

# Feasibility of a low-cost weather sensor network for agricultural purposes

*A preliminary assessment*

Kees van Duijvendijk

---

2015

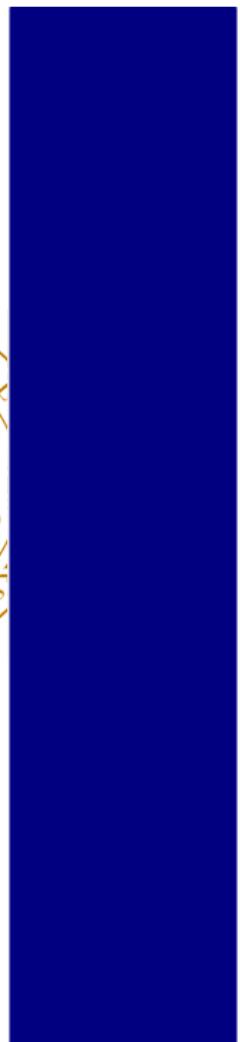
Department of  
Physical Geography and Ecosystem Science  
Centre for Geographical Information Systems  
Lund University  
Sölvegatan 12  
S-223 62 Lund  
Sweden



Seminar series nr XX



RESEARCH PROGRAM ON  
Climate Change,  
Agriculture and  
Food Security



Kees van Duijvendijk (2015). *Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment.*

Master degree thesis, 30 credits in Master in Geographical Information Sciences Department of Physical Geography and Ecosystems Science, Lund University

# **Feasibility of a low-cost weather sensor network for agricultural purposes**

*A preliminary assessment*

---

Kees van Duijvendijk

*Master degree thesis, 30 credits in Master in Geographical Information Sciences  
Department of Physical Geography and Ecosystems Science, Lund University*

**Supervisor 1:**

*Dr. Jonas Ardö*

Associate Professor

Department of Physical Geography and Ecosystem Services

Lund University

Lund, Sweden

**Supervisor 2:**

*Dr. Jacob van Etten*

Senior Scientist, Theme Leader, Adaptation to Climate Change

Bioversity International

Turrialba, Costa Rica



## i Abstract

This study has focused on challenges encountered when setting up a weather-network for agricultural purposes with only low-cost sensors and materials. The study included a set of experiments in a meteorological station and a one-month period of observations in a large coffee plantation with a complex terrain. The experiments with different sensor resolution have shown that the losses in accuracy when using a low temperature resolution are small compared to other losses, which include interpolating temperature in time. Using a high-resolution for humidity observations provides very small improvements over low-resolution, as it provides data with the same accuracy. A second set of experiments has tested different adjustment to PVC tubes, which functioned as shielding for the sensor. All adjustments provided large differences with a certified shielding for the maximum temperature on sunny days. The best coating to limit the impact of radiation was insulating foil, and it is recommended that future experiments focus more on aeration, as this has not yet provide the expected benefits. As no (combination of) adjustments provided data in line with the reference station, different types of data calibration were tested. While direct correction of data by a polynomial regression model provided reasonable results, the main difference between PVC shields and the certified sensor was caused by the faster heating/cooling over time (*thermal-inertia* properties). By creating a linear model between change in time in the PVC and certified shield, a calibration model was developed that has been used to correct data. This has been done by setting an anchor point on each day, to which the corrected change was added/subtracted. With some minor additional calibration, this model provided data that was very similar to data in the certified shield.

After the initial experiments were analysed, one hundred sensors were placed in a nearby coffee plantation with a 500 meter elevation gradient; fourteen sensors were lost and six provided incorrect data. The correlation between temperature and a range of variables was assessed. This included static (*elevation, slope, aspect, canopy height, leaf-area-index and daily radiation*) and dynamic (*hourly radiation*) covariates. On average, 52% of variance in temperature could be explained by static covariates. Including hourly radiation as covariate instead of daily radiation improved this model by 1%. Elevation is by far the most important independent variable ( $\pm 67\%$ ), although this influence is lower during periods with high temperature. A higher daily maximum temperature reduces the strength of elevation/temperature correlation. These are periods during which temperature is harder to predict based on interpolation. A lower correlation between elevation and temperature can be partly compensated by a stronger correlation with other covariates; hourly radiation contributes an average of 20% to the temperature-predicting models during hours with sun (although the models can only predict 54% of variance during these hours). Geostatistical interpolation has been tested for 80, 40, 20 and 8 sensors, with different kriging approaches and sets of covariates. Cross-validation provided the best results for universal kriging with elevation. Dynamic kriging provided smaller errors only with the full 80-sensor network. Co- and Spatio-Temporal kriging provided larger errors in predicting a left-out sensor, while data of the sensors included in the kriging showed least modification. The preferred approach depends on the objective of the network and reliability

of data.



## ii Acknowledgements

I would like to thank my supervisors, Dr. Jacob van Etten and Dr. Jonas Ardö for their support during this study and for providing valuable feedback and lessons. I am also grateful for the support provided by staff of Bioversity International, and especially to Karol Araya for bringing me to the field in terrible weather so that I could start my field work in time. I would also like to thank the people that supported the work during my stay at CATIE, especially Dr. Olivier Roupsard for introducing me to the Barquero family and providing a lot of useful data. My fieldwork would not have been possible in such a short period without the support of Don Alvaro and Alejandro Barquero, who helped in the placing of and retrieving of sensors and provided good food and coffee during the days in Aquiares. I would also like to thank the people that have contributed to the wide variety of R-packages that have been used in this study and for providing answers to my questions. Finally, I would like to thank my wife Laura, for letting me leave her for two months within one month of getting married - and for visiting me afterwards.

Funding for this research has been provided by the CGIAR Research Program on Climate Change, Agriculture and Food Security. The CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS) is a strategic partnership of CGIAR and Future Earth, led by the International Centre for Tropical Agriculture (CIAT). CCAFS brings together the world's best researchers in agricultural science, development research, climate science and Earth System science, to identify and address the most important interactions, synergies and trade-offs between climate change, agriculture and food security. <http://www.ccafs.cgiar.org>. The views expressed in this document cannot be taken to reflect the official opinions of CGIAR or Future Earth.



# Contents

i Abstract	5
ii Acknowledgements	8
iii Lists of Appendices, Tables and Figures	12
<b>1 Introduction</b>	<b>16</b>
1.1 Problem Statement . . . . .	17
1.2 Research Aim . . . . .	17
1.2.1 Aim of Section I: Data accuracy . . . . .	18
1.2.2 Aim of Section II: Spatial aspects of the network . . . . .	18
1.2.3 Aim of Section III: Interpolation strategies . . . . .	18
1.3 Research Questions . . . . .	18
1.4 Thesis Outline . . . . .	19
1.5 Workflow . . . . .	19
<b>2 Background</b>	<b>20</b>
2.1 Climate Science & Sensors . . . . .	20
2.2 Sensors and Standards . . . . .	21
2.3 Implementation of Sensor Networks in Agriculture . . . . .	21
2.4 Sensor Networks and Citizen Science . . . . .	23
2.5 Crop and Disease Thresholds . . . . .	24
<b>3 Data accuracy</b>	<b>25</b>
3.1 Introduction . . . . .	25
3.2 Materials and methodology . . . . .	26
3.2.1 Sensors . . . . .	26
3.2.2 Weather station and shield . . . . .	27
3.2.3 Alternative shields for the iButton sensors . . . . .	27
3.2.4 Experiment 1: Sensor and temporal resolution . . . . .	28
3.2.5 Experiment 2: Shields . . . . .	29
3.2.6 Experiment 3: Data calibration . . . . .	31
3.3 Results . . . . .	34
3.3.1 Experiment 1A: Sensor resolution . . . . .	34
3.3.2 Experiment 1B: Temporal resolution . . . . .	35
3.3.3 Experiment 2: Shields . . . . .	37
3.3.4 Experiment 3: Data calibration . . . . .	40
3.4 Discussion . . . . .	44
3.5 Recommendations . . . . .	47

<b>4 Spatial aspects</b>	<b>48</b>
4.1 Introduction . . . . .	48
4.2 Methodology . . . . .	49
4.2.1 Study area and sampling . . . . .	49
4.2.2 Data correction and validation . . . . .	50
4.2.3 Analysis of covariates . . . . .	52
4.2.4 Selected covariates . . . . .	53
4.2.5 Analyses . . . . .	57
4.3 Results . . . . .	58
4.3.1 Network accuracy . . . . .	58
4.3.2 Correlation between covariates . . . . .	59
4.3.3 Average correlation with covariates . . . . .	60
4.3.4 Daily trends in correlation . . . . .	61
4.4 Discussion . . . . .	65
4.5 Recommendations . . . . .	66
<b>5 Data interpolation</b>	<b>67</b>
5.1 Introduction . . . . .	67
5.2 Methodology . . . . .	68
5.2.1 Data inputs . . . . .	68
5.2.2 Prediction grid . . . . .	68
5.2.3 Geostatistical interpolation . . . . .	69
5.2.4 Selected kriging-approaches . . . . .	69
5.2.5 Analyses . . . . .	70
5.3 Results . . . . .	72
5.3.1 Ordinary Kriging . . . . .	72
5.3.2 Universal Kriging . . . . .	73
5.3.3 Dynamic universal kriging . . . . .	75
5.3.4 Co-kriging approaches . . . . .	76
5.3.5 Spatio-temporal kriging . . . . .	76
5.3.6 Accuracy in predicting the mean temperature . . . . .	78
5.4 Discussion . . . . .	78
5.5 Recommendations . . . . .	80
<b>6 Summary</b>	<b>81</b>
<b>7 Conclusions &amp; Recommendations</b>	<b>83</b>
<b>8 References</b>	<b>85</b>
<b>9 Appendices</b>	<b>92</b>

### iii Lists of Appendices, Tables and Figures

#### List of Appendices

A. Cost of materials (sensor and shield) . . . . .	92
B. Linear models of the change in temperature over time . . . . .	93
C. Humidity statistics in the different PVC shields . . . . .	94
D. Cross-validation of different calibration options . . . . .	95
E. Calibrated temperature in different quantiles (3 sensors) . . . . .	96
F. Zonal statistics for temperature at Aquiares . . . . .	97
G. Strength of temperature-explaining models and relative contribution of covariates . . . . .	98
H. Errors in Ordinary and Universal Kriging of temperature . . . . .	99
I. Errors in different Dynamic Kriging approaches . . . . .	100
J. Errors in different types of Co-Kriging . . . . .	101
K. Errors in different types of Spatio-Temporal Kriging . . . . .	102

#### List of Tables

1 Input of crowd-sourcing projects . . . . .	23
2 Output of crowd-sourcing projects . . . . .	23
3 Time covered by different sensor settings (sensors contains a Ta and RH sensor) . . . . .	26
4 Characteristics of the different coatings of the PVC shields . . . . .	31
5 Differences in temperature at different quantiles using a range interpolation approaches	36
6 Analysis of adjustments on 50mm PVC shields . . . . .	39
7 Analysis of adjustments on 25mm PVC shields . . . . .	39
8 Errors in different types of calibration (Ta) compared to data in a Stevenson shield . .	40
9 Correlation between the selected (static) covariates . . . . .	59
10 Summary of selected interpolation approaches . . . . .	71
11 RMSE (degrees Celsius) of daily mean <i>observed</i> and <i>predicted</i> mean temperature . .	78
12 Cost of different sensors (US\$) when bought in different quantities . . . . .	92
13 Cost of materials for the experiments and complete network . . . . .	92
14 Linear model for to predict change in temperature over time (1 hour) in the PVC shields	93
15 Linear model for to predict change in relative humidity over time (1 hour) . . . . .	93
16 Humidity statistics in different (50mm) sensor shields . . . . .	94
17 Humidity statistics in different (25mm) sensor shields . . . . .	94
18 Cross-validation of change-in-time calibration . . . . .	95

19	Cross-validation of the polynomial temperature calibration . . . . .	95
20	Temperature in different quantiles for a range of calibration-types . . . . .	96
21	Mean, minimum and maximum temperature at different altitude ranges in Aquiares .	97
22	Model strength and relatively contribution of covariates to explaining temperature .	98
23	Errors with Ordinary Kriging . . . . .	99
24	Errors with Universal Kriging (Elevation) . . . . .	99
25	Errors with automatized Universal Kriging (Elevation) . . . . .	99
26	Errors with basic dynamic kriging . . . . .	100
27	Errors with automatized dynamic kriging . . . . .	100
28	Errors with random forest kriging . . . . .	100
29	Errors with co-kriging (elevation only) . . . . .	101
30	Errors of co-kriging (with elevation and radiation) . . . . .	101
31	Errors with co-kriging (with elevation, radiation, slope and aspect) . . . . .	101
32	Errors with spatio-temporal ordinary kriging . . . . .	102
33	Errors with spatio-temporal universal kriging (elevation) . . . . .	102
34	Errors with spatio-temporal dynamic kriging (elevation, radiation, slope and aspect) .	102

## List of Figures

1	Steps taken in this study to reach a robust and low-cost climate network . . . . .	19
2	Top-tube lay-out of the shields . . . . .	27
3	CATIE Meteorological station with PVC shields . . . . .	30
4	High-resolution T sensor-data . . . . .	34
5	Low-resolution T sensor-data . . . . .	34
6	High-resolution RH sensor-data . . . . .	35
7	Low-resolution RH sensor-data . . . . .	35
8	Temperature in the Stevenson (blue) vs. PVC (red) shields . . . . .	37
9	Correction based on actual Ta . . . . .	39
10	Correction on change of Ta in time ( $\Delta t$ ) . . . . .	39
11	Observed temperature in three sensors around the weather station (blue = ‘cool’ sensor) . . . . .	41
12	Corrected change in time (1 hour) based on a calibration model based with 55 values	43
13	Boundaries of the Aquiares estate and placement of the recovered sensors . . . . .	50
14	Leaf Area Index in the Aquiares estate . . . . .	54
15	Canopy height in the Aquiares estate based on LiDAR data . . . . .	55
16	Daily $T_{max}$ vs RMSE . . . . .	58
17	Daily trends in average RMSE . . . . .	58
18	Average RMSE of the 80 recovered sensors in Aquiares . . . . .	59

