range, when taking the closest measurement, is 18.59-29.89 for the high-res sensors and 18.11-30.16 for the low-res sensors. The mean is 22.36 for the high-res and 22.38 for the low-res sensors. The maximum difference at any hour is 0.34 for the high-res and 1.10 for the low-res sensors. With a random starting point, the range for the high-res sensor is 18.57-30.27°C and the mean is 22.35°C. The maximum difference at the hour has gone up from 1.14 to 1.43°C. For the low-res sensor, the range becomes 18.06-30.52°C, with a mean of 22.37°C. The difference between rows stays quite similar with 1.75°C. A summary of the differences caused by these different techniques can be seen in Annex 3. A 2-hour interval is also included in these tables to show how different techniques impact at this larger interval. The mean of al 15 sensors at high-resolution over the week of measurement for the 1-min compared to 2-hour observations (random start-point, linear and spline interpolation) is shown in figure 3. Splines provide a closer match to the original data-set, which is especially clear at higher temperatures. While splines give a good match on most days, it results in a large difference on June 27th, when temperature is highest.

*Table*

**Sensor shielding**

This experiment has been conducted on several days on the meteorological station at CATIE (Turrialba, Costa Rica). The first tests have been done with a larger (50mm diameter) PVC tube, after which a smaller version (25mm) has also been tested. All tubes have been thin white PVC tubes existing of several parts: 1) a 1.75m main tube (25cm to place it in the ground); 2) a PVC elbow attached to the main tube; 3) a small (20cm) second tube that will be attached horizontally to the elbow; 4) a second elbow that will be attached to the horizontal section and that will point down again; 5) a final 5cm tube that will be attached to the elbow and in which the sensor will be placed. The sensor will be placed on plastic mesh that will be located between the elbow and the final part. The material of this net might have an influence on the sensor, but testing different materials for this purpose has not been done at this momet.

As only a relatively short time was available for these experiments, the different shields that have been tested are based on results that have been discussed in existing literature on this subject. Main WMO recommendations have also been taken into account where possible, but have mainly related to the height of the sensor. Another important document has only been discovered after most of the experiments had been conducted (ISO standards on test methods for comparing the performance of thermometer shields). Some of the ISO standards that have been used include: using a test site with a range of meteorological conditions that will also be encountered at the actual field sited; the test site meets the standards of a meteorological station at a height between 1.25 and 2 m; and the shield will be tested under different conditions (in this case only predominantly sunny and cloudy days). The shields were mainly designed to limit the radiation-errors, by increasing the aeration and/or reflectivity. All experiments include at least a ‘control’ blank tube, a tube with holes (aeration), a tube with foil (reflectivity) and a tube with foil & holes. The experiments with 50mm tube also include reflective tape, but these were too difficult (thick) and expensive to also try on the smaller tubes. The graph of the first experiment, with 6 blank 50mm tubes (2 sensors each) and 6 sensors in the official weather station, is shown in figure X.

The first experiment already showed that the peaks in temperature during the peak sun-hours (around noon) are much higher in the station than in the station. There is a strong correlation between the temperature and humidity (-0.98), so the peaks in humidity are similar to the peaks in temperature. For this reason, only temperature statistics have been studied in the different shields. The graphs of the four experiments (50mm vs. 25mm, ‘sunny’ vs. ‘cloudy’ day) are provided in Annex 4. All graphs – except for the 4th graph, have observed 1-minute intervals from midnight to midnight (1,440 observations); the final graph has the same number of observations, but has been measuring the temperature and humidity from 3pm to 3pm due to the weather conditions. The first experiment (50mm with no further adjustment) provides a mean of 24.32°C, compared to a mean of 23.81°C in the station. The minimum is also quite similar, with 20.94°C in the tubes and 20.50°C in the station; the main problem, however, is the maximum: 28.33°C in the station and 31.71°C in the tubes. The standard deviation in the tubes was 0.25°C, compared to 0.06°C in the station. The results from the experiments with different constructions are provided in tables 2A & B.

**Table 2A: temperature statistics in different sensors shields (50 mm PVC) vs. station**

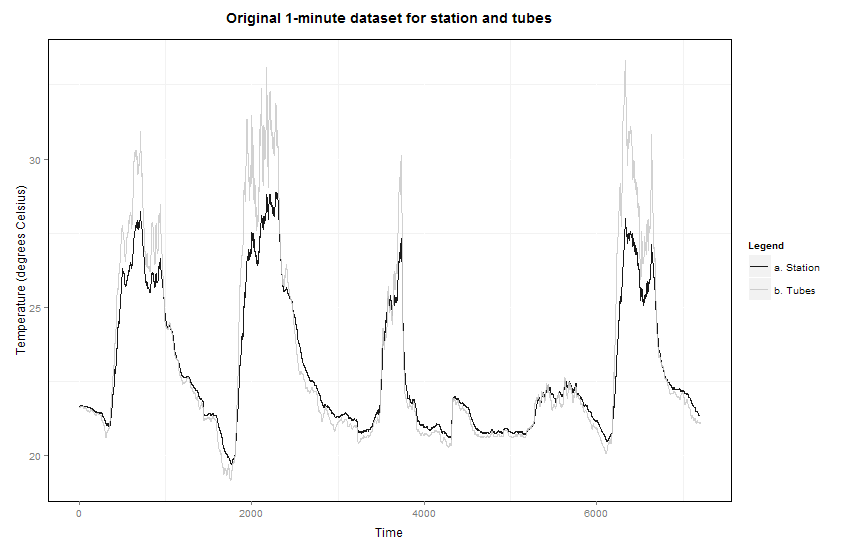
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sunny day | | | | Cloudy day | | | |
|  | Mean | Min | Max | St. dev | Mean | Min | Max | St. dev |
| Station | 23.92 | 19.68 | 28.94 | 0.04 | 21.92 | 20.48 | 27.40 | 0.05 |
| White paint | 24.61 | 19.08 | 33.78 | 0.25 | 21.81 | 20.23 | 30.87 | 0.10 |
| Holes (+- 30%) | 24.38 | 19.16 | 32.47 | 0.21 | 21.68 | 20.20 | 31.82 | 0.24 |
| Insulating foil | 24.61 | 19.48 | 32.17 | 0.22 | 21.92 | 20.27 | 30.20 | 0.08 |
| Holes + foil | 24.76 | 19.36 | 33.24 | 0.24 | 21.80 | 20.23 | 30.31 | 0.12 |
| Reflective tape | 24.70 | 19.20 | 34.42 | 0.42 | 21.77 | 20.27 | 30.17 | 0.20 |

**Table 2B: temperature statistics in different sensors shields (25 mm PVC) vs. station**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sunny day | | | | Cloudy day | | | |
|  | Mean | Min | Max | St. dev | Mean | Min | Max | St. dev |
| Station | 23.23 | 20.42 | 28.12 | 0.04 | 21.40 | 20.68 | 22.53 | 0.04 |
| Nothing | 23.94 | 20.05 | 34.07 | 0.17 | 21.29 | 20.55 | 22.70 | 0.05 |
| Holes (+- 30%) | 23.80 | 19.95 | 33.14 | 0.13 | 21.24 | 20.52 | 22.59 | 0.02 |
| Insulating foil | 23.72 | 20.24 | 32.27 | 0.10 | 21.31 | 20.55 | 22.89 | 0.09 |
| Holes + foil | 23.82 | 20.12 | 33.30 | 0.12 | 21.30 | 20.56 | 22.62 | 0.04 |

The experiments show that sensor housing will not make a large different on a fully cloudy day (2B - right), but even on a cloudy day with very limited sun (2A - right), the peaks caused by the sensors shield can be very significant (> 3°C). Very limited adjustments (nothing, white paint, or tape) result in the largest differences in the maximum value: almost 6°C on a sunny day in a 25mm PVC tube without any adjustment. The lowest differences on sunny days are found for the insulating foil (without holes). Although holes alone perform quite acceptable in both 25 & 50mm tubes on the sunny days (2nd after the insulating foil), this approach is quite labour intensive and will create the risk that water will reach the sensor. Combining holes and foil, however, is providing higher maximum values for the sunny days in both sizes of sensor shields. More detailed discussion will follow later in the *discussion* section. A next step will be to combine experiments 1 and 2, to see whether a reduced temporal interval (possibly in combination with additional smoothing) has the potential to provide better results on the sunny days.

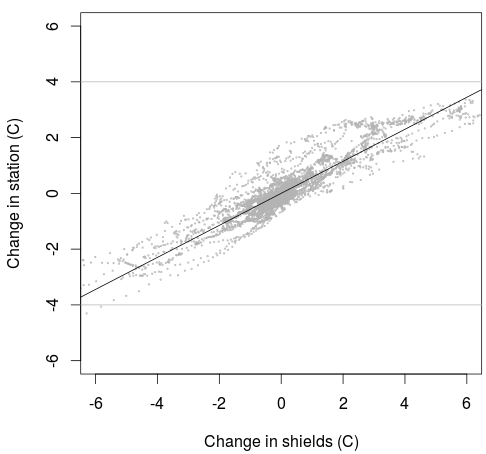
**Figure X: Original 1-minute dataset for the station and the (25 & 50mm) shields**



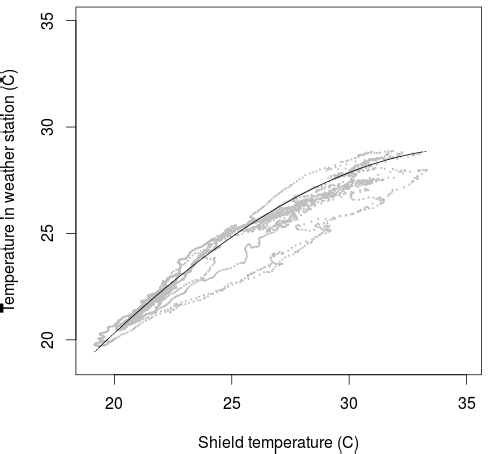
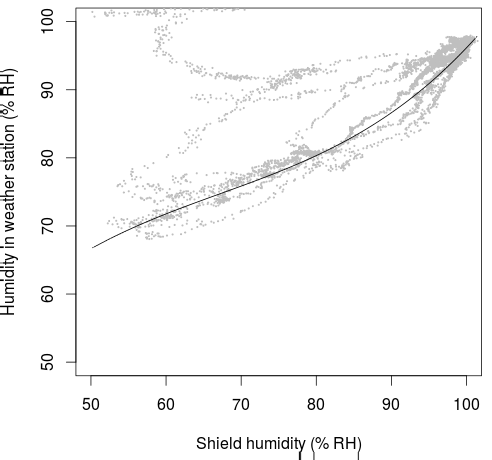
**Data calibration**

As it will be very difficult to increase the performance of the sensor shielding without also significantly increasing the costs of the shields, a regression analysis will have to be performed on the dataset. As is already shown in experiment 1B, increasing the temporal interval from 1 minute to 1- or 2-hours, will result in some data-smoothing. Non-parametric regression, in which the predictor is derived from the data, would be the easiest approach, and can include smoothing techniques such as smoothing splines and loess (local) regression. The limitation with smoothing is that it is a symmetrical approach and does not perform well on datasets in which the ‘peaks’ that will have to be smoothed are predominantly positive / asymmetrical – see figure X. Several smoothing methods have been tested: Turkey’s running median (3RS3S), smoothing splines (smoothing parameter of 0.5) and LOWESS regression (9 hour window), which all show values closer to the temperature in the station (Annex X). For this experiment the five separate days with experiments in different shields vs. the station have been merged to create a five-day period with different types of days (figure X). The temperature in the station is compared with the data in the 25mm & 50mm shields with the least adjustment (painted white or with no adjustment). One day lasts from 3pm to 3pm, which explains the long & short colder periods between days 3 & 4 and 4 & 5. During the warmer (midday) periods, the temperature in the shields is significantly warmer, while the minimum temperature is slightly lower in the tubes.

Smoothing is a solution if the shields cannot reduce the impact of strong radiation during the day, but is not preferred for many reasons. The impact of midday radiation might be reduced easily, but the required smoothing parameters will likely increase the minimum to a level that will cause new problems for further analysis when there will be more different types of days. Parametric regression analysis has many advantages over non-parametric regression, as the relationship can be adjusted to certain conditions. Common models include linear regression, generalized least squares, polynomial regression and general linear models. These can be used for forecasting new values that are not found in the original dataset. The main limitation in parametric regression is that you have to estimate the relationship, which - in these case of this project - will require a period in which the shields will be compared with an original weather station. This has been possible in this project, but might not be feasible in other regions. Another issue that will have to be taken into account is that a wide range of climatic conditions will have to be calibrated at the weather station to avoid values outside of the model. In the current project, it can be expected that the differences between shields and the certified station will be caused by the lack of aeration and reflectivity, which will result in a faster increase in temperature. The change in temperature per hour in the station vs the shields is plotted in figure X. There is a clear linear relationship (∆Tstation = ∆Tshields / 1.5) with a R2 of 0.87. This shows that the problem can be simply explained by the differences in heating/cooling between the structures. Correcting on this, however, will not be easy as this will result in cumulative errors over longer periods. Alternatives will be studied.