19	RMSE vs. hourly correlation of Ta with elevation . . . . .	60
20	Relative importance of different covariates in explaining temperature . . . . .	60
21	<i>Temperature/elevation</i> correlation . . . . .	61
22	<i>Max Ta</i> vs. <i>Ta/elevation</i> correlation . . . . .	61
23	<i>Temperature/slope</i> correlation . . . . .	61
24	<i>Temperature/aspect</i> correlation . . . . .	61
25	<i>Temperature/LAI</i> correlation . . . . .	62
26	<i>Temperature/canopy</i> correlation . . . . .	62
27	<i>Temperature/radiationboxiti</i> correlation . . . . .	63
28	<i>Max Ta</i> vs. <i>mean radiation</i> correlation . . . . .	63
29	Mean hourly linear regression strength ( $R^2$ ) that explains temperature variation . . . . .	63
30	Contribution of different covariates to the best model explaining temperature variation . . . . .	64
31	Ordinary Kriging (80 sensors) . . . . .	72
32	Ordinary Kriging (40 sensors) . . . . .	72
33	Ordinary Kriging (20 sensors) . . . . .	72
34	Ordinary Kriging (8 sensors) . . . . .	72
35	Universal Kriging (80 sensors) . . . . .	73
36	Universal Kriging (40 sensors) . . . . .	73
37	Universal Kriging (20 sensors) . . . . .	73
38	Universal Kriging (8 sensors) . . . . .	73
39	Temperature at 11AM on the first day of measurements with 20 sensors (universal kriging) . . . . .	74
40	Temperature at 11AM on the first day of measurements with 20 sensors (automatized kriging) . . . . .	74
41	Predicted mean temperature based on dynamic kriging and 8 sensors . . . . .	75
42	Mean temperature with Universal Kriging (elevation) at 9AM (80 sensors) . . . . .	76
43	Mean temperature with Co-Kriging (elevation) at 9AM (80 sensors) . . . . .	76
44	Daily temperature variation based on Spatio-Temporal OK (80 sensors) . . . . .	77
45	$T_{mean}$ derived by Universal Kriging . . . . .	79
46	$T_{mean}$ derived by Spatio-Temporal UK . . . . .	79



# 1 Introduction

This thesis discusses the need for low-cost networks of climate sensors and covers several of the issues that have to be taken into account when implementing such sensor-networks. At the same time that advances in the developed world make weather predictions possible at a very high spatial/temporal scale, and models can recommend crop varieties based on trillions of observations (e.g. FieldScripts (Monsanto, 2014)), the costs for maintaining a basic network of weather stations still form a constraint for their use in many low-income countries (Barnett & Mahul, 2007). One of the larger efforts to create high-resolution global climate data is the WorldClim dataset (Hijmans et al., 2005), for which the current climate has been assessed (1950-2000) based on over 14,835 locations (depending on the units measured). This dataset, which also includes the projected climate for four representative concentration pathways (RCPs), is often linked with crop databases (e.g. EcoCrop (FAO, 2013)) to predict the impact of climate change on crops (Hijmans & Graham, 2006; Lane & Jarvis, 2007; Ramirez-Villegas, Jarvis, & Läderach, 2013). Except for downscaling climate data by statistical approaches and interpolation of weather station data (often lacking in rural areas of developing countries), an alternative approach to create climatic data at a scale that is useful for crop models, is by using stochastic weather generators (Wilks & Wilby, 1999). Downscaled climate data is used for many purposes, including short-term local weather prediction (Pielke & Wilby, 2012).

A limitation of current crop models - and the role of climate data within these - is that it is difficult to include the impact of extreme temperatures on specific phenological events (Porter & Christensen, 2013); these are the important crop phases such as leaf initiation, shoot growth, and leaf growth. Another problem with linking local climate and crops in models, is that climate variation at the local level is so high that models have to be validated by data from a number of years, resulting in smoothed datasets (Lobell, Cassman, & Field, 2009). While this is useful in global 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 impact of local climatic extremes on crop production. Climate analogues are expected to move and cause shifts in crops at the global scale (Ramirez-Villegas et al., 2011), whereas options for adaptation will generally depend on the local context (Nhemachena & Hassan, 2007). An important gap in climate science, which will be addressed in this thesis is the difficulty to predict the impact of local climatic changes (Schiermeier, 2010). In this study, local scale is defined as areas  $<1\text{km}^2$ , as data will be linked to performance of crops in field trials at farmers' fields in developing countries. There will be many different micro-climates at this scale, which are influenced by factors such as the canopy height, the openness to wind, general topography and the vicinity of larger water bodies (Ashcroft & Gollan, 2013). While many of the relevant physical processes have already been added to existing models, working at a very small scale with real-time and low-cost data remains an important challenge. Access to weather forecasts is linked to uptake of different adaptation practices (Wood et al., 2014). Taking into account the many local factors that influence climate, and that smallholders mostly cultivate small areas in risk prone areas (Morton, 2007), low cost climate-networks can provide clear benefits to smallholders.

The large number of standards that certified weather stations have to meet (e.g. World Meteorological Organization, 2008) is likely associated with higher costs. This can become a serious limitation to their creation and maintenance in developing countries. As smallholders often cultivate more complex (marginal/mountainous) terrain (Altieri, 2002), information about climatic differences at this scale will be especially relevant to them. Site-specific (climate) 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 assist in providing (top-down) information about market prices, application of agro-chemicals and weather forecasts, but can also be used in more participatory approaches, such as crowd sourcing of crop improvement (Van Etten, 2011). The hypothesis of this research is that reliable networks of climate sensors can be created with very little costs, and that these fit well in the context of climate monitoring and low-cost precision agriculture.

## 1.1 Problem Statement

While studies on the separate subjects of climate downscaling and crop thresholds have received a lot of attention, and micro-sensors are already used in a wide range of environmental studies, there is still a lack of weather stations that provide basic climatic data (temperature and humidity) with high spatial and temporal accuracy in rural areas of developing countries. Downscaled climate data can be useful in agricultural modelling and can accurately be estimated up to hourly intervals (Cheng et al., 2008). A problem, however, remains the limited understanding of impacts at the local level (Jones & Thornton, 2013). Data at this level ( $<1\text{km}^2$ ) can be used for variety recommendations and pest monitoring. One of the related data-gaps, discussed in the 5<sup>th</sup> Assessment Report of the IPCC (Chapter 7), is that most adaptation studies focus on the future climatic conditions, and less on the potential impacts in the current climate (IPCC, 2014). The difficulty to maintain networks of weather-stations in developing countries does not only limit the options to provide smallholders with insurance to weather-related risks (Barnett & Mahul, 2007), but can also limit other important aspects of climate adaptation, such as the uptake of more suitable crop-varieties (Wood et al., 2014).

## 1.2 Research Aim

The main aim of this research is to provide clear recommendations about a range of crucial issues that are faced when setting up a network for climate observations with a budget that is sufficiently low to be used in agronomic trials in developing countries, while still providing significant gains over currently available weather data. Issues include the set-up of the sensor (resolution, temporal interval), shielding of the sensor (PVC tubes with certain adjustments), data correction (different types of calibration), and spatial interpolation of data. All these issues are linked to the initial set-up of the network and aim to reduce the number of sensors that is required for accurate data, while increasing the period of data collection. The secondary aim of the study is to provide a range of uses of the created network in addition to the initial objective of linking climate data to field trials which can make the networks interesting for other extension projects. Aims of the sub-sections include:

### **1.2.1 Aim of Section I: Data accuracy**

(1) Provide information about the trade-offs between costs and accuracy when working at different sensor and temporal resolution, which should remain relevant to assess the short-term extremes. (2) Evaluate approaches of temporal-data interpolation that can be used to create a regular time-series from the irregular time-series that the sensors provide, without losing crucial information. (3) Test a range of adjustments (external coating and aeration) to PVC shields and study how the data from the sensors inside these shields relates to data to a certified (Stevenson) sensor shield.

### **1.2.2 Aim of Section II: Spatial aspects of the network**

(1) Quantify the data accuracy of the network; this includes verification with external temperature and humidity observations and with internal observations (based on a leave-one-out approach). Calculate the linear relationship between (2) several static (elevation, slope, aspect, canopy height, leaf area index) and (3) dynamic (hourly hillshade, cast shadows and maximum solar intensity) covariates and the hourly mean, minimum and maximum temperature.

### **1.2.3 Aim of Section III: Interpolation strategies**

(1) Assess different types of (geostatistical) methods for spatial interpolation of the data. Different types of Kriging (Ordinary, Universal, ‘Dynamic’ Co-Kriging, Spatial - with different covariates) will be compared regarding their (overall, spatial and temporal) accuracy when working with the full data-set and smaller sub-sections (half, one-quarter and one-tenth) of this dataset.

## **1.3 Research Questions**

**Main research question:** Can a robust and accurate weather network - that can provide temperature and relative humidity information at short (hourly) intervals at the plot level (100 meter resolution) - be created with only low costs sensors and materials?

**Data accuracy research question:** What is the optimal sensor and temporal resolution that allows to cover the longest observation period in the field, and which type of low-cost sensor shield would be most suitable to hold these sensors (limit size of corrections)?

**Spatial aspects research question:** Which of the different (static and dynamic) covariates have a strong linear relationship with hourly temperature throughout the study area; at which periods (hours and days) are these relationships strongest?

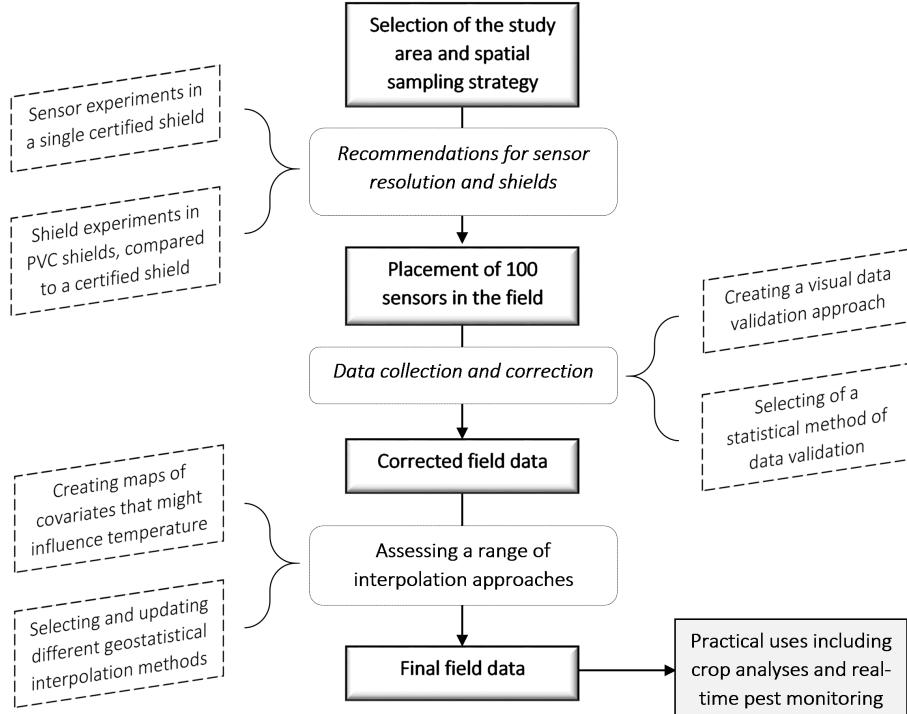
**Spatial interpolation question:** What are the differences between a selection of kriging strategies regarding the overall prediction accuracy and errors in space and time, and what is the kriging approach that provides the best output with the smaller sets of sensors?

## 1.4 Thesis Outline

The thesis is divided into six chapters. The first chapter introduces the research and the second chapter discusses the wider context of weather networks, crop thresholds and crowd-sourcing, including history, limitations and possible uses. The third chapter covers a range of experiments focused at the sensor resolution, shielding, and data calibration. The fourth chapter explains the initial network accuracy and strength and trends of correlation with a set of variables. The fifth chapter focusses on different geostatistical approaches to create a geo-network from the sensor-data, and the sixth and seventh sections provide the summary, conclusions and recommendations of this research. Sections three, four and five all have separate sections for introduction, methodology, results, discussion and recommendations. This should make these sections - in combination with the background - understandable as stand-alone documents.

## 1.5 Workflow

The steps that have been taken during this thesis are provided in *figure 1*, with the main thought-processes at the sides. All scripts are provided at <https://github.com/cornelisvd/thesis>. The thought processes required with each practical step, from selection of the study area to practical implementation, form the basis of the structure of this thesis. Based on the raw data, all steps will be easy to reproduce, which make these steps relevant to other projects using the same sensors. The subsequent practical uses after creation of final field data will be discussed, but not implemented, in this study.



**Figure 1:** Steps taken in this study to reach a robust and low-cost climate network

## 2 Background

### 2.1 Climate Science & Sensors

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) was proposed for real-time data sharing (by telegram in that time) was as early as 1900 (Edwards, 2010), but only once the computer arrived, different formats could easily be merged to use in forecasting. The first weather 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 agricultural science, the main difference between satellites and weather stations is that weather stations provide accurate - but sporadic - data, while for satellites it is hard to measure near-surface phenomena (e.g. precipitation), although the spatial coverage is generally more complete (Mendelsohn et al., 2007). The difficulties with interpolating weather station-data in areas with limited coverage, in combination with the critique they often receive based on their exposure (e.g. Menne, Williams, & Palecki, 2010), has created the need for additional networks of connected weather stations. Examples of weather networks to which users can attach their personal weather station (PWS) include Weather Underground (Geller, 2007) and WeatherLink (DAVIS, 2014b). Despite the fact these networks generally demand certain standards of the stations, basic shields with micro-sensors could increase spatial coverage in areas where such networks still have limited coverage.

The decreasing cost, together with the increasingly small size of sensors and advancements in technology related to data storage and transmission, are making the use of micro-sensors interesting for a wide range of purposes in the agricultural sector (Ruiz-Garcia et al., 2009). The initial objective of many of these sensors is generally related to commercial and closed-environment observations, such as logistics (e.g. Jedermann et al., 2006). Micro-sensors have already been used in clinical studies (Van Marken Lichtenbelt et al., 2006), environmental studies - including habitat monitoring - (Mainwaring et al., 2002) and measuring ocean acidification (Rérolle et al., 2012). Weather observations can be linked to many local-scale recommendations related to variety recommendations and pest management; this includes linkages with IPM modules (e.g. DAVIS, 2014a), and risk-assessment of phytophtora in potato (Baggio, 2005). Another use of high-resolution temperature and humidity data is to analyse the risk of (future) pests during different phenological phases. Different pest-models are discussed in Gaur and Sharma (2014). While basic temperature and humidity data can be used to calculate other relevant issues such as Vapour Pressure Deficit and radiation estimates (*section 2.5 on crop thresholds*), additional information that is needed to provide detailed pest/disease risk and variety recommendations will include soil maps (including soil type, depth, pH), and the geographic distribution of pests/diseases. The current limitations to the creation of networks and related software for precision agriculture in developed countries are expected to include several issues. This includes the costs for the physical network (shields, sensors), and costs of maintaining these. Another constraint to the creation of crop-models is that the diversity of cultivated crops/varieties (and thus required models) is expected to be higher in developing countries than in the developed world.

## 2.2 Sensors and Standards

While climate-monitoring networks can function with many different sensor-types, an important issue to make the provided data robust is not only the sensor itself, but also the shield in which the sensor is placed. The actual *sensors* can be very cheap<sup>1</sup>, while the larger *sensor-housing* in which they are placed adds important features such as memory, connectivity and sometimes robustness to shocks and water damage. The sensor housing - in case of the cheaper sensors - does not add active ventilation or protection from other issues such as solar radiation. This, together with the standards that are set by the WMO (*box 1*), creates the need for a robust shield in which the sensor will be placed (*sensor-shield*). Some of the WMO recommendations are easy to address in a low-cost sensor-shield (including height, inspection, comparison, coordinates, and meta-data), while others are obviously 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. 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, are discussed in the chapter on the *data accuracy*; a number of experiments are analysed that have used different shield constructions with basic materials that can be found in almost regions.

## 2.3 Implementation of Sensor Networks in Agriculture

Real-time climate monitoring and micro-sensors can provide useful information throughout the food sector. As a separate element, micro-sensors can be used to track commodities throughout the value chain, to assure it reaches the right people; sensor networks are, for example, promoted for use in integrated supply chains to get food to the poor in India (Viswanadham, n.d.). Climate monitoring, without using micro-sensors, can be linked to climate risk management (CRM) at a larger scale (an overview of CRM in agriculture is discussed in Selvaraju (2012)), although understanding local differences requires more detailed information. Similarly, small scale yield prediction and crop insurance can be based on satellite monitoring (Seelan et al., 2003; Singh et al., 2002), since satellites can detect vegetation changes (e.g. Vegetation Health Products - NOAA/NESDIS) and provide information at a level which is interesting for crop insurers. Sensors are also not absolutely necessary to study the local microclimates, as issues such as elevation (Charney, 1949), canopy height (Dear-dorff, 1978) and leaf area index (Goudriaan, 1977) can be - and have traditionally been - linked to differences in local climate without having information about the actual atmospheric temperature.

The expected impact of canopy on understory climate (decrease of  $T_{max}$  and increase in  $RH_{max}$ ) has been confirmed by Von Arx, Dobbertin & Rebetez (2012) based on long-term meteorological data. Although the relationship between temperature and humidity is well studied, and can be

---

<sup>1</sup>Online markets (e.g Alibaba.com and Ebay.com) have (temperature and RH) sensors available for less than US\$ 1.

converted based on simple formulae (Lawrence, 2005), humidity indicators are less commonly used in crop models than indicators based on temperature. Humidity data and thresholds are interesting for prediction of plant diseases (Wilks & Shen, 1991), and for this reason could provide a valuable source of information to farmers as addition to Integrated Pest Management programs. Low-cost climate networks can be linked to varietal trials and risk assessment, but research related to biodiversity can also benefit from this, by enabling to better study the relationship between local temperature and plant diversity (Letten et al., 2013) and to locate microrefugia (Ashcroft & Gollan, 2013).

**Box1: 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 2.0m 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 - 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 and 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 (meta-data).

## 2.4 Sensor Networks and Citizen Science

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 (e-Agriculture, 2014). Although e-agriculture applications provide an interesting prospect to generate cost-effective solutions to information delivery, it will be relatively difficult to include farmer participation and avoid the risk of being limited to a top-down approach. One way to include participation in e-agriculture, linked to sensor networks, is by applying a crowd-sourcing approach. Crowd-sourcing is an approach in which tasks are outsourced to (paid or unpaid) volunteers, and has a long history in ecological research (Miller-Rushing, Primack, & Bonney, 2012).

Bioversity International, an international agricultural research institute member of the CGIAR Consortium, has developed an approach in which large numbers of farmers can test different modern and traditional 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. Advanced 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). Involvement of farmers in these projects is not necessarily limited to providing feedback, but could also include the creation of the sensor shields (or more advanced climate-monitoring devices) by farmers (*tables 1 & 2*). After testing whether farmers can create the sensor shielding, other - more complex - instruments could possibly also be created by them. This has the potential to provide a win-win situation by providing off-farm income to farmers and by creating a cost-effective way of increasing the possibility for researchers to make climate measurements in rural areas.

<b>Input:</b>	<b>Output:</b>
<ul style="list-style-type: none"> <li>- Micro-sensors</li> <li>- Shield material</li> <li>- Crop varieties</li> <li>- Mobile phones</li> <li>- Charging station</li> <li>- Mobile application</li> <li>- Financial support*</li> </ul> <p>* If farmers create the PVC shields/ other required constructions.</p>	<p><i>Scientific:</i></p> <ul style="list-style-type: none"> <li>- Varietal trials that can be linked to climate</li> <li>- High-resolution temperature and humidity data</li> <li>- More climate observations in rural areas</li> <li>- Potential to study issues related to microclimates</li> </ul> <p><i>Practical:</i></p> <ul style="list-style-type: none"> <li>- Weather data that can be used in forecasting</li> <li>- Data that can be linked to pest/disease risks</li> <li>- Participation of - and income for - smallholders</li> </ul>

**Table 1:** Input of crowd-sourcing projects

**Table 2:** Output of crowd-sourcing projects

## 2.5 Crop and Disease Thresholds

Thresholds, related to temperature and in lesser extent humidity, are available for a large number of crops. While these are often studied in detail for specific (generally staple) crops, models can be used to assess the impact of several thresholds on a wide range of crops. One of the first models to do this was the PLANTGRO model (Hackett, 1991), while one of the best known records - based on the same concepts - is the EcoCrop database. A well-documented threshold for crops is a daytime temperature of 30°C, while warm nights are also projected to have a negative impact on most crops (IPCC, 2014). The EcoCrop database does not include humidity, while elevation - not having an impact outside of its relationship to temperature - is also excluded in some versions (e.g. the R *dismo* package). There are, however, still *absolute* and *optimal* minimum and maximum temperatures. The impacts will vary for the different phenological phases (Menzel et al., 2006). Studies that include this level of detail have only been made for few crops, such as maize, rice and wheat (Ewert & Pleijel, 1999; Sánchez, Rasmussen, & Porter, 2014). A model that includes the impact of thresholds during specific crop phases is the DSSAT model (Jones et al., 1998), but this model currently includes only 28 crops. The difficulty to model non-linear responses of crops to climate change is expected to result in a underestimation of the overall impact (Porter & Christensen, 2013). Not only does climate change impact on crops in different ways that are hard to accurately model, climate change will also have an impact on plant pests and diseases (Chakraborty, Tiedemann, & Teng, 2000; Rosenzweig et al., 2001). While combining all the different impacts into one model will be very complicated when also considering different crop varieties, it can be expected that - when adjusting the agronomy to minimize risks - certain basic thresholds will be valid for a wide range of crops.

The network that has been created in this projected is not intended to be linked to performance of crops in different phonological phases, but aims to create (with low costs) as many different indicators of crop performance and risk of plant diseases, as possible. Pest and diseases can be linked to temperature and humidity data, while crop performance also requires information related to additional factors, such as soil type, water availability and solar radiation. Information about the soil (including pH, depth, permeability, and water content) is a completely different aspect, which would require at least a one-time soil sampling and continuous soil moisture measurements to include in models. The other factors can - at least partly - be derived from basic temperature and humidity data. With few additional parameters, temperature and humidity data can be related to solar irradiation ( $Rs$ ) (Winslow, Hunt, & Piper, 2001), net evapotranspiration ( $ET$ ) (Allen et al., 1998), and vapour pressure deficit ( $VPD$ ) (Wang et al., 2004). Most of these indicators only require the maximum and/or minimum temperature and humidity as input, which means observations every three hours (assumed to give values relatively close to this) can be sufficient for basic studies of crop performance. While this would also be sufficient to calculate the mean temperature (the WMO standard formula for mean air temperature is provided below), more frequent observations will provide better information of daily extremes and the period above and/or below a certain crop/disease related threshold.

$$T_{mean} = \frac{(T_{min} + T_{max})}{2}$$

### 3 Data accuracy

#### 3.1 Introduction

The objective of this study is to create a low-cost and robust network of climate-sensors. For this reason, it will be important to find a good balance between costs (both for materials as labour for maintenance) and data accuracy. The sensors have to provide high-resolution information about the local climate in the area (interpolated to a geo-network), including daily and seasonal extremes. The number of sensors that is needed for a particular network has a large influence on the total cost. The used sensors have a resolution that provides output that is better than  $0.5^{\circ}\text{C}$  and 5% RH for a large temperature range. This resolution in itself is sufficient to link to most crop and pest/disease thresholds, as explained in *section 2.5*. As sensors cannot be placed everywhere, spatial interpolation is required to estimate climatic data for the locations between the sensors. This will decrease the accuracy of the data. The spatial accuracy of the network has to correspond to the spatial units about which farmers make decisions and where they conduct their experiments. Fields of smallholders are a fraction of a hectare in Central America, where this study was conducted, but accurate data at one hectare ( $100 \times 100$  m) resolution is considered good enough in this project. Another factor that influences the total cost of running a sensor network, is the labour required to collect and process the data. The sensors used in this study do not allow for remote retrieval of the data. A straightforward solution to increase the period that the sensors can be used in the field, is to adjust the resolution at which the sensors make their observations. Another option that is considered is whether it will be useful to include humidity measurements, as these will have some correlation with temperature and can be calculated with a few ( $T + RH$ ) sensors and some basic formulae (Lawrence, 2005).

Shields are also a major cost factor. Sensors are placed inside shields that should ensure that the measured temperature and humidity will be comparable to that in (WMO) certified sensor shields. In case this is not possible, some form of data regression can be done, although this is not preferred. Several studies that have focused on low-cost sensor shielding, and have provided interesting recommendations for this research: Thomas and Smoot (2013) have used PVC tubes, but also included an aspirator unit in the shield. Tarara and Hoheisel (2007) warned for large errors when using open-bottomed shields. Holden et al. (2013) transferred the shields that are normally used beneath a forest canopy to the field by mounting them to poles, and Clark et al. (2006) have created a passively ventilated (inverted-U shaped) shield with PVC pipe. A general conclusion from these studies is that including (low-cost) active aeration and working beneath a canopy will improve the results, as this can reduce the impact of midday radiation. The ISO Standard 17714 (ISO, 2007) '*Air temperature measurements: Test methods for comparing the performance of thermometer shields/screens and defining important characteristics*', is also considered in this experiment. Another issue that will be taken into account is that certain shields have previously been recommended and used by Bioversity International in different projects (Mittra, Van Etten, & Franco, 2013), based on a study in the Alps by Hellström and Mark (2006). The main differences with this study is that it was conducted at higher elevation and thus measured a lower ( $2\text{--}16^{\circ}\text{C}$ ) temperature range, and

used slightly smaller PVC tubes ( $\frac{3}{4}$  inch) compared to the experiments in this study (1 and 2 inch diameter, which is 25 and 50mm). As a first approximation, this study limits itself to a comparison of shields that have been used in previous studies and does not test new *coatings*. Recommendations are made regarding sensor resolution and shielding, as well as possible calibration techniques.

## 3.2 Materials and methodology

### 3.2.1 Sensors

The sensor that has been used in this research is the iButton Hygrochron (DS1923). This sensor was selected because it can measure both temperature and humidity at both low and high resolution. This can help to assess whether high-resolution, as well as humidity, data is required. This sensor brand is also used in similar studies, which have provided good results. This sensor costs  $\pm$  US\$ 80 (*Appendix A*). The cheapest iButton alternative ( $\pm$  US\$ 20) is the DS1921G-F5, which can only measure temperature at a low resolution and has one-quarter of the memory size (Maxim Integrated, n.d.). According to the specifications, the temperature accuracy of the DS1923 is better than  $0.5^{\circ}\text{C}$  for most of the range ( $-10^{\circ}\text{C}$  to  $+65^{\circ}\text{C}$ ), while the accuracy of the humidity measurements is 5% RH. The DS1923 can store the temperature values with 8- or 11-bit resolution ( $0.5^{\circ}\text{C}$  or  $0.0625^{\circ}\text{C}$ ), and humidity with 8- or 12-bit resolution (0.6% or 0.04% RH). The total memory for this sensor (current generation) is 8,192 bytes. By using the lower resolution sensor for both temperature and humidity, the sensor can store double the number of observation at the same time interval (*table 3*). The three options that are most interesting to use in combination with agronomic trials are highlighted; these 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 still be sufficient for many of the important staple crops around the world (at least the modern varieties), including rice, maize and wheat.