As five (quite different) days of 1-minute data from both the weather station and shields (although not all derived the same type of sensor shield) are available, this can be used in a parametric regression which can be used to calibrate the field-data. While the most common approaches are linear, least squares and polynomial, the approach selected here has been quantile regression. Quantiles are sets of a variable, which will divide a frequency distribution into equal groups – each containing the same fraction of the total population. The value 0.5 (2nd quantile) relates to the median, while other values that will be used are the 1st and 3rd (0.25 & 0.75) quantiles. The main advantage of this method over other methods of regression is that it is easy to avoid outliers – which will be one of the main problems in the shields. As can be seen in figure X, the parametric regression can be used to predict values between ± 20 and 34°C, this is based on the {quantreg} R package (Koenker 2011). As the field work will be conducted 200-700 meters higher (the weather station is at ±600m), the minimum temperature might be outside of this calibrated range. This is a limited problem, as the differences at lower temperatures are not as serious (< 0.5°C).



Different quantiles have been selected to fit temperature and humidity data, based on the quantile that seemed to provide the best fit. For temperature, the 3rd quantile has been selected (plotted in figure Xa), while the 1st quantile has been selected to correct the humidity data (figure Xb). Using the 3rd quantile for temperature will result in less smoothing at higher values, but it could result in values of ±30°C, which will be more difficult to get with the other two quantiles. The R2 of a linear regression is high (0.95), but does not capture the complexity of the trend. The above models can be used in corrections of values within the calibrated range. The models are not very accurate, and it would be better to improve the shielding to reduce the problem explained with the faster heating & cooling in the shields. The calibration has been based on a long period, but still faces some problems as other people used/cleaned the station. The high humidity values in the station can be explained by the cleaning (pressure cleaner) of the fence, which resulted in water entering the Stevenson shield and not the shields (which face downwards).

**Discussion**

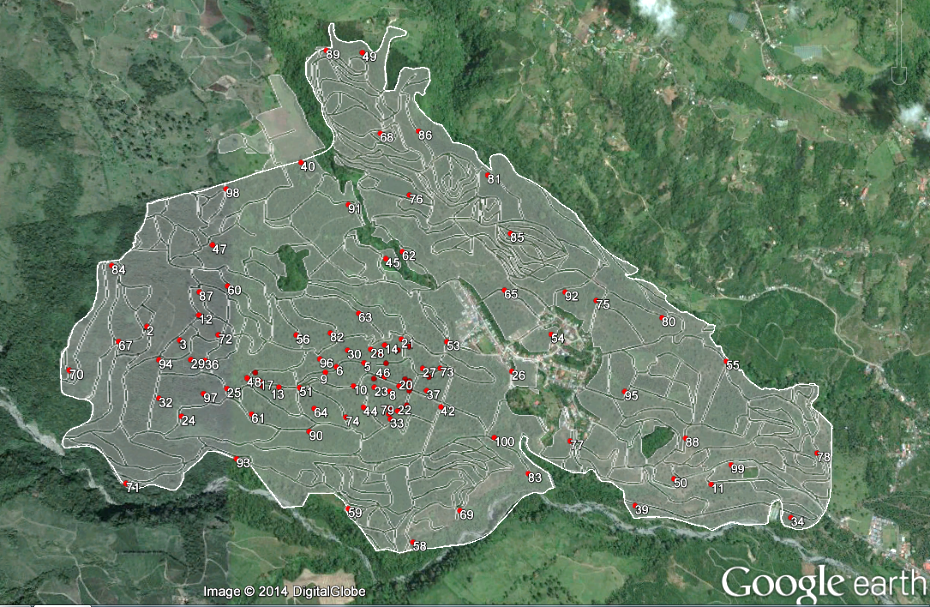
A preliminary conclusion which can be drawn from experiment 1 will be that working at a higher-resolution would not make a lot of sense for both temperature and humidity, as the reduced interval at which observations can be made will likely reduce the quality of the data-set more than this can be increased by the higher accuracy of the higher resolution. Especially for relative humidity, a high-resolution would not contribute meaningfully in the intended projects. The range of observations, as well as the standard deviation between sensors, is higher than for the low-res sensors. In this case, the low-resolution sensor even seems to be better. For temperature, the mean is very similar (within 0.02°C), while there is a large difference between the standard deviation in both datasets. This is not necessarily a big problem, as the maximum standard deviation at any moment is less than 0.4°C. While it is already quite clear that measurements at low-res would be fine for relative humidity, the impact of the low-resolution on temperature requires more study. Experiments with different starting moments and with different interpolation have found differences between the sensors, but only small compared to changing the temporal interval from 1 to 2 hours or when using different shields. There is, however, no clear ‘best’ shield based on the experiments. Another option that will be studied in a later experiment will be whether using the humidity sensors on all the iButtons would be useful at all; an alternative would be to only use this in one out of every *x* sensors, to be able to make more temperature measurements, while not losing a lot of information about the humidity.

Although increasing the temporal interval from 1 minute to 1 hour will result in some smoothing of the data, there does not seem to be a lot of difference between using random start-points or the closest value in the original dataset (unless using intervals of >1 hours). The smoothing that results from increasing the temporal interval is negligible compared to the increased range of the data that results from measuring inside PVC sensor–shields instead of the Stevenson screen, but data in the shields can be corrected with quantile regression when sufficient data (different types of day) is available of the data from the PVC shields and official weather station. Quantile regression is not ideal, as the actual difference between the measured temperatures is caused by the faster change (increase & decrease) of temperature in the PVC tube (around 50%). Correcting this is more difficult, as this requires changing the sensor sensitivity. Using a 2-hour interval in the PVC shields will still provide a larger temperature range than a 1-minute interval in the weather station will provide, but a 2-hour interval has the potential to miss important peaks during the day, and for that reason is not recommended. The main recommendation is to create a script that launches all the sensors at the same moment, while more research into sensor shields that can reduce the effect of midday radiation without adding too much costs, is also something that would provide a lot of benefits to similar projects. In addition to this, it is recommended to place several sensors near a certified weather station for the duration of the project, from which a model can be created with quantile regression. This model can be used to predict the values that have been observed in the field to avoid the impact of outliers on the further analysis.

4. Spatial aspects of the network

**Sampling & spatial distribution**

The Hacienda Aquiares is the largest coffee plantation in Costa Rica (9.95o,-83.72o), covering around 924 hectares, of which 673 are cultivated with (shade-grown) Arabica coffee. The farm is located between the Pacific Ocean (± 70 km) and the Atlantic Ocean (± 55 km). A more important topographic factor is the vicinity of the Turrialba Volcano, which is 3,340 meters and of which the peak is located around 7 km from the upper parts of the farm (close to 12 km from the lower parts). The farm starts just above 800 m and continues to close to 1400 m. The lowest sensors in this study has been placed at 832 meters and the highest at 1399 meters, giving a more than 500 meter altitude range to study. Coffee quality associated with a certain altitude, which will relative to the prevailing climate (temperature, radiation & humidity) and possibly to the different soils at the higher altitudes. The coffee with the best quality is considered to be located > 1,100 meter. This is around 45% of the area, but estimated to only supply 31% of the crop (Cornell University n.d.); this is likely the result of the natural areas that can also be found at the higher altitudes, where the inclination of the terrain is generally also steeper. The presence of natural areas can be seen in a map of the Leaf Area Index (LAI) of the farm, in which areas in the upper regions often have values that can be associated with that of forest (e.g. McWilliam et al. 1993). Sensors have been distributed over the area and sampling has not taken into account all different characteristics, although all altitudes ranges have been covered.



Sampling was based on a hexagonal grid inside one of the watersheds in the farm (*n=50*), while the placing of the other 50 samples was calculated by adding two points to 25 strata's (equal areas) in the study area. This has been based on the {spcosa} R package (Walvoort et al. 2013). Using strata's of equal area – the methodology is described in (Walvoort et al. 2010) - with two samples each, has been selected to reduce the impact of lost sensors. It was estimated that around 20% of sensors would be lost (stolen or in other ways irretrievable), based on similar projects in the region. In the end, the bad weather made it difficult to reach all points, which led to an assessment (together with local farmers) of other suitable locations in the vicinity of the original point. The number of points at different levels (50 in the watershed & 50 in the farm) has remained the same. A GPS has been used to store the location of all points, resulting in the coverage shown in figure X. When collecting the data (which was done by a local farmer), it was discovered that 7 were removed from the shield, while 5 could not be found and 2 other were not found in time for further analysis. Of the returned 86 sensors, 6 were removed from further analysis based on visual analysis. This was mainly the result of contact with water, which resulted in a very stable temperature and high humidity.

**Interpolation strategy**

Interpolation is often applied on spatial datasets which have a limited number of high precision observations that are used to predict reasonable values for a number of unknown points. Some of the best known approaches to multivariate/spatial interpolation include Inverse Distance Weighting, Delaunay triangulation, kriging, and spline interpolation; all these techniques are discussed in Mitas & Mitasova 1999. These are, however, all approaches to 2-dimensional interpolation. The data in this project includes the time dimension, which means an additional factor will have to be taken into account. Two approaches to dealing with the time dimension are reduction and extension (Li & Revesz 2002; Li & Revesz 2004): a reduction approach takes away the time dimension and applies one of the basic (2-dimensional) approaches to interpolation for any set in time; the extension approach will use time as another dimension and only works with a part of the interpolation techniques. While modifications could be made to several types of interpolation, the selection for the method in this research has been based on existing *R* functions to limit the time spent for this complex exercise. The *R-*package {gstat} (Pebesma & Wesseling 1998) includes a function to create different types of spatio-temporal variograms and subsequent use these for ordinary global spatio-temporal kriging, while the {spacetime} *R-*package (Pebesma 2012) includes additional methods and classes for spatio-temporal data. Ordinary kriging is a type of linear kriging that uses the spatial correlation in determining the predictor coefficients. As basic DEM information will also be available for all regions, this could be added in co-kriging of temperature. Including one or more covariates in the dataset (several can be derived from DEMS and the position of the sun) will provide a more smoothed output - reducing the impact of outliers (Odeha et al. 1994).

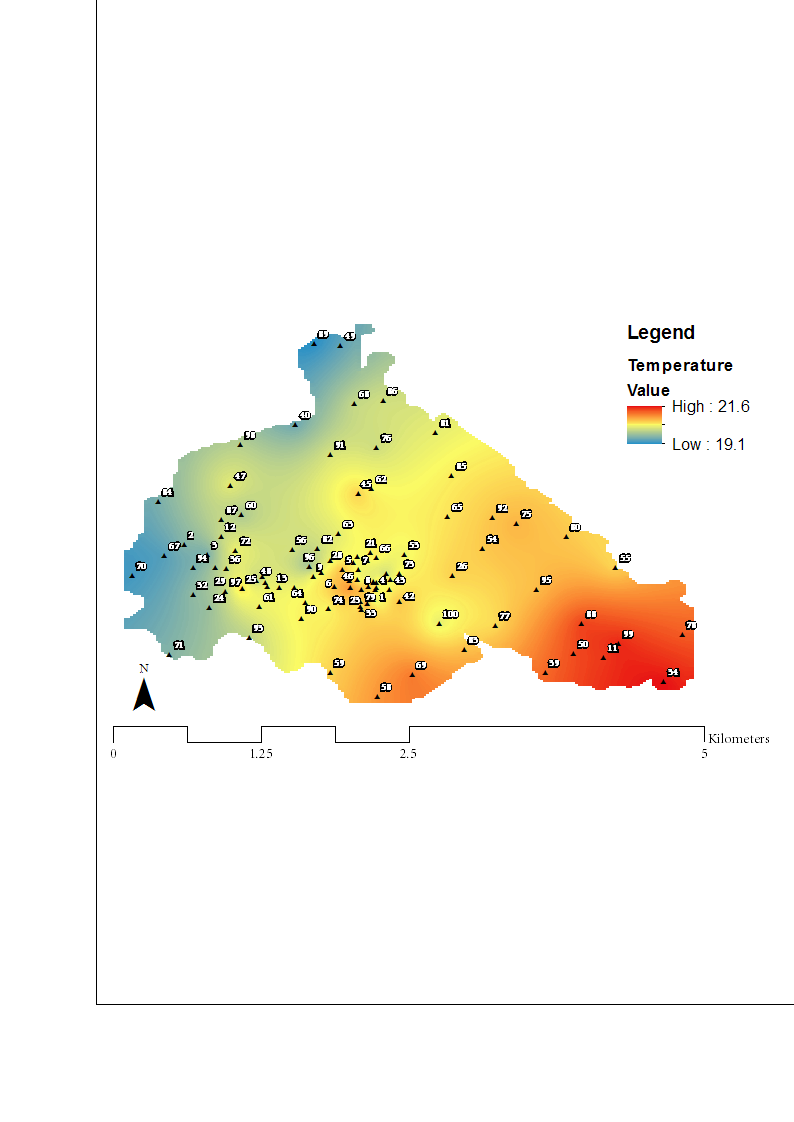
In order to apply spatio-temporal Kriging, variograms will have to be created for both space and time. Different approaches are available in the *vgmST* function of the {gstat} package that have different approaches on how to deal with the separate elements. While selecting values can be done on complex calculations, temperature is quite a basic unit to measure. At any time of day, there will be a correlation in temperature over relatively small regions due to the position of the sun (not taking into account shadows and clouds), while in time temperature will have a more complex relationship. During the night, correlation might be found for several hours, while during the day, this cannot be expected to be for more than two hours in complex terrains. The time-dimension is less important for temperature than the spatial dimension at any moment, which means a reduction model will likely provide equally useful information. When looking at the Pearson's correlation between columns with a) sensor data (spatial); and b) temporal data, the mean correlation is clearly higher for sensors (*r* = 0.99), compared to time (*r* = 0.59). The minimum correlation for sensors is 0.89, while this is < 0.01 in time. The minimum correlation when working with a 1-hour interval is 0.65, which drops to 0.47 for 2-hours and 0.34 for 3-hours. In the matrix below, the correlation matrix is provided when the mean temperatures are clustered per 2-hours. The correlation is high between the night-time (6 pm to 6 am), while during the day – especially between noon and 2 pm – correlation is very limited.

**Matrix of linear correlation coefficient between mean hourly temperatures**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Hour** | 00-02 | 02-04 | 04-06 | 06-08 | 08-10 | 10-12 | 12-14 | 14-16 | 16-18 | 18-20 | 20-22 | 22-00 |
| 00-02 | 1.00 | 0.99 | 0.99 | 0.70 | 0.29 | 0.29 | 0.13 | 0.21 | 0.62 | 0.87 | 0.92 | 0.97 |
| 02-04 |  | 1.00 | 0.99 | 0.74 | 0.31 | 0.27 | 0.09 | 0.15 | 0.55 | 0.81 | 0.88 | 0.94 |
| 04-06 |  |  | 1.00 | 0.74 | 0.33 | 0.29 | 0.11 | 0.18 | 0.57 | 0.82 | 0.88 | 0.94 |
| 06-08 |  |  |  | 1.00 | 0.76 | 0.47 | 0.16 | 0.10 | 0.36 | 0.50 | 0.57 | 0.63 |
| 08-10 |  |  |  |  | 1.00 | 0.76 | 0.46 | 0.34 | 0.39 | 0.27 | 0.28 | 0.29 |
| 10-12 |  |  |  |  |  | 1.00 | 0.83 | 0.70 | 0.64 | 0.41 | 0.40 | 0.35 |
| 12-14 |  |  |  |  |  |  | 1.00 | 0.90 | 0.68 | 0.34 | 0.30 | 0.22 |
| 14-16 |  |  |  |  |  |  |  | 1.00 | 0.82 | 0.49 | 0.44 | 0.34 |
| 16-18 |  |  |  |  |  |  |  |  | 1.00 | 0.88 | 0.83 | 0.75 |
| 18-20 |  |  |  |  |  |  |  |  |  | 1.00 | 0.99 | 0.95 |
| 20-22 |  |  |  |  |  |  |  |  |  |  | 1.00 | 0.98 |
| 22-00 |  |  |  |  |  |  |  |  |  |  |  | 1.00 |

Based on trial & error and some basic assumptions, the values for the variogram for space and time both have been given a sill of 1, while the space variogram includes a spherical model and as range the maximum distance (in kilometers) between any two observations, while the time variogram includes a linear model and a range of 2 (two hours) in the case of hourly observations (*see figure x*). Before continuing with analyzing the correlation between the temperature & humidity rasters with numerous other factors, a few statistical analyses will be undertaken to assess the current accuracy of the network. The *krigeST* function will result in an *STFDF* class from the {spacetime} package; once the spatio-temporal data has been added to this class, it can be converted to a raster (newly created function). This raster can be used for further statistical analysis and to visualize the data (including the possibility to do additional raster calculations) as \*.gif animation or as animation in Google Earth. The animations (both \*.gif and \*.kmz) are added to the online repository (data/animations), as they cannot be visualized in this document.

**Figure X: Mean temperature during the period 6 August – 6 September 2014**

**

Approximate location of a weather instrument

**Network accuracy**

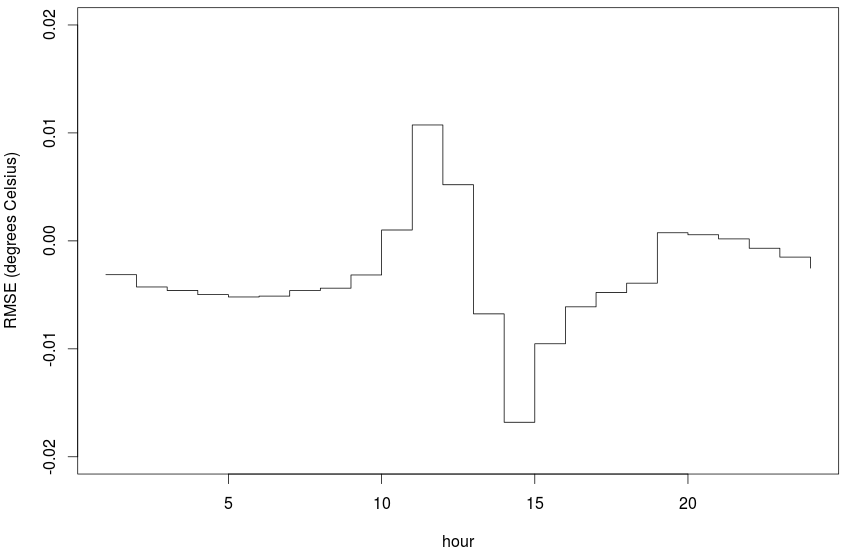
*Outside the network*

Different instruments are available in the study area that measure temperature. This includes instruments that measure air temperature, canopy temperature and soil temperature. The data with which the temperature data will be compared is the "AirTC\_Avg" data, which relates to the devices temperature sensor. The instrument is located close to the main road (see figure X), which is a relatively cool part of the farm. The mean during the period from 6 August to 6 September for this instrument was 19.75°C, with a range of 12.83-26.52°C. While the mean seems to be in line with the mean as measured by the sensors, the minimum and maximum values seem to be quite low. While this could be the result of the quantile regression (using the 1st quantile would result in values closer to these measurements), the location and quality of the instrument is not at a level that would require a renewed correction of the sensors. The instrument is placed on a tower at approximately 26 meters above the surface. The altitude where the instrument is placed is just over 1000 meters, which is 400 meters higher than the instrument at Aquiares. Assuming a similar change in temperature as can be seen within the farm, these 400 meters could result in a 2°C difference in mean. While this mean seems to be in line with what has been measured at Aquiares (± 22°C), the range is quite different from a normal range for this period at CATIE (18.59 – 29.85°C). The biggest difference is the height above surface (1.5 vs 26 meters), which are subject to very different topographical factors such as wind. For this reason, the instrument cannot be used to correct the observations by the sensors (± 15 - 28°C at this location).

*Within the network*

The current number of sensors is very high for a relatively small area and finding correlation between variables - and using regression kriging - can help to reduce the number of sensor that will be required. A first analysis that will be made regarding the network accuracy is on the results of the ordinary kriging. Some of the approaches that can be used to verify the accuracy of the interpolation include bootstrapping, cross-validation and jackknifing, which are explained in Efron & Gong 1983. The selection of the approach to data validation has been based on existing research on spatial-temporal interpolation methods. A publication by Kilibarda et al. 2014, has used a 'leave-one-out' cross validation to assess the accuracy of predicting daily temperatures at the global level; which showed some interesting finding for this research. An important concept in cross-validation is the root-mean-square-error (RMSE). The RMSE is a measure of the difference between predicted and actual values and aggregates the residuals into one aggregated measure. For the dataset with mean temperature, the RMSE has been calculated at 0.12°C, while this is 0.06°C for the dataset with minimum temperature and 0.22°C for the dataset with maximum temperature.

The maximum difference between the predicted and actual temperature is 0.31°C for the minimum temperature, 1.42°C for the mean temperature, and 2.56°C for the maximum temperature. The error is in all situations largest (positive) around 10 am, while it is lowest (negative for mean & max temperature) around noon. This can be the result of the fast increase in temperature in the shield compared to the increase in temperature in the shield. These results are generally in line with expectations, except in the case of the minimum temperature, which can be predicted more accurately than the mean temperature. Even with a smaller range, extreme values are often more difficult to predict than the mean value, as has been the case for the maximum temperature. The residuals, when comparing the predicted and actual value for the mean temperature, are plotted in figure X. This figure shows that the RMSE is close to zero at moments without sunlight, while difference become largest around noon. This is the moment when radiation will have the largest impact on the sensor shields, which can explain the larger errors.



When plotting the RMSE of the mean temperature of a sensor against the distance to its closest to neighbours, the trend shows that the RMSE is higher (both positive and negative) for sensors that have other sensors nearby, than sensors that are located in more remote areas. This is different from the finding in Kilibarda et al. 2014. Similarly, when plotting the RMSE of the mean temperature of each sensor against the altitude at which it has been placed, a similar trend can be seen. At lower altitudes and at higher altitudes, the RMSE is close to zero, while at the altitude of the watershed – in which sensors have been placed at a high density – the RMSE fluctuates more. This shows that it is harder to predict the temperature at sensors that are placed at high density in smaller area. One possible cause is that it is more difficult to find sensors that have not been performing well in areas that have a higher density of sensors, while human activity at this altitude can also play a role (e.g. vicinity of roads, touching sensors). This relatively high RMSE can be found for not only the mean, but also minimum and maximum temperature, and is clearly linked to the short distance to neighbouring sensors. Changing parameters such as range (which currently covers the entire area), when creating the variogram, can change this RMSE and will likely lower this sensors that are located close, but this will cause problems for the sensors that are located in more remote areas. An obvious solution would be to use a similar density throughout the study area and base the interpolation parameters on this.

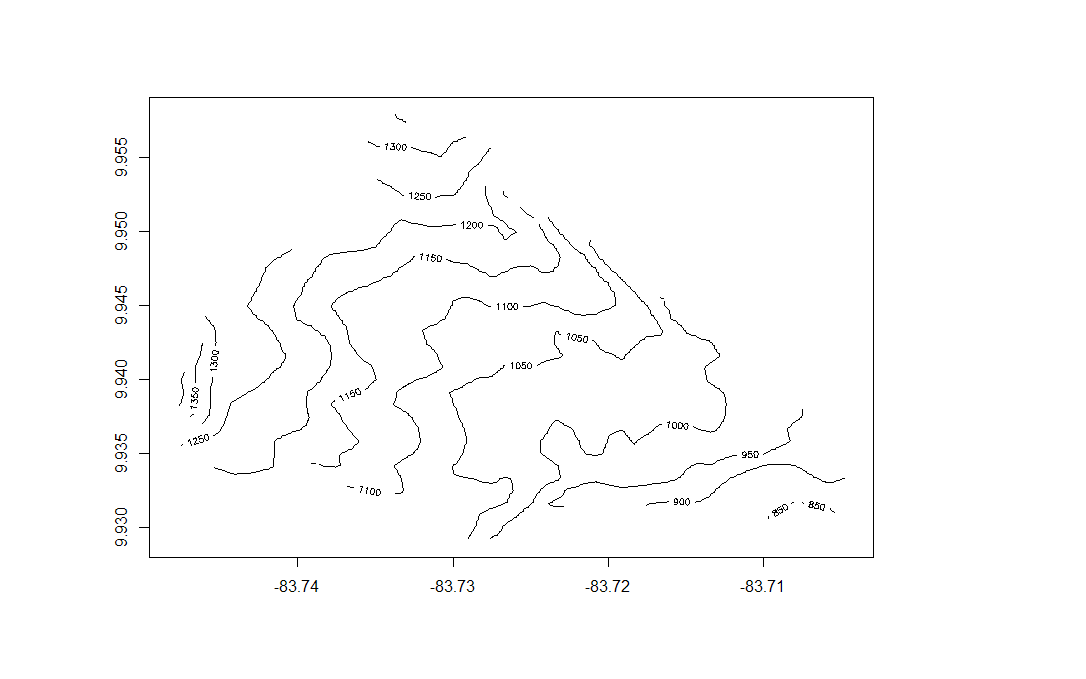
**Static covariates**

Covariates (or co-variables/predictors) are variables that can be used to predict another variable; the dependent variable. The best know covariates that can be used to predict temperature is elevation (Ishida & Kawashima 1993), while other covariates often relate to the position relative to the sun and factors that influence shading, such as the leaf area index (LAI) (Jin & Zhang 2002). Covariates can be used in predicting the temperature in different ways, such as co-kriging. Co-kriging is an interpolation technique that allows for several – generally more samples – variables to be used in predicting the dependent variable; this is most effective when there is a high correlation between the covariate and the prediction variable (Hartkamp et al. 1999). Different covariates can be used to predict the mean, minimum and maximum temperature. There is a general trend that more covariates result in better result (lower RMSE), although the gains become negligible after a certain number; this was five variables for the minimum and six variables for the maximum in a study by (Jarvis & Stuart 2001). Several data sources are available for possible co-kriging in this study; a SRTM Digital Elevation Map (Jarvis et al. 2008), a map of the Leaf Area Index (provided by CIRAD), and a LiDAR map of the vegetation height. In addition to this, two R-packages {insol}: doshade function (Corripio 2014) and {raster}: hillShade function (Hijmans & van Etten 2010), together with a calculator of the sun position, have been used to create a covariate that changes in time. As the maps of minimum, mean, and maximum temperature over a day have been based on a longer period (the middle day being August 22nd), this day has been used in the calculation of the sun position. For every hour the correlation between elevation, leaf area index, vegetation height, and sun/shadow will be calculated to see how this changes during a day.