	Observations (n)	2-Hour interval	1-Hour interval	30-Minute interval	15-Minute interval
<i>One sensor at low-resolution</i>	8,192	683 days	341 days	171 days	85 days
<i>One sensor at high-resolution / two sensors at low-resolution</i>	4,096	341 days	171 days	85 days	43 days
<i>One sensor at low-resolution &amp; one sensor at high-resolution</i>	2,560	213 days	107 days	53 days	27 days
<i>Two sensors at high-resolution</i>	2,048	171 days	85 days	43 days	21 days

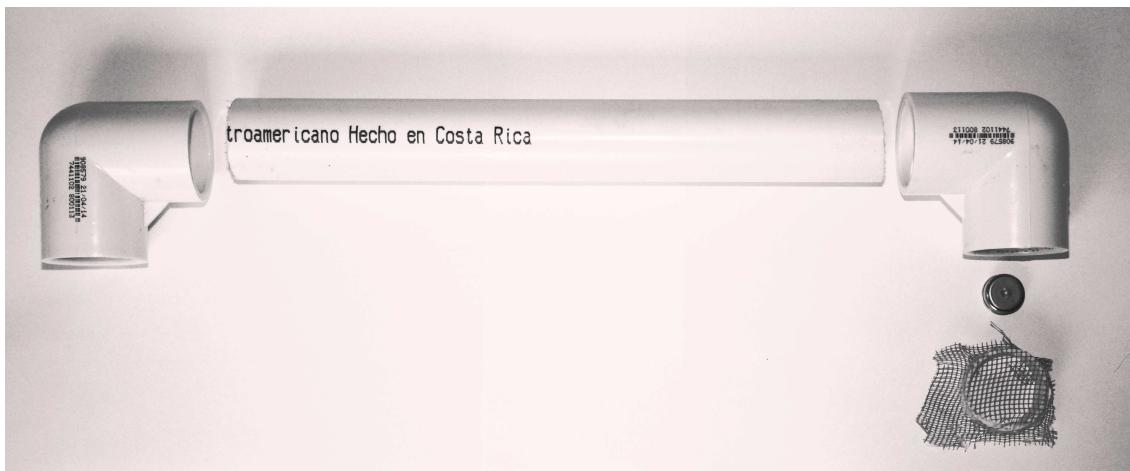
**Table 3:** Time covered by different sensor settings (sensors contains a Ta and RH sensor)

### 3.2.2 Weather station and shield

CATIE (*Centro Agronómico Tropical de Investigación y Enseñanza*) is an agricultural research centre with its headquarters in Turrialba, Costa Rica (Latitude:  $9^{\circ} 53'$ , Longitude:  $-83^{\circ} 38'$ ). The centre has a meteorological station, which has been making measurements since 1942. The station includes two (WMO certified) Stevenson shields, which both house a long-term (analogue) observation instrument. Data derived from sensors in this shield was used to compare with the data derived from sensors in PVC shields to assess the accuracy of alternative PVC-shields. Also, our data can be compared with historical measurements. Additional information - except for temperature and humidity - that this station provides, includes precipitation, solar radiation, solar intensity, evapotranspiration and wind speed. This data will especially be important to compare with field data.

### 3.2.3 Alternative shields for the iButton sensors

In this study, the basic (control) shield will be based on the Bioversity International manual '*Collecting Weather Data in the Field with High Spatial and Temporal Resolution Using iButtons*' (Mittra, Van Etten, & Franco, 2013). The PVC tube that has been selected is  $\pm 1.5$  meter long to be in line with WHO recommendations for weather instruments. It is white to reflect radiation, thin to reduce costs, and can include holes for additional ventilation. The sensor is placed on a fibre-glass (mosquito) mesh at the end of the tube with an open bottom and is partly exposed to reflecting radiation (figure 2). The study on which this construction was based has been conducted at high-elevation, which means the shields were not calibrated to  $>20^{\circ}\text{C}$ . Bioversity International projects that have used similar shields in other regions, provided temperatures that were not in line with what could be expected in these regions. This indicates that the construction needs adjustment, or the data requires calibration. The selected shield-design and exterior 'coating' has been based on the research mentioned in the introduction. Bioversity International has worked with a similar shield-design in other projects, so adjustments to this basic shield was prioritized in this project.



**Figure 2:** Top-tube lay-out of the shields

The reason to use PVC pipes is that these shields are to be used around field crops, where sensors cannot be hung from trees, as is normally done in (agro-) forestry studies where these networks have a longer history. A self-sustained structure is needed, and PVC pipes are a universally available material, so that the shield can be constructed in any part of the world using exactly the same design. All tubes in the experiments were thin (25 to 50mm), white, PVC tubes existing of several parts: 1) a 1.75m main tube (25cm to anchor); 2) a 90<sup>0</sup> PVC elbow attached to the main tube; 3) a small (20cm) PVC 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 is placed. The sensor was placed on a fibre-glass mesh that is located between the elbow and the final tube. The position of the sensor and material of this mesh might influence the measurements, but alternatives were not tested. The PVC was sold per 6 meters (which might depend on the region). Three 1.75m tubes could be created - of which 25cm has been used to firmly place the shield in the soil - with the remainders used for other parts.

### 3.2.4 Experiment 1: Sensor and temporal resolution

#### *Objective*

The aim of *Experiment 1* was to analyse the influence of sensor and temporal resolution and interval scale on the overall accuracy and fitness-for-use of the resulting information. This experiment serves to analyse the trade-off between the time the sensor can be left in the field and the accuracy of the resulting information. While a higher precision can be useful for some purposes, the data for the intended uses of this network do not have to be exact to 0.1<sup>0</sup>C; around 0.5<sup>0</sup>C will be considered satisfactory, as most of the relevant thresholds are provided at a slightly lower resolution (1<sup>0</sup>C: *section 2.5 on thresholds*). A further reduction in accuracy is expected to result from spatial interpolation.

#### *Set up*

Experiment 1 was conducted with a total of 30 sensors, making observations at 5-minute interval during one week in July 2014 at the CATIE meteorological station (observations started at noon). The sensors were placed inside the Stevenson shield to ensure the conditions are very similar for all sensors. This shield, located at ± 1.5 m above the surface, provides protection against precipitation and radiation, while it still allows for sufficient (passive) ventilation inside the shield. For the first experiment, 15 sensors were set at the highest temperature (0.0625<sup>0</sup>C) and humidity resolution (0.04% RH), while 15 sensors were set to measure at low-resolution (0.5<sup>0</sup>C and 0.6% RH, respectively).

#### *Analysis*

The high- and low resolution sensors were placed together throughout the shield to limit the possible impact of differences inside the shield, and linear interpolation was used to attain values at 1-minute intervals. Basic statistics (mean, range, and standard deviation) were subsequently calculated over the resulting dataset (30 \*.csv files with >10.000 observations each). A second analysis that was based on the same dataset of this first experiment focused on the different statistics that resulted from working at different temporal intervals. The number of observations was reduced to hourly or

2-hourly intervals by selecting (*a*) the closest observation to the full hour in the original 5-minute dataset, and (*b*) a random observation in the same time-window. The software that is provided with the iButtons does not enable the user to select an exact start-moment, so it was analysed whether this would make a difference on the accuracy. The user can select a certain delay before measurement starts, but when launching many buttons, this might become complicated and possibly costly when taking into account the time required. A third analysis of *experiment 1* used different interpolation techniques to transform the data to a regular time series. This includes using *linear* and *spline* interpolation approaches, which are available in the *zoo* R-package (Zeileis & Grothendieck, 2005).

Temperature differences in different quantiles were calculated for the separate days, and compared to the same data for the original dataset. Humidity is not analysed separately in this temporal-resolution experiment, as shielding is the same, and the relationship between temperature and humidity will be the same as in the sensor-resolution experiments. The average of the daily differences has been calculated and provided in a table for comparison. Differences between (*a*) the temperature in the original dataset and (*b*) the data that resulted from the adjusted intervals and interpolated data have been categorized into different categories:  $<0.1^{\circ}\text{C}$  is insignificant,  $0.1\text{-}0.25^{\circ}\text{C}$  is *small*,  $0.25\text{-}0.5^{\circ}\text{C}$  is *medium*,  $0.5\text{-}1^{\circ}\text{C}$  differences can be considered *high*, and  $>1^{\circ}\text{C}$  is *problematic*.

### 3.2.5 Experiment 2: Shields

#### *Objective*

The aim of the second experiment was to test different low-cost adjustments to the control PVC tubes, to see if they resulted in robust output. Adjustments were mainly focused at improving the aeration around the sensor, and increase the reflectivity of the shield to avoid high temperature peaks during the periods with strongest solar radiation. No special tools (only tape, scissors, and a drill) were used for assembly. Comparing the data from different adjustments with the Stevenson shield provided some insight in the performance of the shields and possibilities for further calibration. The availability of several decades of climate observations was also used to place the data into context.

#### *Set up*

The second experiment (sensor shields) has been conducted during July 2014 at the meteorological station at CATIE (*figure 3*). As only one month was available, and different types of day are required to provide adequate context, the sensors have been measuring for a long period and were only removed for reading after a day was found to be sufficiently useful to provide new information. Two PVC-diameters (25 and 50mm) were tested. The first set of experiments on both a sunny and rainy day was conducted with the wider (50mm) tubes, after which the cheaper and lighter 25mm tubes were also tested on a sunny and cloudy/rainy day with most of the same adjustments, to analyse whether they had the same impact. While perforation was based on drilling holes, the materials for coating have been selected based on their cost and likely global availability. The materials have included reflective (red/white) tape, gutter tape, 6mm thermal insulation foil, fibre glass mosquito mesh, and different types of white paint. Galvanized metal sheets and metal mesh were also bought, but were



**Figure 3:** CATIE Meteorological station with PVC shields

found too difficult to cut/attach to the shields. The costs of all purchases are provided in *Appendix A*. The coatings and holes were selected to provide the same functions as can be found in the certified Stevenson shield (*table 4*). As the time for this experiment was limited, and different types of day (sunny, cloudy/rainy) had to be analysed to assess the performance of the shields, some aspects could not be studied during this research. This includes placement of the sensor at different locations in the shield (now relatively exposed to reflective radiation) and the inverted-U shaped shield that has been recommended by Clark et al. (2006). As the shields were expected to provide incorrect data under periods of strong solar radiation, different approaches to data regression were also studied.

#### *Analysis*

As, based on existing literature (*section 3.1: introduction*), shields were expected to have an exaggeration of the midday temperature in low-cost (PVC) shielding, different approaches were tested to reduce these peaks. Smoothing is expected to reduce the peaks in the dataset, but these techniques (e.g. smoothing splines, LOWESS, and Turkeys running median) are symmetrical approaches and hence reduce the range of the lower temperatures as well, which are relatively similar to the data in certified shields. While settings were adjusted to match the maximum temperatures in the Stevenson shield, the resulting minimum temperature was a lot higher as well, which gives an incorrect

**Table 4:** Characteristics of the different coatings of the PVC shields

Shielding	Description
<i>Stevenson shield</i>	The reference station is a Stevenson shield that has been used for temperature and humidity measurements at CATIE for several decades. The shield is placed at 1.5 m above the surface and contains an analogue temperature/humidity logger.
<i>White paint</i>	Paint has been applied to the top-tube of the shield, as well as the elbows. Spray paint (Bosny Blanco Brillante) was used, but this did not provide a reflective layer on the tube and could be considered a matt paint layer with limited reflectivity.
<i>Perforations</i>	Perforations have been made by an electric hand-drill and the size of the holes was different for the larger and smaller tubes. The exact size of the holes is not known, but it was assessed (visually) that approximately 30% of the top-tube was drilled out.
<i>Insulating foil</i>	The top-tube, as well as an extra 10cm on both sides of this tube, was covered by a 6mm insulating foil with a white and silver side. The foil, with the silver side facing outwards, was tightly tied to the top tube at two places with thin adhesive tape.
<i>Insulating foil + perforations</i>	This combination included the 30% holes, and the same insulating foil as above. The foil was more loosely attached to the tube, the holes at the bottom were still completely open. Holes at other locations might have had restricted air flow.
<i>Reflective tape</i>	Two types of tape have been used. The first (red/white reflective) is normally attached to trucks to increase visibility, while the other was a silver-coloured tar-based tape that can be used to close holes in roofs. This tape was very thick.

picture of the local extremes. For this reason, smoothing was not further attempted in this study. Two approaches that were tested are polynomial and quantile regression. An important advantage of quantile regression is that it can provide more robust data with less impact of outliers (Yu, Lu, & Stander, 2003). Studying the required data-correction did not only focus on the actual temperature/humidity in the reference shield versus the temperature/humidity in the test shields, but also analysed whether differences can be explained by looking at the rate of change over time ( $\Delta t$  in Stevenson vs. PVC shield). The latter will be more difficult to apply in a correction, as there is a risk of cumulative errors - and it will have to be done for all separate days (using an anchor-hour). An algorithm was tested (*experiment 3*) to assess whether correction based on change over time provides good results. Both temperature and humidity were analysed, as shield-adjustments were expected to have a different impact on both types of data, due to differences in thermal inertia.

### 3.2.6 Experiment 3: Data calibration

#### *Objective*

After the sensor and temporal resolution and impact of different shields were analysed, different types of data calibration were done to find models that can best predict the actual (Stevenson) air temperature based on the temperature in the PVC shields. This was only done for temperature, after which the previously found correlation between temperature and relative humidity assessed in the Stevenson was used to predict the relative humidity as well. The objective was to find a model

that resulted in the least errors for all hours in the day, which could also be applied to the field data.

#### *Set-up*

The set-up of this experiment is partly the same as in *experiment 2*, with the five different days that were analysed in different PVC pipes. The data that has been calibrated is the ‘control’ data versus the Stevenson data, as the control dataset contains the most observations for calibration and the ‘control’ shields have already been used for experiments in different regions as well. Both sample and control dataset contain 7,200 minute observations, and 120 hourly observations, which have been analysed. In addition to this, three sensors were placed close to a weather station in the main study area (Aquiares), which helped to assess if the calibration worked in different regions. These sensors measured at a 20 minute resolution to find out how calibration performed over a longer period.

#### *Analysis*

Two main approaches were tested in this experiment: (1) calibration based on the actual measured temperature in the PVC shields versus the Stevenson; and (2) calibration based on the hourly change in temperature in the different shields. The first approach has applied a polynomial model (1<sup>st</sup> to 3<sup>rd</sup> order) to fit the PVC data to the Stevenson data during the five days of experiments at the CATIE meteorological station. This calibration used the temperature at the minute (full days), which means all (2 times) 7,200 observations have been used to create the polynomial model. For each day the errors at every hour have been squared and subsequently the mean of the square root of these errors has been calculated. The mean of these hourly square-root errors has been compared between the different polynomial and change-in-time calibration models to find the model with the lowest errors.

In order to create the model based on change in time, first two anchor points have been created based on the five days with observations every 5-minutes. The smallest absolute difference in temperature between the PVC tube and the Stevenson was calculated within two hours of the time of sunrise ( $\pm 5:30\text{ AM}$  in the month of July). This is expected to be the moment from which differences between the PVC and Stevenson can become problematic due to the impact of solar radiation. The median of the moments with lowest difference has been selected as the anchor point, and this moment (at a given minute) has been rounded to the closest hour. The same has been done for the evening: the lowest difference within two hours of sunset was calculated and rounded to the closest hour. This has created a certain time-window that could be calibrated, while the rest of the data was not adjusted. For all five days, the calibrated data has been created by creating a vector with the morning, after which several hours of calibrated data were calculated, followed by the data for the evening. Data for the morning and evening has also been calibrated by testing the impact of different adjustments ( $\pm 0\text{--}1^{\circ}\text{C}$ ) of the hourly temperature. The main calibration (during hours with sun) was based on the change in time, for which a simple linear model was created.

The linear model for the change in time was based on hourly observations during the period that had to be calibrated (five times 12 hours). The differences over time during this period could be calculated for the 11 hours in between these hours. The difference in time in the PVC shields were

predicted by a linear model, creating different values, which were cumulatively added to the anchor point. New values are thus created for a number of hours every day, based on a linear model of the change in time. Several parameters in this calibration have been tested. The first is a certain multiplication of the change in time that resulted from the prediction model. Values between 0 and 2 (steps of 0.01) have been tested to see which value provides the lowest errors with the control (Stevenson) data. The other two parameters that have been tested are the addition/subtraction of a certain temperature in both the hours before the first anchor-point and after the second anchor-point (after the multiplication of change in time was tested). Values between -1 and 1 were tested in steps of 0.01 to find the closest match to the control data. This tuning has resulted in three parameters that were added to the calibration model. The original root-mean-square errors (RMSE) against which the data is compared has a value of  $0.66^{\circ}\text{C}$  during the five days, although the errors are significantly larger during the day (7am - 5pm:  $1.21^{\circ}\text{C}$ ), compared to the night ( $0.27^{\circ}\text{C}$ ). Model validation has been done by leave-one-out cross-validation, in which the model and parameters has been calibrated based on  $n-1$  days, after which the RMSE for the left-out day was measured. The results have been used to create a preliminary model based on the physical properties of the PVC shields.

This analysis provided the RMSE for the two types of calibration based on five days of testing in unshaded (not necessarily sunny) positions at a meteorological station. Field-testing, however, was done in an area with more shading and a more complex terrain. For this reason, different polynomial models and calibration based on change in time, were also tested for the field data to see which provides the best match to a local weather station. Of the hundred sensors that were placed at the Aquiares coffee plantation (*discussed in more detail in the next chapter*), three were placed close to a weather station, which contains a *Campbell* sensor placed at 26 meters above the surface. Data for the studied period is available at 30-minute interval, and the hourly values will be compared with the calibrated hourly values of the three closest sensors, to see which adjustment fits best with the data at this weather station. The air temperature derived from this sensor has a range of  $16.0\text{-}25.3^{\circ}\text{C}$  and a mean temperature of  $19.8^{\circ}\text{C}$  during the period that has been studied. Relative humidity is also available and has a range of 55.1- 99.9% and mean of 87.7%. As the sensor is placed high above the surface, the calibrated data will also be compared to data at CATIE (taking into account the standard/wet adiabatic lapse rate) and data for August in the WorldClim dataset.

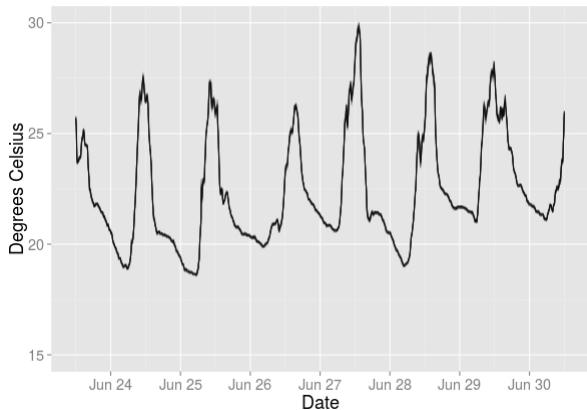
The absolute minimum and maximum temperature (average 2003-2013) at CATIE in August are  $17.3$  and  $30.5^{\circ}\text{C}$ . Taking into the  $\pm 400$  meter elevation difference between both locations and a adiabatic lapse rate of  $6.4^{\circ}\text{C}/\text{km}$ , the range at the location in Aquiares can be expected to be around  $14.8\text{-}27.9^{\circ}\text{C}$ . This is a significantly larger range than was provided by the temperature logger in Aquiares. The WorldClim dataset (Hijmans et al., 2005) contains current, past and future conditions at 30 arc-second resolution, based on 24,542 locations that measured temperature (60% of which also included minimum and maximum temperature observations). In the study area, the maximum temperature is on average  $4.8^{\circ}\text{C}$  higher than the mean temperature, while the minimum temperature is  $4.8^{\circ}\text{C}$  lower in August. This range ( $9.6^{\circ}\text{C}$ ) is comparable to the range provided by the weather station in Aquiares ( $9.3^{\circ}\text{C}$ ) and is smaller than the range at the CATIE station ( $13.1^{\circ}\text{C}$ ).

### 3.3 Results

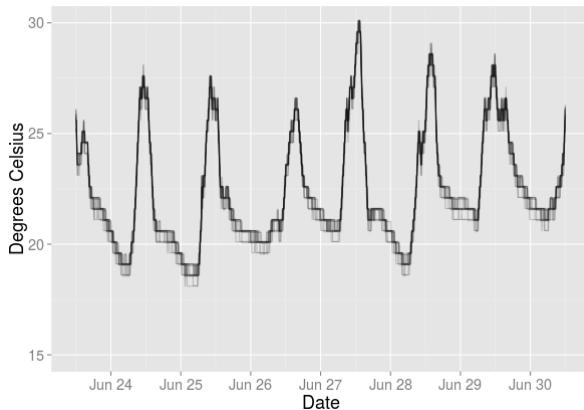
#### 3.3.1 Experiment 1A: Sensor resolution

After the one-week experiment with 15 high- and low-resolution sensors, the (uncorrected) differences for the resulting temperature and humidity data (1-minute interval) were studied. The temperature data of the separate (15) sensors for each resolution are plotted in *figure 4 and 5*. The temperature graphs show that the lower temperature resolution provides a higher maximum difference and standard error at any moment: the maximum difference between any two sensors is  $1.02^{\circ}\text{C}$  for the low-resolution and  $0.65^{\circ}\text{C}$  for the high-resolutions sensors, and the standard error for the low-resolution is  $0.26^{\circ}\text{C}$  compared to  $0.05^{\circ}\text{C}$  for the high-resolution sensors. The differences between the low-resolution sensors are highest during the period without sun (which rises around 5.30AM local time). The minimum temperature also shows the largest difference between the low- and high-resolution sensors ( $0.43^{\circ}\text{C}$ ), compared to  $0.02^{\circ}\text{C}$  for the mean and  $0.10^{\circ}\text{C}$  for the maximum.

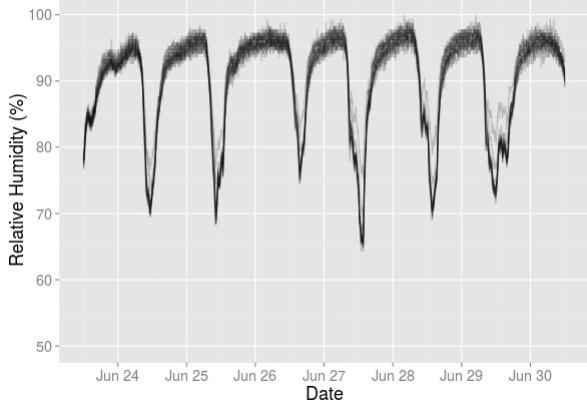
The graphs of the humidity data (*figure 6 and 7*) do not show the same clear differences between the low- and high-resolution sensors as the temperature graphs did. Part of this can be explained by the still relatively high resolution (0.6% on a 40% range compared to  $0.5^{\circ}\text{C}$  on a  $15^{\circ}\text{C}$  temperature range), but also by the micro changes in humidity that occur close to the maximum humidity. The large fluctuations at high humidity - which shows no relation with the different sensor resolution - results in fudging of the data to a  $\pm 5\%$  RH window. The relative humidity range of the low-resolution sensor is 63.9-99.9%, which is 64.3-100.2% for the high-res sensors (>100% values can occur when the sensors are in contact with water). The mean humidity derived from the different resolutions is within 0.5%, with 90.4 for the low-res and 90.9 for the high-resolution sensors. The standard deviation is comparable, with 1.1% for the low-resolution sensors, and 1.2% for the sensors at high resolution. It has to be noted here that high-resolution humidity is stored in 12-bits, while high-resolution temperature data is stored in only 11-bits. This difference in memory size used to store observations could reduce the difference between low-and high-res data for both units.



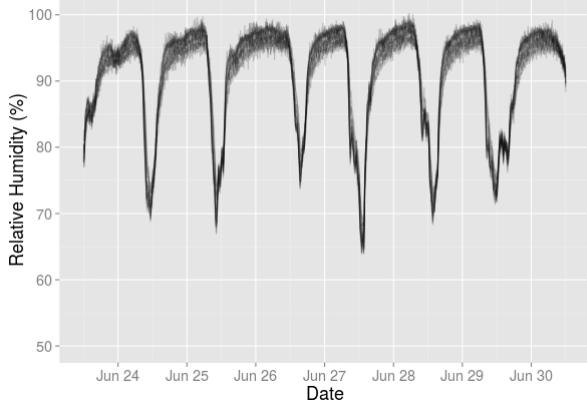
**Figure 4:** High-resolution T sensor-data



**Figure 5:** Low-resolution T sensor-data



**Figure 6:** High-resolution RH sensor-data



**Figure 7:** Low-resolution RH sensor-data

Correlation between the temperature and humidity measurements is strong for both the low- and high resolution sensors. This can be used to reduce the number of the more costly sensors that can measure both temperature and humidity in future networks. The mean correlation over the seven-day period is  $r = -0.91$  for the low resolution sensors (never weaker than 0.86) and  $r = -0.92$  for the high resolution sensors (always stronger than -0.89). This results in linear models with an R-squared of 0.84 (low-res) and 0.86 (high-res), which are strong models that can be used to predict the humidity data based on temperature data alone. The errors of using different resolutions for temperature and humidity observations has been calculated with different methods for both individual sensors and the mean data for all sensors, and is consequently larger between the different temperature resolutions than the low- and high humidity resolution. The average of the absolute maximum difference at any hour is  $0.27^{\circ}\text{C}$  (max:  $1.3^{\circ}\text{C}$ ) for the low-res and  $0.05^{\circ}\text{C}$  (max:  $0.8^{\circ}\text{C}$ ) for the high-res sensor, while for humidity the average absolute maximum difference is 1.2% (max: 11.8%) for the low-res and 1.4% (max: 9.4%) for the high-res observations. While the high-resolution sensors provide better information in all cases, the difference is less significant for humidity than for temperature data.

### 3.3.2 Experiment 1B: Temporal resolution

The assessment of the impact of changing the temporal resolution was based on the same (low- and high-resolution) dataset as the previous experiment, but with a reduced number of observations and with different moments the sensors were launched. The temperature range for any of the low-resolution sensors during the week of experiments was within  $18.11 - 30.12^{\circ}\text{C}$ , which was slightly smaller ( $18.54 - 30.02^{\circ}\text{C}$ ) for the sensors at high-resolution. Over the week of observations, the mean temperature (averages of the 15 sensors) was  $22.37^{\circ}\text{C}$  for the low-resolution sensors and  $22.35^{\circ}\text{C}$  for the high-resolution sensors. The average minimum temperature was  $19.44^{\circ}\text{C}$  for the low-resolution sensors and  $19.81^{\circ}\text{C}$  for the high-resolution sensors. The average maximum temperature was higher for the low-resolution sensors, with  $28.34^{\circ}\text{C}$ , while this was  $28.10^{\circ}\text{C}$  for the low-resolution sensors. This shows that the range for the low-resolution sensors is around  $0.5^{\circ}\text{C}$  larger, while the mean temperature remains similar for both sensor types. *Table 5* shows the impact of different temporal

intervals, interpolation approaches and launching moments on the resulting temperature in different quantiles. It is clear that from the minimum temperature up to the median/mean temperature, differences are very small, and longer intervals generally result in slightly higher minimum temperatures. Differences become larger in the 3<sup>rd</sup> quantile (0.75), although in most cases still remain small.