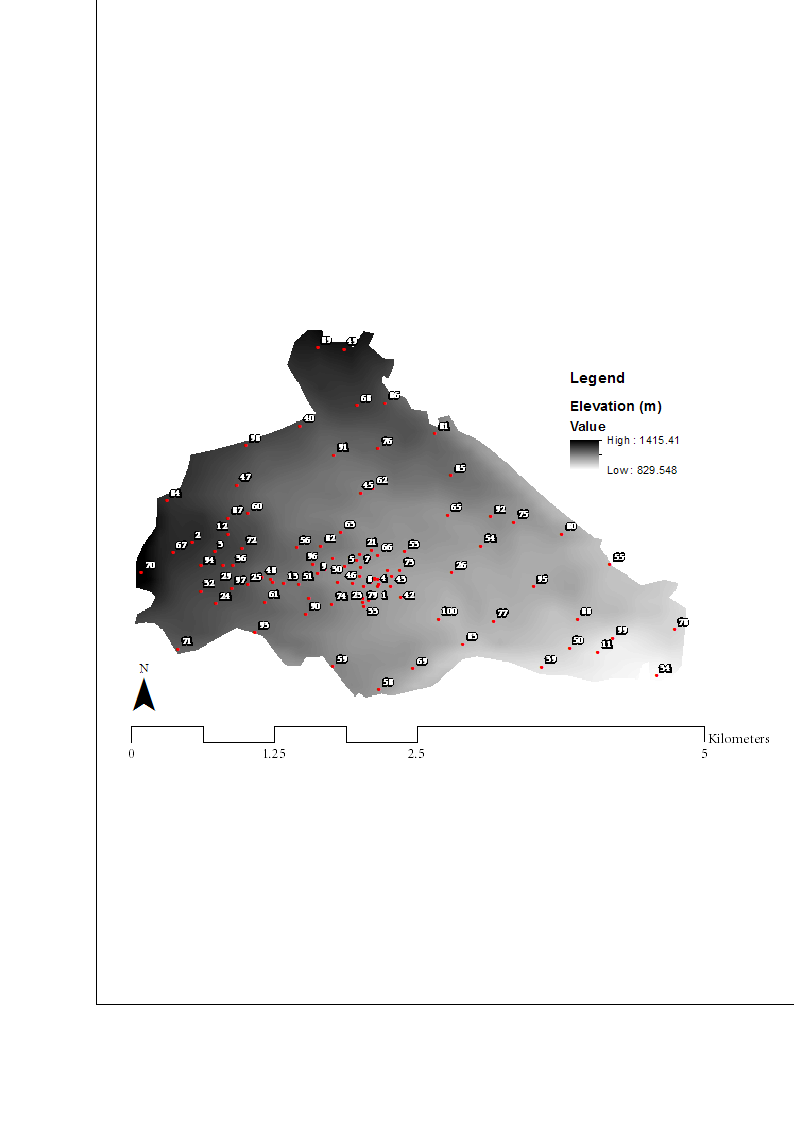
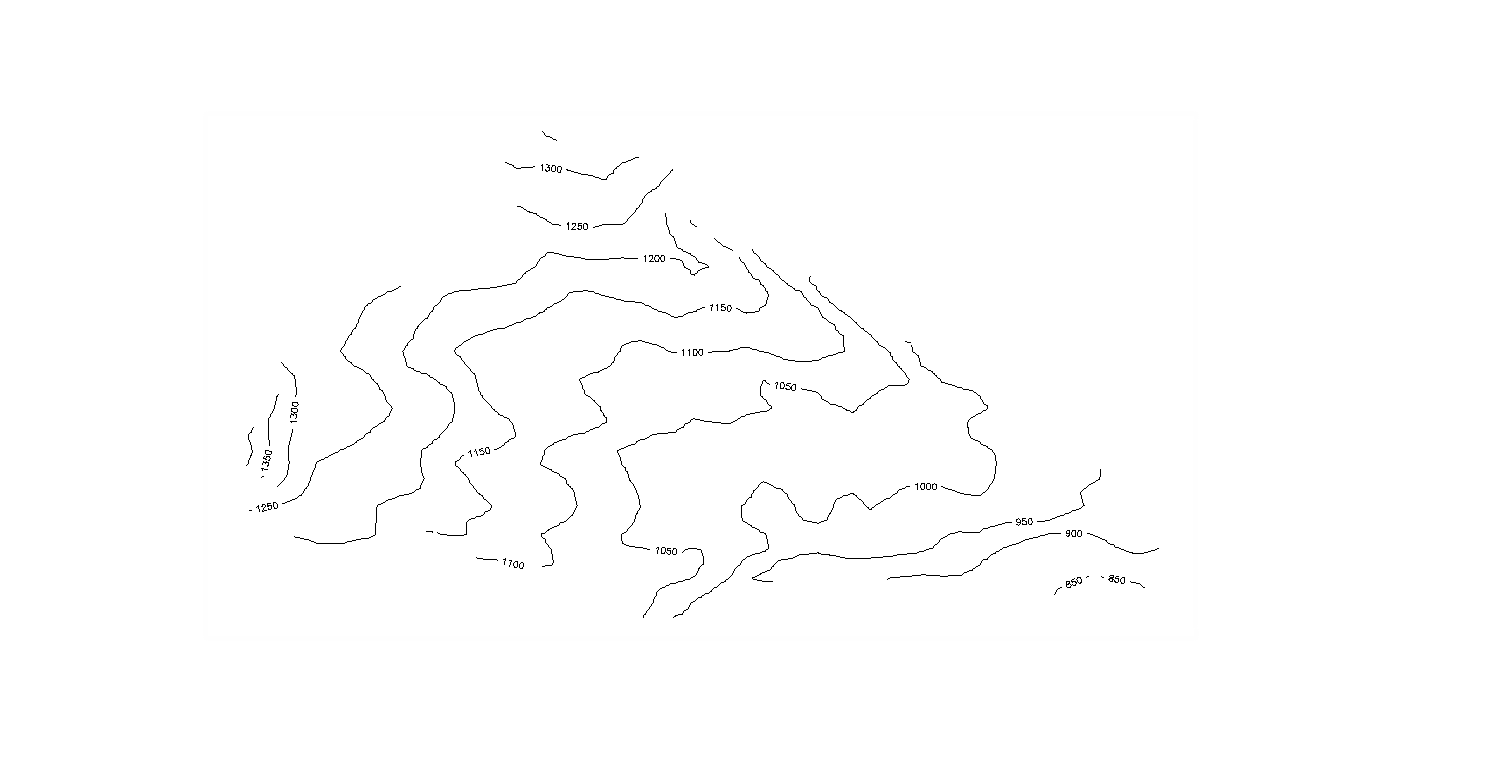
*Altitude*

The elevation has been taken as the provided value by the GPS and not by an extracted value from a disaggregated Digital Elevation Model, as the GPS value is expected to be more accurate in the complex terrain. The difference between the lowest and highest parts of the sampled part of the farm is over 500 meters; assuming a standard environment – the International Standard Atmosphere (Cavcar 2000), the mean temperature would decrease by just over 3 degrees Celsius (0.5 \* 6.5) from the lower to higher parts. No specific studies have been found on the impact on altitude on absolute temperature, but some factors that have to be taken into account include the higher solar UV radiation (Blumthaler et al. 1997) and topographic/vegetation aspects, such as more natural vegetation in the steeper parts and more impact of wind in the higher areas. The correlation (Pearson's *r*) of mean temperature with altitude has a range of -0.25 to -0.97, with a mean of -0.68. As expected, the correlation is strongest when there is no – or limited – influence of the sun (6pm to 7am). According to the day-length function in the {insol} package, the sunrise on August 22 was up just before 6am, while the sunset was just before 6pm. The plantation can be divided into two main altitude ranges; low altitude (<1,100 meters) and high altitude (> 1,100) meters.

The best quality (Aquiares Estate) is selected from the higher altitude regions, which covers just below half (48.6%) of the farm. Different in quality can relate to a number of characteristics, such as the soil type, but climate and related diseases can also play a role. The mean temperature in the higher areas is 19.9°C (sd = 0.33), while this is 20.7°C (sd = 0.34) in the lower parts of the farm. The humidity at the higher altitudes has a mean of 89.7% RH (sd = 1.26), while the mean in the lower areas is 90.5% RH (sd = 1.40). While the relationship between temperature and altitude is clear (less so for altitude and humidity), the differences in the area are relatively small. It is unlikely that the differences in temperature alone (<1 °C) between the low and high altitudes can explain differences in quality.



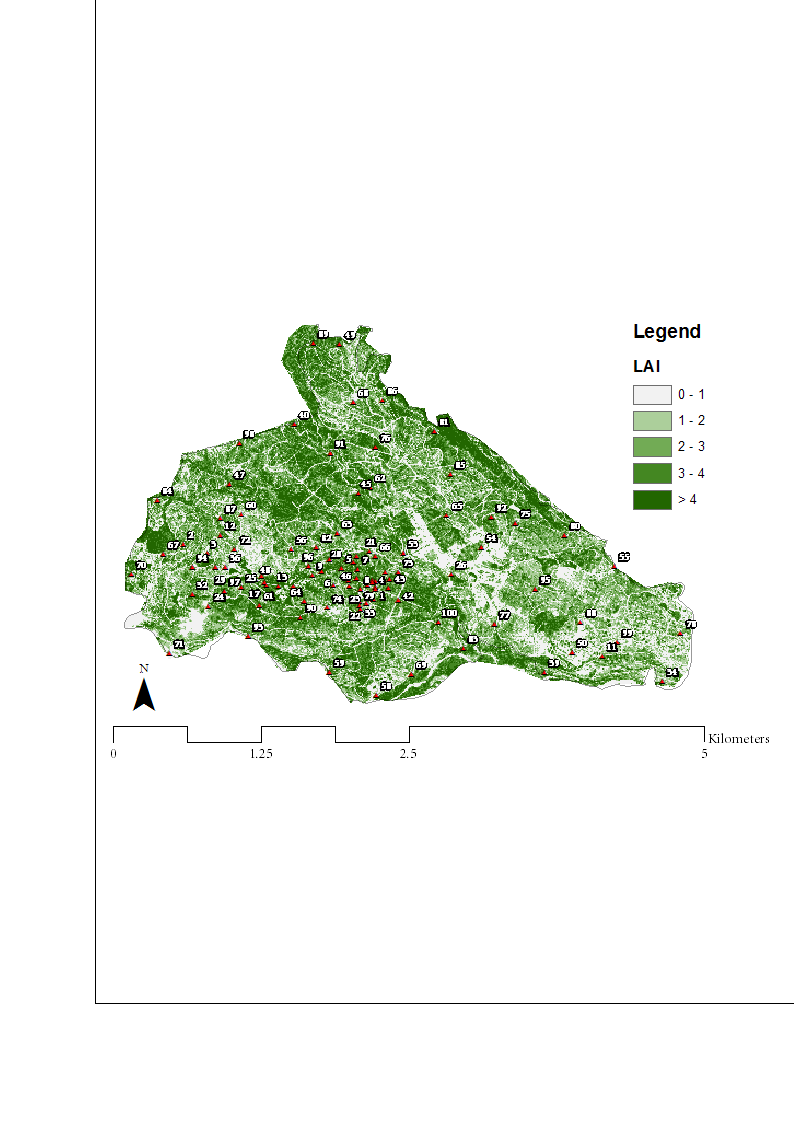
**Figure x: Digital Elevation Map (resamples SRTM4) of the Aquiares Estate**



For minimum temperature, the correlation with altitude has a range of -0.42 to -0.98 and a mean *r* of -0.78, which is a stronger correlation than for the mean. The trend during the day is the complete opposite of that of mean temperature, with the lowest correlation during the nighttime, while the correlation is high during the day. For the maximum, the range of the Pearson's *r* has been 0.09 to -0.95, with a mean of -0.53. The trend is the same as for the mean temperature, with highest correlation during the night, but low correlation between 8am and 6pm (-0.16), compared to -0.51 for the mean temperature and -0.90 for the minimum temperature. Correlation between temperature and humidity has been 0.32 to -0.99 (mean = -0.94) for the mean, while the minimum and maximum temperature are better correlated with the other extreme of humidity. Maximum temperature and minimum humidity have a mean *r* of 0.91, which is -0.77 for maximum humidity with minimum temperature.

*Leaf area index*

The Leaf Area Index (LAI) is a variable that provides information about the total area of one-sided photosynthetic tissue per unit of ground surface over which it is measured (Watson 1947); this is index is dynamic and changes depending on season and age/management practices of the crops (e.g. Welles & Cohen 1996) The approximate range is from close to zero for deserts to > 20 for boreal evergreen forests; tree plantations are generally the biomes with the highest LAI with an average of 8.7 (Asner et al. 2003). For coffee in the study area, the LAI has been measured at 4.0 (+- 2.9) in 1982 (Ewel et al. 1982), although parts of the farm are replanted at a frequent interval. The currently grown variety (Caturra) is characterized by relatively limited vertical growth (< 2m) and planted in rows that are 2m apart; other trees are also planted in the farm (Dauzat et al. 2001). While the impact of leaf area index on near-surface climate has been well studied (Buermann et al. 2001), the impact on temperature at/above the canopy level has received less study. Above and inside canopy climate and its relationship to the LAI has been well studied in Goudriaan (1977) and includes – amongst others – canopy geometry, incoming radiation, and optical properties. In this study only the actual LAI is available over a larger areas. As the sensors are placed adjacent to the crops, there will not be a shading impact of the coffee trees (shading is possible by other larger trees). The impact of LAI on temperature can both be positive and negative; the areas with higher LAI can provide more shelter to wind, while higher LAI can also result in more reflection of incoming radiation (Sellers 1985). As LAI can be estimated by the spectral reflectance of red and near infrared light (Fuchs et al. 1984), and infrared reflectance can also be used to assess surface temperature (Petitcolin & Vermote 2002), there might be correlation between temperature and LAI.



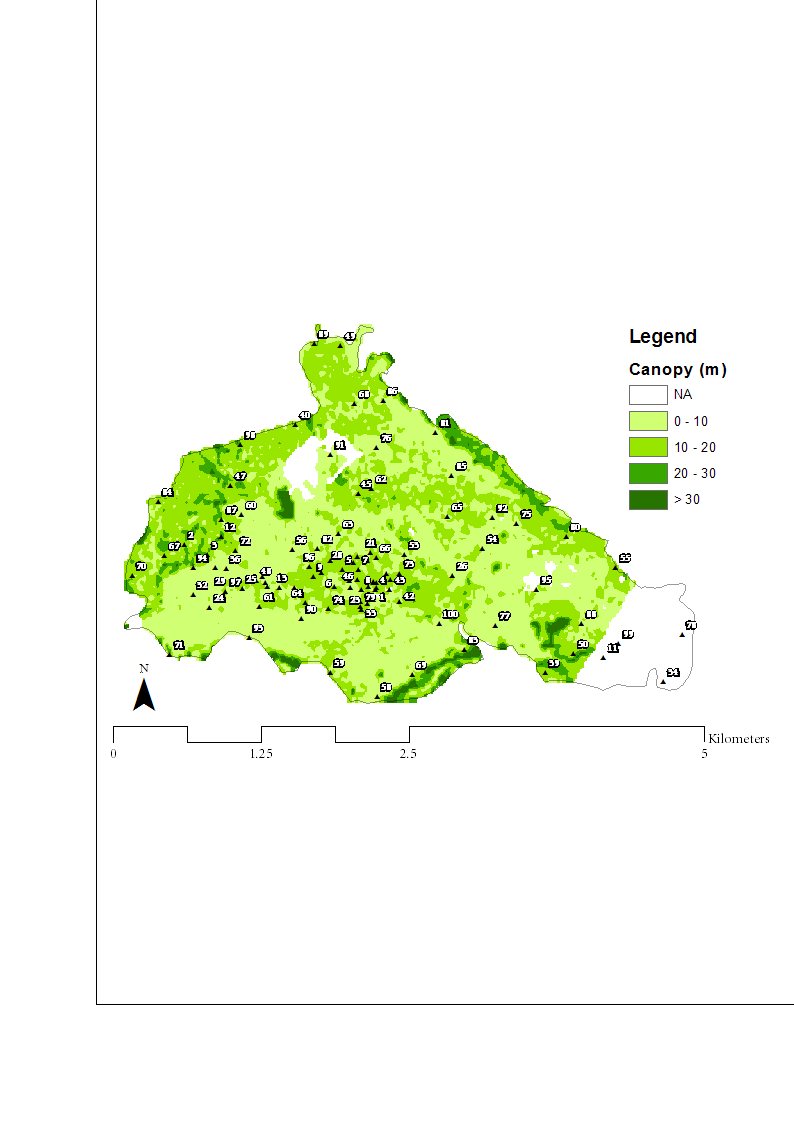
The map of LAI that has been used is based on High Resolution Multispectal Images (MODIS) image together with field verification. Earlier studies based on the LAI in Aquiares have found that the LAI varies from 2.4 to 4.4 depending on the season (Taugourdeau et al. 2014); this study also the inter-annual variations that can be caused by pruning and plot renovation – which can have an impact on this study on the relation with local climate. While correlation between leaf area index and temperature would make sense in multistory landscapes, the average LAI in the Aquiares farm in the used raster is only 2.56 with a minimum of 0 and a maximum of 8.98. This map already removed the roads and village that are located inside the farm. An average LAI of 2.56 is comparable to that of grassland (Asner et al. 2003) and is thus very low for a tree plantation. As the sensors have been placed at 1.20 meters and have been placed next to the coffee crop, the only impact on temperature would come from larger trees. The coffee crops are replaced every x years (reference), which also reduces the reliability of the data of the LAI map.

The mean Pearson's r that has been found (25m radius around the sensor) between LAI and mean temperature is -0.07 (range: -0.01 to -0.15), -0.02 (range: 0.04 to -0.13) for minimum temperature and -0.05 (range: 0.06 to -0.14) for the maximum temperature. At no moment the correlation between any of the temperature statistics and the temperature has been higher than -0.15. There seems to be some correlation between the leaf area index and temperature (a higher density results in a lower temperature), but this is relatively limited and – in combination with the fact that maps of LAI are not easy to get – make this variable of limited use as covariate in co-Kriging in similar projects.

*Vegetation height*

The height of vegetation is another factor that can be related to the climate in an area, which might partly overlap with the leaf area index, as more natural areas will likely have a high LAI and canopy height. Trees are often grown together with coffee crops in shade-grown coffee systems. Benefits include protection from wind, an improved microclimate and the addition of organic matter to the soil (Budowski 1980). In addition to this, shading can reduce the temperature stress (Butler 1977), reduce pest outbreaks (Staver et al. 2001), and overall provide a refuge to biodiversity (Perfecto et al. 1996). The data in this study is derived from the Laser Vegetation and Ice Sensors (LVIS) (Blair et al. 1999), which is a scanning laser altimeter instrument. The scanner can create waveforms that provide information about the canopy height (canopy top reflection and the ground reflection). The images provide the mean ground elevation of the lowest detected mode in each waveform as well as the elevation of the highest detected altitude in this waveform. As this data includes the absolute minimum and maximum in the study area, the values are an exaggeration of the actual canopy, but show a similar trend than would be the case if the median height would be taken; for correlation this will not make a difference. The used dataset is a 2005 Costa Rica dataset (Blair et al. 2006) that covers the area around the Turrialba volcano close to which the farm is located. Some processing of the data had to be done, as the points are not regular (as these are the locations of the highest points). The resulting raster is shown in figure x, and shows that vegetation is highest in the areas around the rivers and in the highlands. This is caused by the difficulty of the terrain for growing coffee in these areas and additional environmental policy that might be in place; 150 hectares of the farm are covered by a protected natural forest (Anon 2008). The average height in the farm is 10.9 meters with a maximum 76.5 meters. The vegetation height is in general higher towards to boundaries of the farm, which could be caused by the difficulty of the terrain in these areas or might be done on purpose as management strategy (natural farm limit and protected area for certifications – which includes a certification by the Rainforest Alliance). It has to be noted that measurement errors can play a large role in this map, as it does not only show the canopy height, but also the complexity of the terrain as the height difference within the collection size (10 to 25m) is also included in the calculated height.

The average vegetation height in a 25 radius around the sensors is just over 10 meters, with a minimum of 4.3 meters and a maximum of 21.5 meter. This might be a slight exaggeration of the actual height as the highest value from each cell has been taken, but the trends in the map are in line with what has been seen in the field. The areas near the rivers and at higher altitudes have taller vegetation, as there is more natural vegetation in these areas that are hard to access. The correlation between the vegetation height and mean temperature has a range of -0.02 to -0.30 (mean *r* = -0.20). For minimum temperature this range has been -0.08 to -0.32 (mean *r* = -0.20) and for maximum the range has been 0.13 to -0.32 (mean *r* = -0.15). This is, after altitude and leaf area index, a third variable with a negative correlation on temperature. The correlation seems to be stronger during the night, but does not show clear trends during the day (figure x). The correlation with maximum temperature is lowest during the day, indicating a low impact of the vegetation on shade cooling. This shows that shading will be hard to estimate from these maps when not taking into account the angle of the sun. Due to the similarities between the map of slope and vegetation height, the correlation between slope and temperature has also been calculated; the mean correlation is only *r*=-0.10, but the range is higher, with values up to -0.39 around the time the sun goes down.



**Dynamic covariates**

*Introduction*

The previously discussed variables are all assumed to be relatively static during a short period; leaf area index and vegetation height will change during a season, but will generally remain similar during a day. Temperature and humidity are both dynamic variables that are to a large extent influenced by the altitude and position/intensity of the sun. The position of the sun can be related to different covariates of temperature, which include: 1) hillshading; 2) cast shadows; and 3) solar radiation and resulting insolation at each hour. This will be of relatively little use when excluding cloud cover, but the data against which the influence of solar position will be tested is the average of a longer period and the area is sufficiently small to assume cloud conditions will remain similar. Very complex modifications can be made to the dynamic covariates to capture their impact on certain variables (e.g. additive models), as explained in Hong et al*.* (2014), but this study will limit itself to simple linear models. How to combine static and dynamic variables in species distribution models (climate change)) is discussed in Stanton et al. (2012). This aim of this section is to find temperature predictors at moments when the impact of altitude will be limited.

*Sun position*

The sun can have an influence on the temperature in different ways; the position of the sun during the day can result in positions that have more shading than others, while the solar radiation during different times of day will also result in different rates of heating. The time of highest insolation (amount of solar radiation on a given surface) is the solar noon; the moment when the sun crosses the meridian at the selected site. For the studied day (August 22nd 2014), the solar noon was at 11.38 am (NOAA n.d.). Both the temporal and spatial distribution of insolation can be understood and modelled; important factors include latitude, time of year, cloud cover and topographical factors such as slope and aspect. Cloud cover data can be derived (real-time) and processed from different satellites (e.g. NOAA’s CLAVR-x system (Thomas et al. 2004)), while modelling the impact of slope and aspect on solar radiation in GIS already has a long history (e.g. Dubayah & Rich 1995). Solar radiation can be related to daily minimum and maximum temperature (Bristow & Campbell 1984) and to productivity in tropical ecosystems (Monteith 1972). Modelling the relationship between solar radiation and temperature range requires different empirical coefficients and for that reason is different to implement in models. Calculations of the solar radiation at a certain moment at a certain latitude is relatively straightforward and could be combined with DEMs to model the insolation and shadow during different times of day. The studied farm is relatively small and located on a south-east facing slope with a mean slope of 12 degrees. The aspect and lack of large objects on the south side ensure that the farm gets sun for most of the day; only in the late afternoon there will be an impact of the volcano in the north.

*This section will mainly be based on the functions in two R-packages and will verify the outcomes with data from the CATIE meteorological station and calculations with additional functions, such as the sunPosition function (SO answer by Josh O'Brien - based on Spencer 1989). The packages that will be used are the {insol} package (Corripio 2014), and the {raster} package. Both packages include a function to calculate hill shading which both have very different outcomes. As the calculated insolation depends on the solar intensity perpendicular to the surface, this function will provide the most important results.*

*Hill shading*

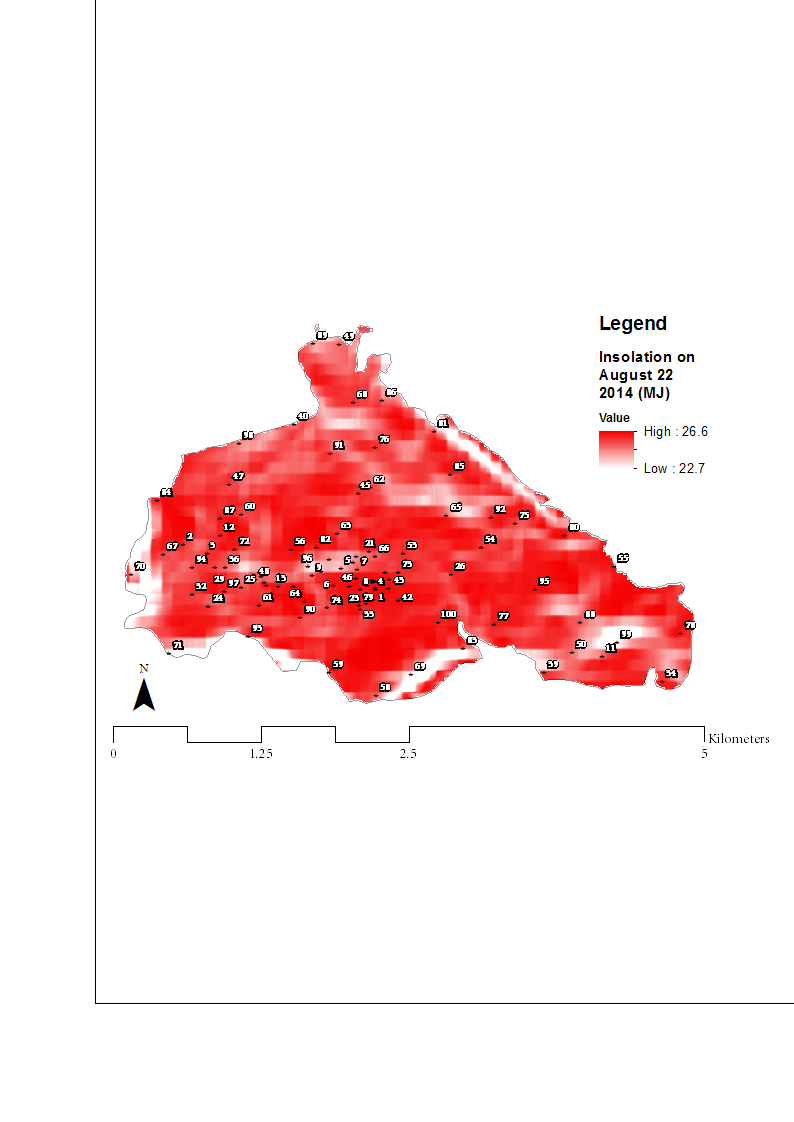
Calculating hill-shade is a well-known functionality in many GIS software, but is often mainly used for aesthetic improvements of maps. The actual meaning is often not discussed, but will have significant implications for the amount of insolation at a given location. The ArcGIS Resource Center summarizes the *Hillshade* as a function that ‘creates a shaded relief from a surface raster by considering the illumination source angle and shadows’. The output of the algorithm is similar to the one used in the {raster} package; values range from 0 to 255, but are often converted to an easier to interpret range, such as -1 to 1. Values of 0 (or -1 in a non-normalized calculation) relate to shadow areas, while values of 255 (or +1) can be found in areas with complete illumination. The calculation (simplified) used for this function in the {raster} package is hillShade = int(cos([Slope] / 180 \* 3.1415927) \* 255). An approach that is often used to convert this value to the insolation is I = S \* Hillshade/255, in which I is the insolation and S is an insolation constant. Working with this formula provides the same results as working with a non-normalized version of the hillshade function and resetting the negative values to zero.