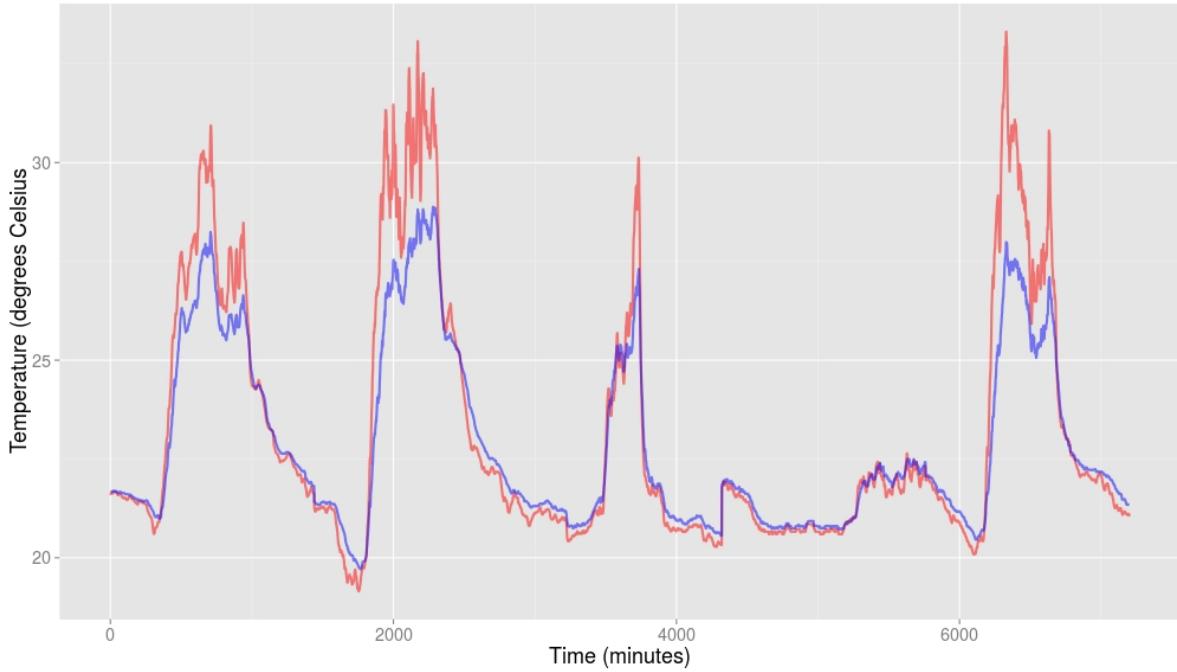
The first medium ( $>0.25^{\circ}\text{C}$ ) differences can be seen for some of the 2-hour data that have been created with the high-resolution sensor. At the 0.9 quantile, all values have at least a small difference, and in this case the low-resolution sensors show more medium differences as well, which can be found more for the 1-hour than for the 2-hour intervals. At the 0.95 quantile, the differences become smaller again, with the four 1-hour/low-resolution datasets providing insignificant differences. This is the case for the splined 1-hour high-resolution datasets as well. Only the closest observation datasets at 2-hour (both sensor types, both interpolation approaches) provide medium differences with the original dataset. The differences with the maximum temperature during the seven days of measurements is in almost all cases medium to large. The only small difference is found for the high-resolution sensors with random start moment and 1-hour interval. Differences are large for all - except for one (high-resolution sensor with random start moment) - of the 2-hour interval datasets. The standard deviation is small (although in line with the original dataset) for the low-resolution sensors, while this is insignificant for the high-res sensors (closest), and small for the high-res sensor (random). The maximum differences at any moment are especially problematic for the data with *random* start-moment, where they are in all cases larger than in the *closest* dataset.

**Table 5:** Differences in temperature at different quantiles using a range interpolation approaches

	Min	0.05	0.1	0.25	0.5	Mean	0.75	0.9	0.95	Max	St.dev	Diff
<b>Low-resolution</b>	19.44	19.94	20.04	20.54	21.60	22.37	23.97	25.84	26.84	28.34	0.26	1.07
<i>Closest</i>												
1-hour linear	+0.01	0.00	-0.02	0.00	-0.01	+0.01	-0.16	+0.27	+0.06	-0.37	0.26	0.93
1-hour spline	0.00	-0.01	-0.02	0.00	0.00	+0.01	-0.16	+0.27	+0.06	-0.32	0.27	1.01
2-hour linear	+0.01	0.00	-0.04	0.00	+0.03	+0.03	-0.13	+0.18	+0.40	-0.63	0.26	0.92
2-hour spline	0.00	0.00	-0.05	0.00	+0.03	+0.03	-0.12	+0.18	+0.40	-0.59	0.26	0.96
<i>Random</i>												
1-hour linear	+0.07	+0.01	+0.08	+0.04	0.00	0.00	-0.11	+0.15	-0.02	-0.31	0.27	1.18
1-hour spline	-0.02	-0.01	+0.05	+0.01	-0.01	0.00	-0.09	+0.25	+0.06	-0.23	0.28	1.36
2-hour linear	+0.16	+0.07	+0.16	+0.12	+0.02	+0.01	-0.23	+0.11	-0.10	-0.90	0.32	1.44
2-hour spline	+0.03	-0.02	+0.04	+0.03	-0.02	+0.01	-0.18	+0.25	+0.21	-0.62	0.33	1.58
<b>High-resolution</b>	19.81	19.98	20.10	20.59	21.60	22.35	23.95	25.80	26.87	28.10	0.05	0.46
<i>Closest</i>												
1-hour linear	+0.06	-0.01	-0.01	-0.01	-0.04	+0.01	-0.20	+0.23	-0.10	-0.33	0.04	0.32
1-hour spline	+0.06	-0.01	-0.01	-0.01	-0.04	+0.01	-0.20	+0.23	-0.09	-0.32	0.05	0.32
2-hour linear	+0.09	-0.01	-0.02	+0.04	-0.04	+0.03	-0.25	+0.10	+0.46	-0.63	0.04	0.31
2-hour spline	+0.09	-0.01	-0.02	+0.04	-0.04	+0.03	-0.25	+0.11	+0.47	-0.62	0.05	0.31
<i>Random</i>												
1-hour linear	+0.07	+0.02	0.00	0.00	-0.02	+0.01	-0.15	+0.22	-0.12	-0.31	0.10	0.64
1-hour spline	+0.05	0.00	0.00	0.00	-0.02	0.00	-0.13	+0.21	-0.06	-0.13	0.10	0.94
2-hour linear	+0.12	+0.08	+0.04	0.00	-0.01	0.00	-0.20	+0.15	-0.10	-0.61	0.19	1.03
2-hour spline	-0.02	+0.01	0.00	+0.06	-0.06	+0.01	-0.27	+0.29	+0.18	-0.46	0.19	1.41

### 3.3.3 Experiment 2: Shields

An initial test to compare the temperature in the control group (tubes with the least external coating) with the temperature in the Stevenson shield clearly shows much higher peaks in temperature in the PVC shields during the midday period (*figure 8*). This can be seen for all days, except the fourth day (which runs from 3PM to 3PM) where temperature does not surpass  $23^{\circ}\text{C}$ . This was a completely overcast day with very limited temperature variation. Whereas the temperature rises faster (and peaks are higher) in the PVC shields, it also drops faster, resulting in lower temperatures in the PVC shields than in the Stevenson shield when the temperature is lowest (after the sun has set). This results in both a higher maximum temperature and a lower minimum temperature in the PVC shields, although the difference is smaller - and not problematic - for the minimum temperatures.



**Figure 8:** Temperature in the Stevenson (blue) vs. PVC (red) shields

A visual analysis of the  $T_{\text{PVC}}$  vs.  $T_{\text{Stevenson}}$  graph (*figure 9*) shows that temperature in the Stevenson shield and the control PVC shields are similar up to  $24^{\circ}\text{C}$ , which are often periods without sun (*figure 8*). The maximum temperature is several degrees higher in the PVC shield, while the minimum temperature is a bit lower. After the point where temperature in the Stevenson and PVC shields is relatively similar in the morning, a more curvilinear pattern is present. During periods of high temperature, more (short) sunny periods occurred, which warm up the PVC shields quicker than the Stevenson shield. The change in time in the Stevenson versus the control tubes is shown in *figure 10*. The linear models to predict change over 15 and 30 minutes, as well as 1 and 2 hours, are provided on the next page. The linear models of the change in temperature over time in the Stevenson versus the change in the different PVC shield adjustment is provided in *Appendix B*.

$$15 \text{ mins} : T_{PVC} = 1.83 \times T_{station} \quad (R^2 = 0.75)$$

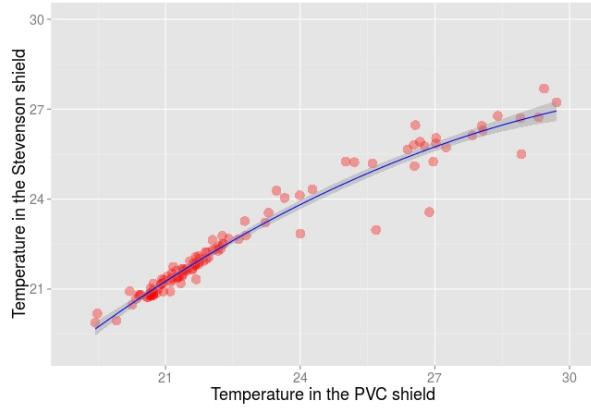
$$30 \text{ mins} : T_{PVC} = 1.65 \times T_{station} \quad (R^2 = 0.81)$$

$$1 \text{ hour} : T_{PVC} = 1.52 \times T_{station} \quad (R^2 = 0.87)$$

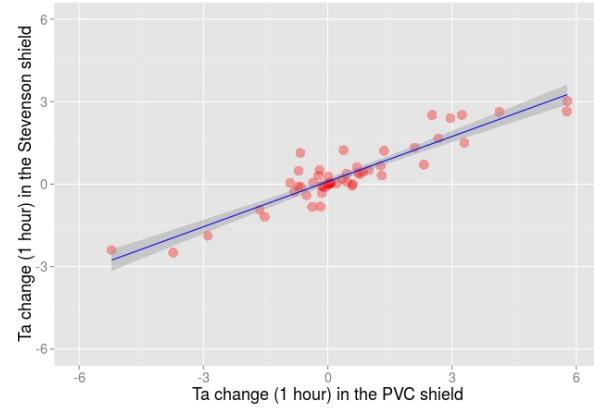
$$2 \text{ hour} : T_{PVC} = 1.46 \times T_{station} \quad (R^2 = 0.90)$$

The linear models that provide the slope closest to 1 and the highest R<sup>2</sup> can be expected to provide the most reliable results, but it has to be taken into account that there will be trade-offs with costs as well. The ‘best’ adjustment to the shields on which temperature-calibration can be based is adding insulating foil, which has an R<sup>2</sup> that is 0.05 higher than the control tube, and a slope that is 0.09 lower (*Appendix B*). The control group (no adjustments) provided the largest difference in rate of heating/cooling (>50% faster than in the Stevenson shield). A problem that occurred with the later experiments (smaller tubes) is that, during the experiments, the fence around the meteorological station was cleaned by a high-pressure hose, which resulted in erroneous measurements. On the fifth day of experiments, the relative humidity in the Stevenson shield was above 100% for over 40% of the day, which is likely the result of cleaning of the meteorological station. On the fourth day of experiments, humidity also showed a different trend from the other days, but in this case a similar trend (limited fluctuation over the day, but large over a short period) is also seen in the PVC tube, with 32% of the day having a relative humidity >100%. This could be the result of a fully rainy day. The day with incorrect measurements in the Stevenson is removed, the other four days are used.

The linear models and coefficients of determination (*Appendix B*) only explain part of the impact of the different coatings in dealing with different circumstances; *tables 6 and 7* provide the temperature statistics on different types of day and in PVC tubes with the smaller (25mm) and larger (50mm) diameter. The same analysis for humidity is provided in *Appendix C*. The insulating foil provides the closest match for the maximum temperature on sunny days, although this difference is still >3°C in both types of tubes. Insulating foil also provides the closest match to minimum temperature on the sunny days, and is <0.2°C. On the cloudy days, the overall differences are smaller, and only the shield with holes clearly provides worse results than the other adjustments (larger standard deviation and difference with maximum temperature). The standard deviation, which is based on the different shields with the same coating (each with 2 sensors), is in the same range for most coatings on a sunny day (except for reflective tape, which shows more deviation). Holes show both the highest and lowest deviation on the two cloudy days, indicating that - while the benefits of increased aeration can be significant - there is a risk of sensor contact with water when working with low-cost shielding.



**Figure 9:** Correction based on actual Ta



**Figure 10:** Correction on change of Ta in time ( $\Delta t$ )

The analysis of humidity data in the different shields (*Appendix C*) has shown that the maximum relative humidity in the shield surpasses 100% on every day, while this is *normally* not the case in the Stevenson shields. The cloudy day on which the 25mm tubes were tested, however, shows a very high humidity (114% in the Stevenson). This was the day that the border around the meteorological station was cleaned, and water has obviously infiltrated the Stevenson (where water could not easily escape). This also indicates that there is a risk of not having open-bottomed shields. The mean humidity that is provided by the shields is relatively similar, while insulating foil seems to increase the maximum humidity (and standard deviation) on sunny days, and holes (in combination with foil) has provided clear incorrect data on the cloudy day when the 50mm was tested.