The *hillshading* function in the {insol} package provides a similar (-1 to 1) range, but negative values are set to zero inside the function. One important issue to take into account with this function and which can cause slowing down of calculations, is that the DEM has to reprojected to have squared cells (*this was not clear in the package vignette, but this was clarified by the package author*). In this case the DEM has been reprojected to the ESPG:3857 (WGS84 Web Mercator) projection with 100m resolution. Calculating the direct insolation based on this function is done by an easy hillshade \* direct irradiance calculation. When adding the hourly values that are provided by the hillshade function in a DEM with a 10km buffer around the study area, the normalized function divided by 255 gives daily values between 5.0 and 7.9, with a mean of 7.6; the hillshade function in the {insol} package provides values between 4.3 and 7.9 (mean =7.6). There are some implications of these minor differences for the calculation of daily insolation; the direct irradiance at noon on August 22nd was 747 kW/m2, while diffuse radiation was 242 kW/m2. The range and mean of insolation in the study area, based on complete or direct-only radiation, is provided below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | CATIE Weather station average daily radiation | {insol} Complete (MJ) | {insol} Direct (MJ) | {raster} normalized  Complete (MJ) | {raster} normalized  Direct (MJ) | |
| Mean | 16.9 | 26.1 | 17.8 | 26.2 | 17.8 |
| Min | 12.8 | 23.8 | 15.5 | 23.6 | 15.3 |
| Maximum | 20.8 | 26.7 | 18.4 | 26.7 | 18.4 |

*Cast shadows*

Except for the impact of the local relief on the insolation, there will also be an impact of cast shadows (sky view factor). Due to the relatively small size of the study area and its orientation, this is not expected to be a significant issue at this level, but even at this scale there will be places where the cast shadows will have a larger impact throughout the day. This will be places close to larger objects (e.g. mountain tops) or in areas with steep slopes (e.g. rivers). The output of a layer of cast shadows provides values ranging from 0 (full shadow) to 1 (no cast shadow), and for this reason can easily be multiplied by a hillshade raster when calculating the insolation. Cast shadows have been calculated at 10 minutes intervals throughout a day, taking into account objects within a 10km buffer of the study area to include all the larger mountains and volcanos in the region. When dividing the sum of the shadows at 10 minute intervals by the length of the day (12.3 \* 6), the range of the periods in direct sunlight becomes 0.81 to 0.95. This indicates most of the area has a clear sky view for most of the day. The areas with a lower value are located at the border of the farm (river) and at the high altitudes where the impact of neighboring objects might be clear. Based on the figure of the shadows, it seems that there is some correlation with the Leaf Area Index and vegetation height. The correlation between the three (hill shade only, cast shadows only and combined insolation) static maps and temperature throughout the day is in most cases positive, but varies a lot throughout the day. When only considering the hill shade, correlation with mean Ta ranges from -0.14 to 0.12 (mean = 0.05), with minimum Ta this ranges from -0.04 to 0.23 (mean = 0.08) and with maximum Ta the range is -0.18 to 0.15 (mean=0.03). For cast shadows the correlation with mean temperature the range is 0.09 – 0.34 (mean = 0.25), with minimum it is 0.10-0.37 (mean = 0.24) and with maximum temperature the range of correlation is -0.12 to 0.38 (mean = 0.18). Insolation is basically a combination of both types of shadow and provides a correlation of -0.02 to 0.32 (mean = 0.14) with mean Ta, -0.04 to 0.37 (mean = 0.19) with min Ta and -0.18 to 0.34 (mean = 0.12) with max Ta. Correlation with maximum temperature – which is the biggest problem – remains outside a useful level.

**

*Hourly insolation*

The correlation of minimum and mean temperature with the DEM will be useful to assess the climate at a local level, but correlation with maximum temperature has not provided any useful correlation so far. As the maximum temperature has often been linked with the midday radiation – which should be especially relevant for the sensor shields that have been created, it can be expected that there will be a link of maximum temperature with insolation at a given moment. For this reason, two types of dynamic correlation will be studied. The first will focus on the modelled insolation at any hour and link this with the 80 sensors in the field. The second will be a cumulative model that adds the insolation of one hour to the insolation so far on the day, to see whether the correlation becomes better when adding the insolation throughout the day. The main period of interest are the two hours before and after noon, as the temperature at different sensors cannot be well-explained by any of the static covariates. The first analysis will be that of a possible correlation of insolation with min, max and mean temperature. The matrices that will be compared to find correlation contain 24 rows (hours) and 80 rows (sensors). When creating a linear model of all values, a clear positive trend can be seen (17.74 + 2.33x, R2 of 0.91), which indicates that a higher insolation (in MJ/h) is related to a higher temperature. As the insolation is modelled and based on the solar position, the values are highest around noon, which is the same for the temperature. The Pearson’s correlation at each hour (between 5am and 6pm) is relatively small, with the highest value when the sun comes up (*r* = 0.39 at 6am) and in the afternoon (*r* = 0.40 between 1 -4 pm). Between 10 am and noon, correlation is very small with an average r of -0.03. A model of cumulative radiation can be expected to have a better relationship with temperature in the morning, while this will be limited in the afternoon when the cumulative value will be high, but temperature will go down. The cumulative model has only been able to provide a similar correlation at 6am (*r* = 0.39), after which the correlation becomes low; at noon the correlation is negative with a Pearson’s *r* of -0.33.

Of the static and dynamic covariates, the two most interesting are the elevation and the insolation. Elevation has a well-studied relationship to temperature, but cannot predict differences during peak sun-hours. Solar radiation has a clear impact on the temperature inside the sensor shields. The extremes have already been corrected in an earlier step, but the values that had extreme values will still have higher values than sensors that already had normal values. The standard function for ordinary global Spatio-Temporal Kriging (krigeST – {gstat}) accepts adjusted formulas that include linearly independent variables. Making adjustments to this function will not be possible without taking the time-dimension out. A solution would be to create a stack of rasters with different covariates at different times. As altitude has the strongest correlation with mean, min and maximum temperature at almost all moments (and is only marginally passed by other variables at the other moments), including osanly a DEM as independent variable would be the best solution. This will be the easiest solution – as existing functions can be used and DEMs are available from all regions, but this will also be the safest options, as other correlations are less well-studied.

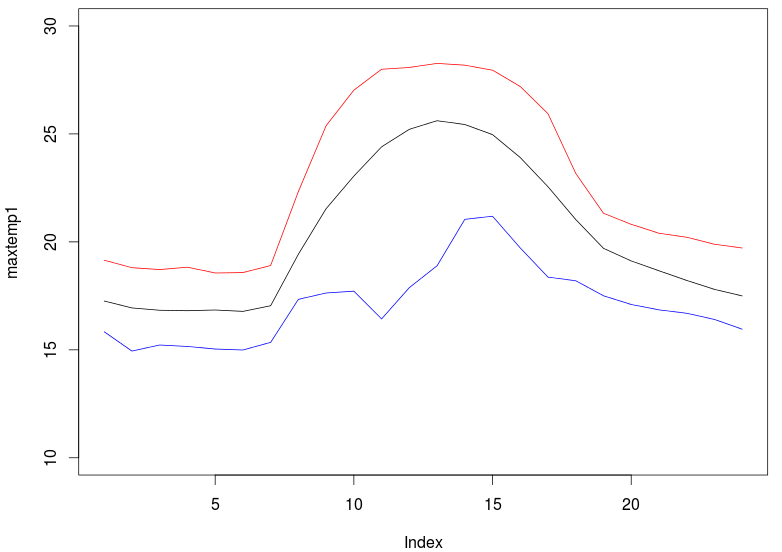
*Discussion*

The accuracy of the network seems to be quite good when using a ‘leave-one-out’ approach, which is better for minimum than for mean and which is worst for the maximum. Of the available covariates, the altitude provides the best correlation with the temperature data. The average Pearson’s *r* for altitude with mean temperature is *-*0.68, -0.78 for minimum temperature and -0.53 for maximum temperature. Using the z-value from the GPS (> 7.5 higher on average) provides almost the same correlation as an aggregated DEM, based on data with a 90m resolution. The correlation with leaf area index and vegetation height with temperature is negative, which might be expected, but also statistically insignificant with a maximum correlation of *r* = -0.20. As these datasets are not easy to find for other regions, using this data as covariate will be of limited use in similar projects. The position of the sun at different days and throughout the day can result in the creation of daily maps of cast shadows, hill shades and insolation (when multiplying the orientation of the surface by the solar radiation at a given moment). These variables can be created for both full days as per hour and are relatively simple to compute based on a DEM and four functions (hill shade, cast shadow, sun position, and insolation). While this was expected to result in clear positive correlation with temperature, and especially maximum temperature, the correlation has been relatively limited. Of the daily maps, the strongest correlation has been with cast shadows and mean temperature (*r* = 0.25), while the mean correlation has in all cases been positive. The impact of solar radiation depends on the sensor shielding and the objects in the vicinity that can provide shading, so this will be difficult as a covariate in studied. Cast shadows will be interesting to include, but this will include large adjustments to the existing functions that perform space-time interpolation.

*The above section on covariates will be significantly reduced and more attention will be given to different interpolation techniques. The two main alternatives are directly working with rasters and ignoring the time-dimension (which allows for covariates) or interpolate in space and time (ignoring covariates) and only convert to rasters later. Both have advantages and disadvantages that will be discussed. Raster is probably faster and works with less packages (although interpolation is done in other packages), while using the {spacetime} package is more specialized for interpolation, but less useful for the subsequent calculations. Tests will be done on user friendliness (speed, number of packages, clearness) and accuracy (RMSE, points required, incorporation of covariates).*

5. Practical implementation

This section will deal with the practical implementation of a network of micro-sensors and will provide the findings from a trial with 100 sensors in the Aquiares coffee farm in Costa Rica. The farm is located at an altitude of 800-1300 meters at an approximate latitude of 9.94o and longitude of -83.73o. The farm has been selected because it covers a large altitude range and is close to the CATIE; this has resulted in a lot of studies in the area which can be used as input to this research or to compare the results. According to the official website, the size of the farm is 917 hectares, of which 675 is planted with coffee. In this study, a polygon has been drawn that covers an area of 840 ha, which includes the village in the center of the farm, but does not include some of the upland areas where no coffee was grown. This area has been selected based on the knowledge of the boundaries of some of the local farmers. The temperature and humidity has been measured during one month (August 2014). This is a month that is normally characterized by constant high temperatures with a daily maximum temperature range between 25-30oC. The month of August is part of the rainy season, which means the sky is cloudy most of the time. Humidity is also high. The (corrected) measurements during this research are well in line with the expected climatic conditions; figure x shows the minimum, mean and maximum temperature (average of the 80 sensors) during the observation period.



*- Introduction to Aquiares*

*- Introduction to coffee*

*- Climate & coffee quality*

*- Climate & pests/diseases*

*- Discussion*

The data derived from a network of micro-sensors can be used in different ways; it can improve the understanding of the local climate and be used in more accurate interpolation, but in this case an example will be provided of the possible use in more accurate decision making at farm-level of the selected varieties and of periods in which pest-monitoring or application of agro-chemicals would be best based on the climate. This is not at the level of precision agriculture that is possible in many developed countries, but would be a clear improvement in areas where high-resolution satellite data and/or weather stations are not available or not accurate enough for the intended uses.

*Coffee and alternative crops*

Two main types of coffee exist: *Coffea Arabica* and *Coffea canephora* (robusta). These are adapted to different growing conditions, which can be summarized as lowland conditions for robusta and highland conditions for Arabica. The annual temperature for both crops has to be between 15-300C, while coffee tolerates shade and requires soils with a permeable soil (Pohlan and Janssens, 2012). The coffee at Aquiares is arabica, which is close the optimal range for this crop. As the tolerated climatic range and altitude range for Arabica is quite broad (<2800 m, 14-280C optimal temperature and 1400-2300mm optimal rainfall), this is less a problem for coffee than pests and diseases. Pest populations generally stay within tolerable limits due to the perennial nature of the crop and lack of climatic extremes, which results in a good level of biological control by diseases and natural enemies of the pests (Bardner 2006). The main risks to coffee in Costa Rica include rust and nematodes (Waller et al. 2007). Most pests and diseases are linked to certain climatic thresholds, which can be analyzed for the farm in Aquiares. Different solutions can be recommended in areas with high risk to numerous risks and diseases, which range from basic Integrated Pest Management (Natali de Oliveira et al. 2003 to changing to different crops if the situation is already problematic and can be projected to become even worsened in the light of climate change. The latter is drastic, but some alternatives will be recommended for the region, by linking the climatic range to other crops that can be grown in the region.

For this analysis, the EcoCrop dataset (1710 crops in the {dismo} packages) will be used, as well as a newly created dataset with 10 pests and diseases that can create problems to coffee crops. Not all pests and diseases are related to both temperature and humidity, so calculations will be done on different (stacks of) rasters, including altitude and rainfall (linked to EcoCrop) and temperature and humidity (related to both EcoCrop and pests & diseases). The diseases that will be modelled in the studied farm are summarized in table x. Modelling pests is also possible, but is generally more difficult, as they are often less linked to certain climatic conditions than fungi and other diseases. Altitude information is more often available (e.g. Barrera n.d.), but is less useful as this is more general information. The first analysis will look at the study areas latitude and rasters of elevation, rainfall, and mean temperature during the month of August (and compare the range with all crops in the EcoCrop database to find the top 10 crops that would be suitable based on these characteristics. The second analysis will look at a list of coffee pests and diseases and some of the data that is provided about these (including altitude, temperature and humidity per hour), and will create a fuzzy overlay of all the observations during a certain period. This will be explained in the *calculations* section.

*Crop thresholds (EcoCrop data)*

List of possible crops that can be grown

*Mean data*

*Hourly data*

Risk map of pests and diseases

*Pest & Disease thresholds*

Elevation (DEM)

&

Rainfall

Elevation (DEM)

Humidity

Temperature

|  |  |  |
| --- | --- | --- |
| *Table X: List of some coffee pests and diseases and related thresholds* | | |
| Coffee Rust/  Coffee stem rust | *Hemileia vastatrix* | Rain plays a crucial role in disease development, while only few climatic thresholds are provided. These include a lower risk at temperatures <100C and > 350C (Ferreira and Boley, n.d.), which is a wider limit than the coffee crop itself. The optimal conditions for infection is between **800-900 m**, while nighttime temperatures **<150C** and **>280C** will reduce the risk of germination (Trujillo et al, 1995). |
| Coffee berry disease | *Colletotrichum kahawae* | Altitude **> 1000 meter**, **high humidity**, temperatures between **20-22 degrees** (these factors can be modelled). Shading trees can reduce the disease in two ways; by reducing the direct sunlight (Bedime et al. 2008) and by providing protection to rainfall (Mouen Bedimo 2001). This could be incorporated in more advanced models (including location of shade-trees), but will left out for this analysis. |
| Koleroga (black rot) | *Ceratobasidium noxium* | Koleroga or web blight is caused by Ceratobasidium koleroga. The pathogen is aggressive on shaded coffee during warm, humid weather (**27-300C** and **100% RH**). |
| American leaf spot | *Mycena citricolor* | American leaf spot develops best at an altitude between **1,100 and 1,550m**, and especially well on slopes with a lot of sunlight (Avelino et al. 2007). The pathogen thrives in shade-coffee plantations with temperatures **< 230C** (Trujillo et al, 1995). |
| Berry blotch / iron spot | *Mycosphaerella coffeicola* | This disease occurs throughout Central America and can cause defoliation, dropping of berries before they are mature and cause damage during the ripening process. The problem is most severe during night-time temperatures of around **250C** and high humidity (Trujillo et al, 1995). High humidity will be taken at **> 90% RH.** |
| Pink disease | *Erythricium salmonicolor* | Pink disease is a relatively minor disease, but has a global coverage. It is related to koleroga and the disease can spread to the berries and cause damage to these. The conditions under which the disease is favored is a wet climate in densely planted and shaded plantations with temperatures around **280C** (Trujillo et al, 1995). |
| Coffee berry borer | *Hypothenemus hampei* | A large number of parameters are available for the coffee berry borer from the CLIMES model, including absolute and optimal lower and upper temperature limits. For this exercise, the lower (**230C**) and upper (**300C**)temperature limits will be used. |
| Stem Borers | *Plagiohammus spp.* | Of the different species of Plagiohammus that exist, the one that has been reported in Costa Rica is P. maculosus (Bates). The problem is largest **> 1000m** (Barrera n.d.) |

*Calculations*

**Raster calculations**

Raster calculations itself is a relatively straightforward process and can include standard arithmetic operators. Calculations can be done on a single layer or a stack of layers. Raster calculations can take quite some time, so it should be important to reduce the number of cells to a number that can relate to the level at which farmers can make decisions. Providing recommendations about crops or pest risks at a level of a few square meters is of limited use in large farmers, while maps of pest risks at the level of >1 hectare might be of limited use for smallholders. In the case of the studied farm, the total number of hectares is larger than 800 hectares, so recommendations at 1 hectare would be sufficient. As the resolution of the DEM is 90m, it would be easiest to use this as the resolution for the other variables as well. This means creating a prediction grid with this resolution and – when required – changing the resolution of the rainfall raster (provided by the CGIAR-CSI) to have the same resolution as the DEM.

**Boolean logic** *(Ecocrop example)*

Boolean algebra – first introduced by George Boole in 1947, is the method in which values are set to either true (1) or false (0) and is very common in computer science and GIS analysis. Boolean logic is also used in the raster calculator of the larger GIS software such as ArcGIS and includes the main operations AND, OR and NOT. This is a relatively fast and simple approach to analyze whether a certain area in a raster meets the provided condition. There are obviously limits to this study, which will be discussed in the section on fuzzy logic, but Boolean logic can be used to analyze for the absolute limits in an area. This would be suitable to link with the Ecocrop dataset, as there are so many alternatives, that Boolean logic can be used to create a much smaller subset of crops that fall within the absolute (temperature, rainfall, latitude and altitude) limits, after which the optimal temperature limits (mean) can be linked to the mean temperature in the area to provide a further ranking of the remaining crops. The EcoCrop dataset in the {dismo} package includes the following variables that can be used: optimal and absolute temperatue, rainfall, altitude, and latitude (not all these fields are always completed). The problem is that the latter two variables are not provided in the EcoCrop dataset in {dismo}, which means a new dataset will be created which includes the top 25 crops cultivated in Costa Rica. The table (see Annex X) will include the above mentioned variables. The first analysis will be based on Boolean logic, after which the remaining crops can be analyzed in more detail (*done later*).

*Example of a Boolean analysis based on EcoCrop data*

**Fuzzy logic** *(pests & disease example)*

While there are many advantages to Boolean logic, especially in computer models, they are of limited use when explaining many real-life examples, as many crops, pests and diseases are not so exact bound to certain limits. Fuzzy logic was first introduced by Zadeh (1965) and creates a set of ‘fuzzy’ classes with different levels of membership. This will result in a certain level of reclassification based on clusters in the data. Fuzzy logic and the fuzzy modelling/overlays that can be derived from this can be used in numerous studied in the environmental sciences. Studies of fuzzy logic in GIS include studies on climate suitability (Craig et al. 1999), crop land suitability (Nisar Ahamed et al, 2000), and soil mapping (Zhu et al 2001). Some of the fuzzy expert systems that are most relevant to this research include studies on pest management (Saini et al. 2002; Mahaman et al. 2003; and Yang et al. 2003). The initial idea to this link with pests and diseases is to identify the areas where and period when the risks are highest; this could be linked to increased monitoring, of which the results could be returned to create a more detailed map of the actual state. For now, only the risks will be mapped based on some fuzzy rules. This will result in maps with values between 0 and 1 for each cluster, which can be added to get an overview of the cumulative risks of pests and diseases in the area. Areas with the most favorable conditions can be monitored at a higher frequency to avoid outbreaks of diseases by applying pesticides when the risks are largest. While fuzzy rules can be created automatically, the large number of different thresholds related to pests (absolute, optimal, minimal maximum) and the limited number of pests allow for the manual creation of fuzzy limited for the pests provided in table X. The diseases that relate to characteristics during part of the day (e.g. nighttime) will be corrected so that the value can still be 1. This table will serve as input in a complex raster calculation (in R) which is explained in the next section.

**Table x: thresholds and their weights**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Constraint** | **Factors** | **Fuzzy Class** | **Value** | **Weight** |
| *Coffee rust* | 1. < 15 and > 180C nighttime 2. Nighttime > 15 and < 180C 3. Altitude between 800-900m | Masked  1 (medium)  2 (high) | 0.00  0.50  1.00 | 30 |
| *Coffee berry disease* | 1. Altitude < 1000 meters 2. Altitude > 1000 meters 3. Temperatures between 20-220C and > 90% RH | Masked  1 (medium)  2 (high) | 0.00  0.50  1.00 | 20 |
| *Coffee berry borer* | 1. Temperatures <15 and > 320C 2. Temperatures 15-230C and 30-320C 3. Temperatures 23-300C | Masked  1 (medium)  2 (high) | 0.00  0.50  1.00 | 20 |
| *Koleroga* | 1. Temperatures <270C and >300C 2. Temperatures between 27-300C with humidity < 95% RH 3. Temperature 27-300C with > 95% RH | 1 (low)  2 (medium)  3 (high) | 0.25  0.50  1.00 | 10 |
| *American leaf spot* | 1. Temperatures > 230C 2. Temperatures < 230C and an altitude <1100 and > 1550 meters 3. Temperatures < 230C and an altitude >1100 and < 1550 meters | 1 (low)  2 (medium)  3 (high) | 0.25  0.50  1.00 | 5 |
| *Berry blotch* | 1. Nighttime temperature < 24 and > 260C and humidity < 90%RH 2. Nighttime temperature > 24 and < 260C and humidity > 90%RH | 1 (low)  2 (high) | 0.25  1.00 | 5 |
| *Stem Borers* | 1. Altitudes < 1000m 2. Altitudes > 1000m | 1 (low)  2 (high) | 0.50  1.00 | 5 |
| *Pink disease* | 1. Temperatures < 280C 2. Temperatures > 280C and RH < 90% 3. Temperatures > 280C and RH > 90% | 1 (low)  2 (medium)  3 (high) | 0.25  0.50  1.00 | 5 |

**Risk map of Aquiares**

This will be done after the interpolation is accurate and fast enough for this type of calculation.

**Discussion**

This will be done after the other sections have been finalized.