**Table 6:** Analysis of adjustments on 50mm PVC shields

	Sunny day				Cloudy day			
	Mean	Min	Max	Dev	Mean	Min	Max	Dec
Stevenson	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	34.43	0.42	21.77	20.27	30.17	0.20

**Table 7:** Analysis of adjustments on 25mm PVC shields

	Sunny day				Cloudy day			
	Mean	Min	Max	Dev	Mean	Min	Max	Dec
Stevenson	23.23	20.42	28.12	0.04	21.40	20.68	22.53	0.04
No adjustment	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	0.20

### 3.3.4 Experiment 3: Data calibration

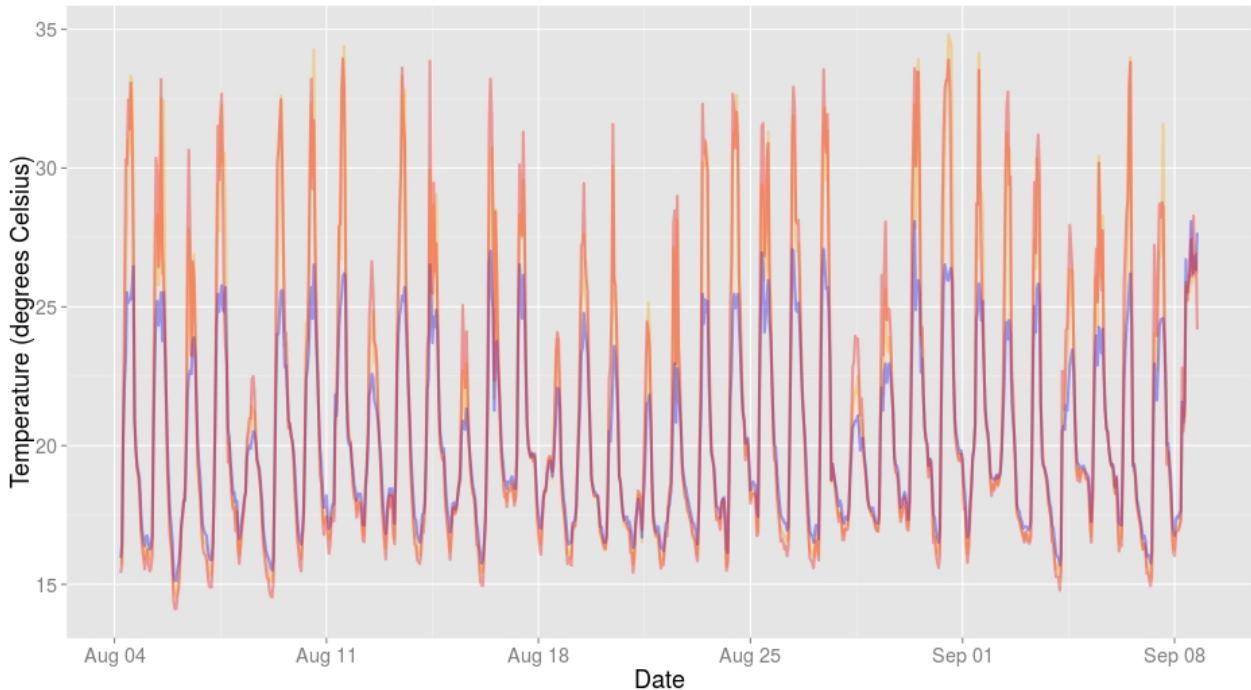
The median moment with the smallest difference between the (control) PVC shields and the Stevenson shield - within a 2-hour window from the sunrise (5.30AM) during the five days of experiments - was 5.55AM, which is rounded to 6AM in further calculations. The moment in the evening with the lowest difference during the experiments was 4.56 pm, rounded to 5PM in the model. The number of hours that has to be calibrated every day, in the case of calibration on the change in temperature over time, is 10 hours: the first calibrated hour is 7AM and the last is 4PM. The hours between midnight and 6AM, as well as 5PM to midnight, will not be calibrated. The temperature outside the calibrated range is slightly lower in the PVC shields than in the Stevenson, (*figure 8*), which means a certain fixed temperature will be added to the PVC-temperature to get a closer match to the control temperature. Addition of a temperature of  $0.22^{\circ}\text{C}$  in this ‘night’ period showed the best match to the Stevenson temperature, and has reduced the RMSE during this period to  $0.16^{\circ}\text{C}$ .

The performance of the different models, compared to *no calibration*, are provided in *table 8*. The polynomial models perform clearly better when moving from 1<sup>st</sup> (linear) to 2<sup>nd</sup> (exponential) order, while performance of the 3<sup>rd</sup> order polynomial model is comparable to that of the 2<sup>nd</sup> order polynomial model. The fourth order polynomial model has an average RMSE of 0.25, which is a very small improvement compared to the extra order that is used in the model. The polynomial model is a clear improvement over the data without calibration, especially for the day-time temperature, maximum difference and also the overall RMSE. Calibration based on change in time, which has a linear model - calibrated over 55 hourly observations - with an  $R^2$  of 0.88, is done for 10 hours. With no additional parameters (adjustment of night-time temperature and fitted change in time), this provided only a small improvement over the non-calibrated data. The RMSE during the night remains the same (no data-modification), and there is only a limited improvement during the day and for the maximum difference between the data sources. The best-calibrated model multiplies the fitted change in time by a factor of 1.35 (after which  $0.11^{\circ}\text{C}$  is added), while the night-time temperature in the PVC is increased by 0.22 during the other 14 hours. The result of this model has a RMSE similar to 2<sup>nd</sup> order (and higher) polynomial models, and performs well during day and night. The main difference with the polynomial model is that the maximum difference at any moment remains higher. Cross-validation of both calibration approaches is provided in *Appendix D*.

**Table 8:** Errors in different types of calibration (Ta) compared to data in a Stevenson shield

	No calibration	1st order polynomial	2nd order polynomial	3rd order polynomial	Delta-T, no parameters	Delta-T + parameters
RMSE overall	0.66	0.33	0.26	0.26	0.57	0.26
RMSE night	0.27	0.22	0.15	0.16	0.27	0.16
RMSE day	1.21	0.48	0.42	0.40	1.00	0.39
Max error	4.83	2.12	2.13	2.06	3.01	2.41

Calibrating has been done under homogeneous conditions, with all sensors at the same elevation and with similar exposure to sun and wind (which is the objective of a meteorological station). The 2<sup>nd</sup> order polynomial has been selected as the best polynomial model, with overall good statistics with a relatively low-order. Calibration of the change in time, with the previously provided parameters, is selected as the best model to predict temperature based on change over time. These models were validated with a leave-one-out approach, which has resulted in similar model parameters and also a low ( $<0.5^{\circ}\text{C}$ ) RMSE for the day that was left out (*Appendix D*). This indicates that the model can explain the warming in the shields (*thermal inertia*) during the studied period. This is a preliminary model that was also tested in another area. Three sensors are placed at close distance to a certified weather station (*sensors 31, 41 and 57*). The uncorrected data of these sensors is plotted in *figure 11*, and shows a range of  $14.1 - 34.8^{\circ}\text{C}$ . The problem that becomes clear from this figure is that sensors in comparable shields and at a similar location can still have very different temperature ranges. In this case, two sensors provide clear extreme ( $> 30^{\circ}\text{C}$ ) temperatures during the period around noon - which is in line with the previous experiments at the CATIE weather station - while the other sensor shows a much stabler temperature. The most obvious explanation for this difference would be either a position of permanent shading or contact of the sensor with dirt or water. The trend of this *cool* sensor has not been encountered during any of the previous experiments, and does not show any indication of contact with water or other errors, as there is still sufficient fluctuation. The mean is also quite different between these sensors; the sensors with high peaks have a mean temperature of  $\pm 21^{\circ}\text{C}$ , while this is a low  $20^{\circ}\text{C}$  ( $20.2^{\circ}\text{C}$ ) for the *cool* sensor, which means they are still within  $1^{\circ}\text{C}$ .



**Figure 11:** Observed temperature in three sensors around the weather station (blue = ‘cool’ sensor)

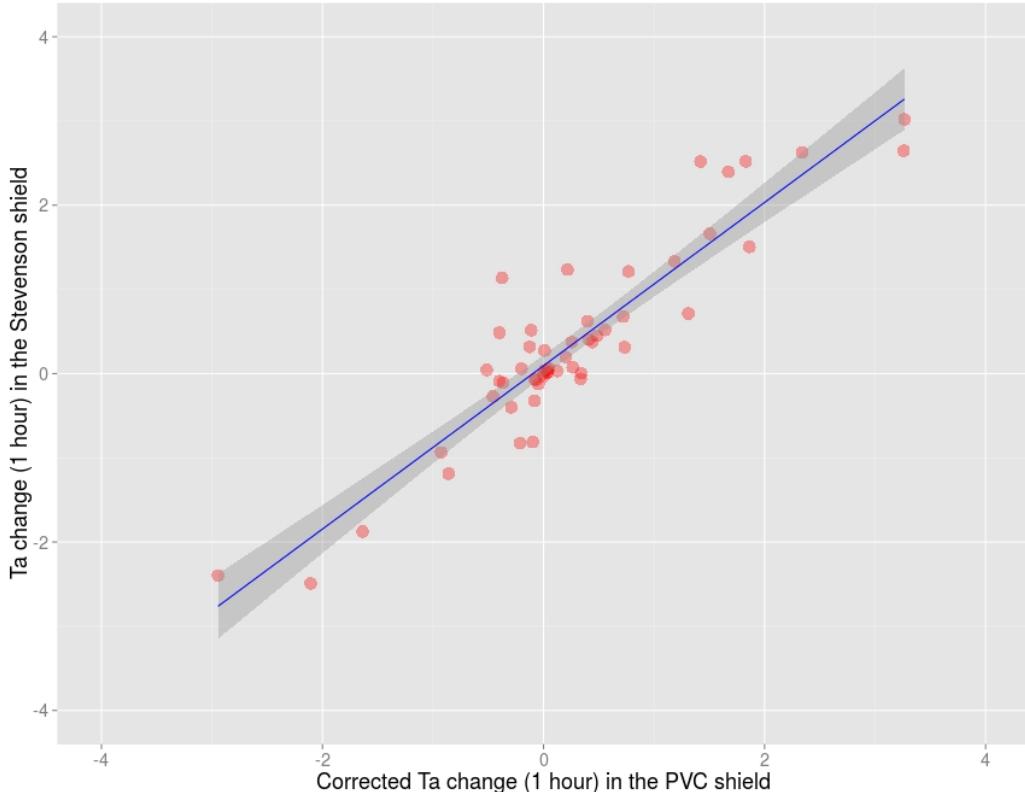
*Figure 11* also shows that extremes are caused by the moment of day, as even on relatively cool days the temperature around noon can differ several degrees between the different sensors. Based on the visual analysis and earlier experiments, it seems that most of the sensors have performed similar to the two sensors that showed the  $>30^{\circ}\text{C}$  temperatures, which means correction will mainly focus on these sensors and aims to limit the impact of data-correction on the *cool* sensors. Based on the temperature range of the shaded sensor in *figure 11* ( $15.1 - 28.1^{\circ}\text{C}$ ), it can be expected that this will be in line with the temperature range as measured by certified stations. The results of different calibrations (*Appendix E*) generally results in a minimum temperature below  $15^{\circ}\text{C}$  during the studied period, which is  $1^{\circ}\text{C}$  lower than in the local weather station.

The linear calibration, however, provides a  $>16^{\circ}\text{C}$  minimum, which could be considered closer to the temperature measured at the local weather station. The minimum of the three other calibration approaches is generally still close to that of non-calibrated data (which has performed well in the CATIE-experiments), and also closer to the temperature ( $14.8^{\circ}\text{C}$ ) that can be expected based on a standard adiabatic lapse rate. The median temperature in the different calibrations ranges between  $18.9$  and  $20.2^{\circ}\text{C}$ . The maximum temperature fluctuates even more, which has a minimum value of  $23.7^{\circ}\text{C}$  for the *cool* sensor when calibration on  $\Delta t$  (change-in-time) - without calibration parameters, and a highest value of  $31^{\circ}\text{C}$  for a sensor in the calibration based on a linear relationship of the temperature in the PVC shield and in the Stevenson. The two best approaches for CATIE data-calibration (exponential and change in time with parameters) provide very different outputs when using them for the actual field data, which becomes especially clear for the higher values. The maximum of the change-in-time (with additional parameters) provides a maximum that can be considered impossible for the study area - which is not the case for the exponentially corrected data. Only from the 0.97 quantile, however, the data for the  $\Delta t$  method becomes higher than the polynomial corrected data. At the 0.9 quantile, polynomially corrected data is still over  $1^{\circ}\text{C}$  higher than data derived from  $\Delta t$  correction. This means only a small part of the data that is corrected by the latter method can be considered to be incorrect. One option to avoid these extreme values would be to set a certain quantile as the maximum temperature. Assuming change-in-time correction faces problems for 2% of the data, it could be decided to reset data above the 0.98 quantile to the 0.98 quantile. This is an easy way to avoid outliers in the dataset, although it also smooths the data.

A final test has assessed whether the calibrated values still show the expected *adiabatic lapse rate* over the study area, and whether they show similarity to the minimum, mean and maximum temperature assessed in the WorldClim dataset. The difference in mean, minimum and maximum in the WorldClim data for August (current climate) over the study area is around  $2^{\circ}\text{C}$  ( $\pm 600\text{m}$ ). This is similar for the four tested calibrations (linear, exponential,  $\Delta t$  with/without parameters), with a range between  $1.75-2.25^{\circ}\text{C}$  (*Appendix F*). The maximum value for the different mean temperatures (at the lowest elevation), differs by up to  $1.5^{\circ}\text{C}$  (linear vs  $\Delta t$  without parameters). The trend for the minimum temperature in all cases shows a lower minimum in the higher regions. The differences are very small for the change-in-time calibrations (which do not adjust the minimum temperatures), with values around  $0.8^{\circ}\text{C}$ . While this is a very small difference when considering the adiabatic lapse

rate, this is the absolute minimum that was assessed during one month of observations under very wet conditions (especially during the night-time). There could also be local differences caused by mountain flows (McNider & Pielke, 1981); different factors can influence the collection of cooler air in the valleys (Barr & Orgill, 1989). Measurements of the minimum were quite accurate during the earlier experiments at CATIE; differences in observations are thus not a significant problem.