6. Conclusion & recommendations

*- Minimum sensors required*

*- Options to improve accuracy*

*- Possibilities for co-kriging*

*- Trade-off cost and accuracy*

*- Links to mobile applications*

1. Altieri, M., 2002. Agroecological Principles for Sustainable Agriculture. In *Agroecological innovations : increasing food production with participatory development*. pp. 40–46.
2. Anon, 2008. Hacienda Aquiares. Available at: http://www.cafeaquiares.com/ [Accessed October 1, 2014].
3. Ashcroft, M.B. & Gollan, J.R., 2013. Moisture, thermal inertia, and the spatial distributions of near-surface soil and air temperatures: Understanding factors that promote microrefugia. *Agricultural and Forest Meteorology*, 176, pp.77–89.
4. Asner, G.P., Scurlock, J.M.O. & Hicke, J.A., 2003. Global synthesis of leaf area index observations : *Global Ecology & Biogeography*, 12, pp.191–205.
5. Baggio, A., 2005. Wireless sensor networks in precision agriculture. *Precision Agriculture*, pp.1–22. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.46&amp;rep=rep1&amp;type=pdf.
6. Beardsley, T.M., 2014. It’s 2014... Do you know where your data are? *BioScience*, 64, p.559.
7. Blair, J.B., Rabine, D.L. & Hofton, M.A., 2006. Processing of NASA LVIS elevation and canopy (LGE, LCE and LGW) data products, version 1.01. Available at: http://lvis.gsfc.nasa.gov.
8. Blair, J.B., Rabine, D.L. & Hofton, M.A., 1999. The Laser Vegetation Imaging Sensor: A medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, pp.115–122.
9. Blumthaler, M., Ambach, W. & Ellinger, R., 1997. Increase in solar UV radiation with altitude. *Journal of Photochemistry and Photobiology B: Biology*, 39, pp.130–134.
10. Bristow, K.L. & Campbell, G.S., 1984. On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*, 31, pp.159–166.
11. Brown, N., 1993. The Implications of Climate and Gap Microclimate for Seedling Growth-Conditions in a Bornean Lowland Rain-Forest. *Journal of Tropical Ecology*, 9, pp.153–168. Available at: <Go to ISI>://1993LP12500002.
12. Budowski, G., 1980. *The Place of Agro-Forestry in Managing Tropical Forests*,
13. Buermann, W. et al., 2001. Evaluation of the utility of satellite-based vegetation leaf area index data for climate simulations. *Journal of Climate*, 14, pp.3536–3550.
14. Butler, D.R., 1977. *Coffee Leaf Temperatures in a Tropical Environment*, Koninklijke Nederlandse botanische vereniging.
15. Cavcar, M., 2000. *The International Standard Atmosphere (ISA)*,
16. CGIAR/CCAFS, 2011. Analogues online toolkit: calculate climate analogues. A glimpse of tomorrow’s climates, today. Available at: http://gismap.ciat.cgiar.org/analogues/.
17. Cornell University, I.P.-C. of A. and L.S., Aquiares Estate: Coffee & Community Since 1890. Available at: http://ip.cals.cornell.edu/undergrad/docs/study-abroad-prog-Aquiares-Estate.pdf.
18. Corripio, J.G., 2014. Package “insol.” *R package ver. 2.0–8*, p.37. Available at: http://www.meteoexploration.com/R/insol/index.html.
19. Corripio, J.G., 2003. Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain. *International Journal of Geographical Information Science*, 17, pp.1–23.
20. Daly, C. et al., 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Research*, 22, pp.99–113.
21. Dauzat, J., Rapidel, B. & Berger, A., 2001. Simulation of leaf transpiration and sap flow in virtual plants: Model description and application to a coffee plantation in Costa Rica. *Agricultural and Forest Meteorology*, 109, pp.143–160.
22. Dubayah, R. & Rich, P.M., 1995. Topographic solar-radiation models for GIS. *International Journal of Geographical Information Systems*, 9, pp.405–419. Available at: <Go to ISI>://WOS:A1995RQ49300005.
23. Easterling, W.E. et al., 1992. Preparing the erosion productivity impact calculator (EPIC) model to simulate crop response to climate change and the direct effects of CO2. *Agricultural and Forest Meteorology*, 59, pp.17–34.
24. Edwards, P.N., 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, Available at: http://www.amazon.com/Vast-Machine-Computer-Climate-Politics/dp/0262013924.
25. Efron, B. & Gong, G., 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37, pp.36–48.
26. Erickson, C.O., 1964. Meteorological Satellites. *Science (New York, N.Y.)*, 143, pp.612–613.
27. Van Etten, J., 2011. Crowdsourcing crop improvement in sub-Saharan Africa: A proposal for a scalable and inclusive approach to food security. *IDS Bulletin*, 42, pp.102–110.
28. Ewel, J. et al., 1982. Leaf area, light transmission, roots and leaf damage in nine tropical plant communities. *Agro-Ecosystems*, 7, pp.305–326.
29. FAO, 2000. Bridging the rice yield gap in the Asia-Pacific Region. *Food and Agriculture Organization of the United Nations*, p.215. Available at: http://coin.fao.org/coin-static/cms/media/9/13171760277090/2000\_16\_high.pdf.
30. FAO, 2013. EcoCrop. Available at: http://ecocrop.fao.org/ecocrop/srv/en/home.
31. Fuchs, M. et al., 1984. Leaf area estimates from measurements of photosynthetically active radiation in wheat canopies. *Agricultural and Forest Meteorology*, 32, pp.13–22.
32. Gakuru, M., Winters, K. & Stepman, F., 2009. Innovative Farmer Advisory Services Using ICT. In *IST-Africa 2009 Conference Proceedings*. pp. 1–11.
33. Goudriaan, J., 1977. *Crop micrometeorology: a simulation study*. Wageningen. Available at: library.wur.nl/WebQuery/wda/104086.
34. Govindasamy, B., Duffy, P.B. & Coquard, J., 2003. High-resolution simulations of global climate, part 2: Effects of increased greenhouse cases. *Climate Dynamics*, 21, pp.391–404.
35. Hijmans, R.J. et al., 2005. WORLDCLIM - a set of global climate layers (climate grids). *International Journal of Climatology*, 25, pp.1965–1978. Available at: http://www.worldclim.org/.
36. Hijmans, R.J. & van Etten, J., 2010. raster: Geographic analysis and modeling with raster data. *R package version*, 1, p.r948.
37. Ishida, T. & Kawashima, S., 1993. Use of cokriging to estimate surface air temperature from elevation. *Theoretical and Applied Climatology*, 47, pp.147–157.
38. Van Ittersum, M.K. et al., 2013. Yield gap analysis with local to global relevance-A review. *Field Crops Research*, 143, pp.4–17.
39. Jarvis, A. et al., 2008. Hole-filled seamless SRTM data V4. *International Centre for Tropical Agriculture (CIAT)*, p.available from http://srtm.csi.cgiar.org.
40. Jarvis, C.H. & Stuart, N., 2001. A Comparison among Strategies for Interpolating Maximum and Minimum Daily Air Temperatures. Part II: The Interaction between Number of Guiding Variables and the Type of Interpolation Method. *Journal of Applied Meteorology*, 40, pp.1075–1084.
41. Jedermann, R. et al., 2006. Applying Autonomous Sensor Systems in Logistics - Combining Sensor Networks , RFIDs and Software Agents. *Sensors and Actuators A: Physical*, 132, pp.370–375.
42. Kilibarda, M. et al., 2014. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres*, 119, pp.2294–2313.
43. Koenker, R., 2011. Quantile regression in r: a vignette. *Online*, 15, pp.1–21. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.161.133&amp;rep=rep1&amp;type=pdf>.
44. Letten, A.D. et al., 2013. The importance of temporal climate variability for spatial patterns in plant diversity. *Ecography*, 36, pp.1341–1349.
45. Li, L. & Revesz, P., 2002. A Comparison of Spatio-temporal Interpolation Methods. In *Geographic Information Science*. pp. 145–160.
46. Li, L. & Revesz, P., 2004. Interpolation methods for spatio-temporal geographic data. *Computers, Environment and Urban Systems*, 28, pp.201–227.
47. Lin, B.B., 2007. Agroforestry management as an adaptive strategy against potential microclimate extremes in coffee agriculture. *Agricultural and Forest Meteorology*, 144, pp.85–94.
48. Lobell, D.B., Cassman, K.G. & Field, C.B., 2009. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. *Annual Review of Environment and Resources*, 34, pp.179–204.
49. Mainwaring, A. et al., 2002. Wireless Sensor Networks for Habitat Monitoring. In *Proceedings of the 1st {ACM} International Workshop on Wireless Sensor Networks and Applications*. pp. 88–97. Available at: http://doi.acm.org/10.1145/570738.570751.
50. Van Marken Lichtenbelt, W.D. et al., 2006. Evaluation of wireless determination of skin temperature using iButtons. *Physiology and Behavior*, 88, pp.489–497.
51. McWilliam, A.-L.C. et al., 1993. Leaf Area Index and Above-Ground Biomass of terra firme Rain Forest and Adjacent Clearings in Amazonia. *Functional Ecology*, 7, pp.310 – 317. Available at: http://www.jstor.org/stable/2390210.
52. Mendelsohn, R. et al., 2007. Climate analysis with satellite versus weather station data. *Climatic Change*, 81, pp.71–83.
53. Merriam-Webster Dictionary, 2014. Merriam-Webster Dictionary. *Merriam-Webster Dictionary*, p.Accessed 23 Sept. 2014. Available at: http://www.merriam-webster.com/dictionary/.
54. Miller-Rushing, A., Primack, R. & Bonney, R., 2012. The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, 10, pp.285–290. Available at: <Go to ISI>://000307146600004\nhttp://www.esajournals.org/doi/pdf/10.1890/110278.
55. Mitas, L. & Mitasova, H., 1999. Spatial interpolation. In *Geographical Information Systems: Principles, Techniques, Management and Applications*. pp. 481–492. Available at: http://skagit.meas.ncsu.edu/~helena/gmslab/papers/hgint39.pdf.
56. Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, 9, pp.747–766. Available at: http://www.jstor.org/stable/2401901.
57. Morton, J.F., 2007. The impact of climate change on smallholder and subsistence agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 104, pp.19680–19685.
58. Nhemachena, C. & Hassan, R.M., 2007. Micro-Level Analysis of Farmers ’ Adaptation to Climate Change in Southern Africa. *Food Policy*, p.2. Available at: http://www.ifpri.org/publication/micro-level-analysis-farmers-adaptation-climate-change-southern-africa-0.
59. NOAA, NOAA Solar Calculator. Available at: http://www.esrl.noaa.gov/gmd/grad/solcalc/ [Accessed October 1, 2014].
60. Odeha, I.O.A., McBratney, A.B. & Chittleborough, D.J., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, 63, pp.197–214.
61. Perfecto, I., Rice, R. & Greenberg, R., 1996. Shade Coffee: A Disappearing Refuge for Biodiversity. *BioScience*, 46, pp.598– 608.
62. Perfecto, I. & Vandermeer, J., 1996. Microclimatic changes and the indirect loss of ant diversity in a tropical agroecosystem. *Oecologia*, 108, pp.577–582.
63. Petitcolin, F. & Vermote, E., 2002. Land surface reflectance, emissivity and temperature from MODIS middle and thermal infrared data. *Remote Sensing of Environment*, 83, pp.112–134.
64. Popa, M. & Iapa, C., 2011. Embedded weather station with remote wireless control. In *2011 19th Telecommunications Forum, TELFOR 2011 - Proceedings of Papers*. pp. 297–300.
65. Ramirez-Villegas, J. et al., 2011. Climate Analogues : finding tomorrow’s agriculture today. *CGIAR research program on climate change, agriculture and food security*.
66. Rérolle, V.M.C. et al., 2012. Seawater-pH measurements for ocean-acidification observations. *TrAC - Trends in Analytical Chemistry*, 40, pp.146–157.
67. Ruiz-Garcia, L. et al., 2009. A review of wireless sensor technologies and applications in agriculture and food industry: state of the art and current trends. *Sensors (Basel, Switzerland)*, 9, pp.4728–50. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3291936&tool=pmcentrez&rendertype=abstract.
68. Satyanarayana, T., Kaushik, M. & Biradar, D.P., 2011. *New approaches and tools for site-specific nutrient management with reference to potassium*, Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=caba6&AN=20113079706.
69. Schiermeier, Q., 2010. The real holes in climate science. *Nature*, 463, pp.284–287.
70. Sellers, P.J., 1985. Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, 6
71. Shah, R.C. & Rabaey, J.M., 2002. Energy Aware Routing for Low Energy Ad Hoc Sensor Networks. In *Proceedings of 10th International Conference on Communications and Networks(WCNC ’02)*. pp. 350–355.
72. Singh, S., 2012. *E-Agriculture and Rural Development*, Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84898362180&partnerID=tZOtx3y1.
73. Staver, C. et al., 2001. Designing pest-suppressive multistrata perennial crop systems: Shade-grown coffee in central america. In *Agroforestry Systems*. pp. 151–170.
74. Taugourdeau, S. et al., 2014. Leaf area index as an indicator of ecosystem services and management practices: An application for coffee agroforestry. *Agriculture, Ecosystems and Environment*, 192, pp.19–37.
75. Thomas, S.M., Heidinger, A.K. & Pavolonis, M.J., 2004. Comparison of NOAA’s operational AVHRR-derived cloud amount to other satellite-derived cloud climatologies. *Journal of Climate*, 17, pp.4805–4822.
76. Tilak, S., Abu-Ghazaleh, N.B. & Heinzelman, W., 2002. A taxonomy of wireless micro-sensor network models. *ACM SIGMOBILE Mobile Computing and Communications Review*, 6, pp.28–36.
77. Walvoort, D., Brus, D. & Gruijter, J. de, 2013. Package “spcosa.” *R*, pp.1–25. Available at: http://cran.r-project.org/web/packages/spcosa/spcosa.pdf\nhttp://www.icesi.edu.co/CRAN/web/packages/spcosa/spcosa.pdf.
78. Walvoort, D.J.J., Brus, D.J. & de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. *Computers and Geosciences*, 36, pp.1261–1267.
79. Watson, D.J., 1947. Comparative physiological studies on the growth of field crops I. Variation in Net Assimilation Rate and leaf area between species and varieties, and within and between years. *Annals of* Botany (London), 11, pp.41–76.
80. Welles, J.M. & Cohen, S., 1996. Canopy structure measurement by gap fraction analysis using commercial instrumentation. Journal of Experimental Botany, 47, pp.1335–1342. Available at: http://jxb.oxfordjournals.org/content/47/9/1335.abstract.
81. Whiteman, C.D., 2000. Mountain Meteorology: Fundamentals and Applications, Pacific Northwest National Laboratory, Richland, WA (US).
82. World Meteorological Organization, 2008. Guide of Meteorological Instruments and Methods of Observation. In *WMO*. pp. I.15–1 to I.15–11–II.2–10 to II.2–11.
83. Zeileis, A. & Grothendieck, G., 2005. zoo : An S3 Class and Methods for Indexed Totally Ordered Observations. *Wirtschaftsuniversität Wien*.