For the maximum temperature, the differences between the lower and higher areas of the study area are lowest for the exponential calibration ( $1.8^0\text{C}$ ) and the  $\Delta t$  method without additional parameters ( $3.0^0\text{C}$ ), while the other two calibration approaches provide differences of  $4.2^0\text{C}$  (linear model) and  $4.7^0\text{C}$  ( $\Delta t$  with parameters). The calibration approach which provided the lowest RMSE during the test at the CATIE station ( $\Delta t$  with parameters), provides maximum temperatures that can be considered unrealistic. Not including parameters, however, provides a good match to the data of the weather station, while it also continues to show clear trends for the maximum temperature and does not change the minimum. This correction, however, performed worst during the experiments at CATIE. After calibrating the hourly temperature change with  $\Delta t$  calibration, the new values were close to the values in the Stevenson. Hours during which temperature only changes slightly show more outliers than the hours with larger changes (*figure 12*). This shows that this calibration is useful to predict periods of large heating/cooling, which are the most important to assess the extremes.



**Figure 12:** Corrected change in time (1 hour) based on a calibration model based with 55 values

### 3.4 Discussion

#### *Experiment 1: Sensor resolution*

The high-resolution sensors show a smoother temperature curve than low-resolution sensors, resulting in a lower standard deviation and a lower maximum difference between any two sensors in the Stevenson shield at any moment. The humidity, however, shows similar graphs for the low- and high-resolution, which - together with a similar mean and standard deviations - indicates that high-resolution data is not necessary for measuring the relative humidity. The strong correlation between temperature and relative humidity ( $R^2 > 0.84$ ) will make it possible to use only few sensors with both temperature and humidity sensors. Data from these sensors can be used to create a linear model that will enable prediction of relative humidity at locations where only temperature will be measured. Changing the temporal resolution - and applying different corrections - shows that a shorter (1-hour) interval generally provides better results than a longer (2-hour) interval. While this could still be acceptable in some cases, the long interval will make it very difficult to calibrate data based on the change in time ( $\Delta t$  calibration), which could be a good way to correct data.

Another result of the first set of experiments is that using the closest measurement to the actual hour provides much better results (at the hour), than when estimating this value at the hour based on (linear/spline) correction of measurements taken at random moments in the hour. Linear interpolation of measurements performs well - in most cases even better than splines - when working with hourly measurements. When measuring only once every 2-hours, splines have the potential to provide a better match to maximum temperature, although it can also result in a lower minimum compared to the results derived by using linear interpolation. The methods that provide relatively large ( $>0.5^0\text{C}$ ) - and possibly problematic - errors, mainly occur at the 2-hour intervals, while a random start moment generally increases the maximum difference at any moment. All 1-hour intervals are acceptable, whereas it will depend on the use of the network whether longer intervals can be used as well. Long-term studies that assess basic indicators can be done with 2-hour intervals, while calibration of the data is limited with this interval. Although differences in the maximum differ by up to  $0.9^0\text{C}$ , the differences at the 0.95 quantile are always smaller than  $0.5^0\text{C}$  compared to the temperature in the original dataset, making this data more reliable than the maximum.

#### *Experiment 2: Sensor shields*

The experiments with different adjustment to the PVC shield showed that none of the tested alternatives provides an acceptable temperature that is in line with the temperature in a certified shield. This is especially a problem on sunny days and is similar in structures of different (25 and 50mm) diameter. Differences in temperature can be close to  $6^0\text{C}$ , while maximum humidity surpasses 100% in all experiments - reaching up to 106% RH. Whereas none of the adjustment perform sufficient for direct use (without calibration) in the field, the best coating to reduce the maximum temperature in the experiments has been insulating foil. This has provided the closest match to the temperature in the Stevenson shield in most cases. The second-best performing adjustment has been the addition

of holes, while a combination of holes and insulating foil has provided surprisingly bad results in the tests. The best adjustment to provide accurate humidity data is not very clear from the experiments. The best performing shielding differs on most days, with both the best and worst match for insulating foil in different experiments.

While the tested PVC shields are not good enough to provide data that can be used without calibration, the different coatings show a clear linear relationship when it comes to the change of temperature over time. This ranges from a 43 to 52% faster increase in temperature in the shields versus the Stevenson, with the strongest models for shields with holes and with insulating foil. The strongest model ( $R^2$  of 0.92) is found for insulating foil, which means calibration based on a change of time ( $\Delta t$ ) could be feasible. The insulating foil - which is available in most regions - is normally used to insulate houses, and has been used by different research organizations in similar structures in the study area. The smooth surface of the foil, together with the silver colour, can be expected to reflect a significant part of the radiation heat. While the reflection of the coating plays an important role, the possible impact of surface radiation has not been well studied in the experiments. In all experiments, the sensor has been placed above the surface on a glass fibre mesh, which does not prevent a possible impact of reflecting radiation on the sensor. The small diameter of the PVC tube and the different absorptivity and thermal conductivity compared to the Stevenson shield (which is a specialized and certified structure) can result in a different impact due to thermal inertia, which is '*the degree of slowness with which temperature of a body approaches that of its surroundings and which is dependent upon its absorptivity*' (Merriam-Webster dictionary).

Based on the four days of experiments, insulating foil can be considered the best performing coating. Ventilation by holes, however, only provides slightly better results than no adjustment/white paint, which indicates that the type and location of holes in the shield has to be improved. This is also clear from the results of the holes + foil adjustment, which does not show the cumulative benefits of the separate adjustments. The main problem for humidity observations is derived from its link to temperature, which is often exaggerated in the shields. Another problem is the danger of water intrusion, which is especially a risk for shields with holes and no tape around loose parts.

### *Experiment 3: Data calibration*

Calibration of the PVC shields in the experiments at the CATIE meteorological station showed that a low error during both night and day can be achieved by different types of models. A calibration that changes the PVC temperature at any moment to the average Stevenson temperature that can be associated to this, is an easy approach that can quickly be done for complete datasets. The disadvantage is that this polynomial relationship is calibrated on a certain set of days at a different location and thus is not well suited to correcting temperature and humidity outside of the calibrated range. Another limitation is that this approach is not influenced by the physical properties of the tube. A temperature above 30°C can be an error due to the thermal inertia differences in the PVC tube on one day, while it could be the actual temperature on other days. A 2<sup>nd</sup> order polynomial

model can reduce the overall RMSE to 0.26, with values of 0.15 during the night and 0.42 during the day. The advantage of this model is that it is easy to understand and implement, and - while errors may occur when using the model outside of the calibrated range by changing the studied region or period of observation - the daily errors can be expected to be relatively low, as each temperature is adjusted without building up errors along the way.

Calibration based on the change in time is an approach that is based on a strong and significant linear model between the changes in the PVC tubes and the Stevenson shield, which is caused by the different dimensions, absorptivity, and resulting thermal inertia inside the PVC shields. Calibration can be done by creating an anchor point on every day, which is a moment around sunrise/sunset when differences between the Stevenson and PVC shield are still relatively low. The observations made during the night do not have to be calibrated, although a small adjustment could be done if the temperature has a small, constant offset. Applying this  $\Delta t$  calibration, based on a linear prediction model without additional parameters, by building up the temperature for every day starting at an anchor point, shows only a small improvement over the non-calibrated data. Adding additional correction in the model (three basic parameters) can reduce the overall RMSE to a lower level than that of the polynomial model. Validation based on a leave-one-out cross-validation shows that the model is very accurate in prediction the temperature in the day that has been left out. The disadvantage is that this model has been tested in a location without any shelter, which could change the performance of the shielding. The corrected change in time is provided in *figure 12*.

Applying the different calibration approaches to field data indicated additional differences. The *control* temperature data against which the shield-temperature is compared shows a small range, which is different than would be expected based on data at the CATIE meteorological station - taking into account a standard adiabatic lapse rate. Testing the performance of shields around the weather station in Aquiares showed that different calibration approaches can provide very different output. While the polynomial model and change-over-time model provide similar results at the CATIE weather station, the results from applying these corrections to data from the field show larger differences. The mean remains similar between these approaches, but the minimum that results from polynomial correction is exaggerated, as this lower range has not occurred during the calibration period. The maximum temperature is also lower in the dataset that has been corrected with the polynomial model. This is mainly a result of the low maximum temperature that has been measured in the Stevenson, which means temperature is always corrected to  $<29^{\circ}\text{C}$ . While the 2<sup>nd</sup> order polynomial correction provides data that is best in line with data from the local temperature logger, the lack of calibration in the study area, as well as the lack of relationship to the physical properties of the PVC tubes that cause the faster warming, are an important limitation to this approach. The preferred approach is the calibration based on change in time. This does not correct night-time temperature and is based on the thermal inertia (faster heating) of the PVC shields. The main constraint is that errors can be large on separate days, due to possible cumulative errors.

### 3.5 Recommendations

Based on the experiments that have been conducted - and taking into account the accuracy at which most crop and disease thresholds are provided - a low-resolution of the tested sensors, for both temperature and humidity, will be sufficient for most agricultural studies. The strong correlation between temperature and humidity also makes it possible to reduce the number of sensors that measure both temperature and humidity, although the ratio of T to T+RH sensors will depend on the available budget and the use that humidity data has in the project (mainly important for disease thresholds). The interval that has to be used can either be 1 or 2 hours, which will mainly be influenced by the type of calibration and crops that will be studied. With annual crops (most cereals) that are only in the field for half a year, a 1-hour resolution will be sufficient to cover the full growing season, while for perennial crops, a 2-hour resolution can cover a period of almost one year and would thus be favoured in these cases (periods are based on sensors that measure both units at low-resolution). Creating a script that can launch all sensors at the same time is expected to be a relatively small exercise for someone who is experienced with this, and can be expected to be worth the investment by improving the accuracy of the data and enabling longer intervals.

The current adjustments to the shields cannot adequately reduce the impact of radiation during the midday period and shows large differences with the control (Stevenson) shield on sunny days. This exaggerated temperature is linked to a very low humidity, which together indicate a situation of crop-stress which is not actually present. While insulating foil shows the best results, more study has to be done with combinations of aeration (holes) that reduce the impact of thermal inertia in the relatively small structure. Regarding the structure, a recommendation is to place the sensor in a position where it cannot be reached by reflecting radiation (and water does not built up), which was not the case in the experiments. Placing the sensor inside an inverted *U-shaped* tube (attached to a main structure) could increase aeration, and will make it possible to cover the full structure with insulating foil. These shields can be constructed with the same costs, although not all problems might be fully solved. Calibration based on a polynomial model is relatively easy and can correct the data at any moment without building errors along the way. The limitation is the difficulty to apply this approach in different areas than were calibration has originally taken place. Calibration based on change-in-time ( $\Delta t$ : 1-hour) can result in larger differences on certain days, but is based on the physical properties of the PVC shield and is thus expected to provide more accurate data.

It is recommended that future studies will mainly focus on improved shielding of the sensors, as this will reduce the need of calibration and hence of expected errors. Priority should thus be given to robust shielding, especially (a) reducing the impact of radiation through ventilation and reflection; and (b) reducing the risk of humidity intrusion through appropriate sealing of the area around the sensor. As shields might not function the same in all cases, it is also recommended to always have a set of sensors nearby a certified weather station for the period of study. Having access to a certified shield can be used to assess the relationship between temperature and humidity, which can be used to create a temperature-humidity model to predict humidity and thereby reduce need for RH-sensors.

## 4 Spatial aspects

### 4.1 Introduction

The objective of this chapter is to assess the relationship between the temperature/humidity data derived from a sensor-network and a number of variables that can be derived from open-access satellite imagery (especially Digital Elevation Models). Understanding the correlation between a number of these variables and temperature/humidity, can help to reduce the required number of sensors in a certain study area, by enabling geostatistical interpolation approaches that can use this relationship to predict values in areas where sensors have not been placed. The main objective is thus to find a scientific basis - based on the accuracy and correlation of the data derived from the points (*sensors*) - on which spatial interpolation can be based to create a geospatial network. The variables that will be included in this analysis include topography (elevation, aspect, slope), land-use (leaf area index - LAI, canopy height), and factors related to the sun-position (hill shade, cast shadow and irradiation). The influence of topography on climate - especially rainfall and temperature - is well studied and the relevant literature has been used to understand what strength of correlation can be expected to be useful in co-kriging/universal kriging (*next chapter*). Maps that cover canopy height and LAI are not as easily available as DEMs, but are expected to play an important role in climatic differences at the scale that is studied in this thesis. Factors related to the sun can be derived from some advanced solar calculations. Based on the clear peaks in temperature that the (uncorrected) sensors provide during the period of solar noon, it is expected that this can explain some of the variation in temperature in a terrain where cast shadows play a large role during certain hours.

In this chapter, the data that has been collected in a coffee plantation during a period of over one month will be studied. The sampling strategy, as well as a detailed background of the topological/vegetation-related conditions on the study area, are explained. The data has been analysed in several ways to assess the accuracy of the data. Basic interpolation approaches have been applied (Inverse Distance Weighted) to assess the accuracy of the network, based on a leaving-one-out (LOO) cross-validation approach. A period of 31 days has been used for this cross-validation. More advanced spatial interpolation approaches are analysed in *section 5*; in this chapter the RMSE was analysed for a period of one week to reduce the time required for complex computations. After an analysis of the initial network accuracy, the correlation between temperature (humidity is not analysed due to the strong correlation with temperature) and six static covariates, as well as two dynamic covariates was assessed. This was done by analysing the correlation at every hour during the one-month of observations and subsequently calculating the average hourly correlation during this period. The relationship between the *errors* and the *strength of correlation* of temperature with covariates is assessed, to understand the types of days on which the accuracy is highest/lowest. It is also assessed which variables have the largest impact, and which selection of covariates has to be included at different hours in order to create the strongest linear model that can explain temperature variation.

## 4.2 Methodology

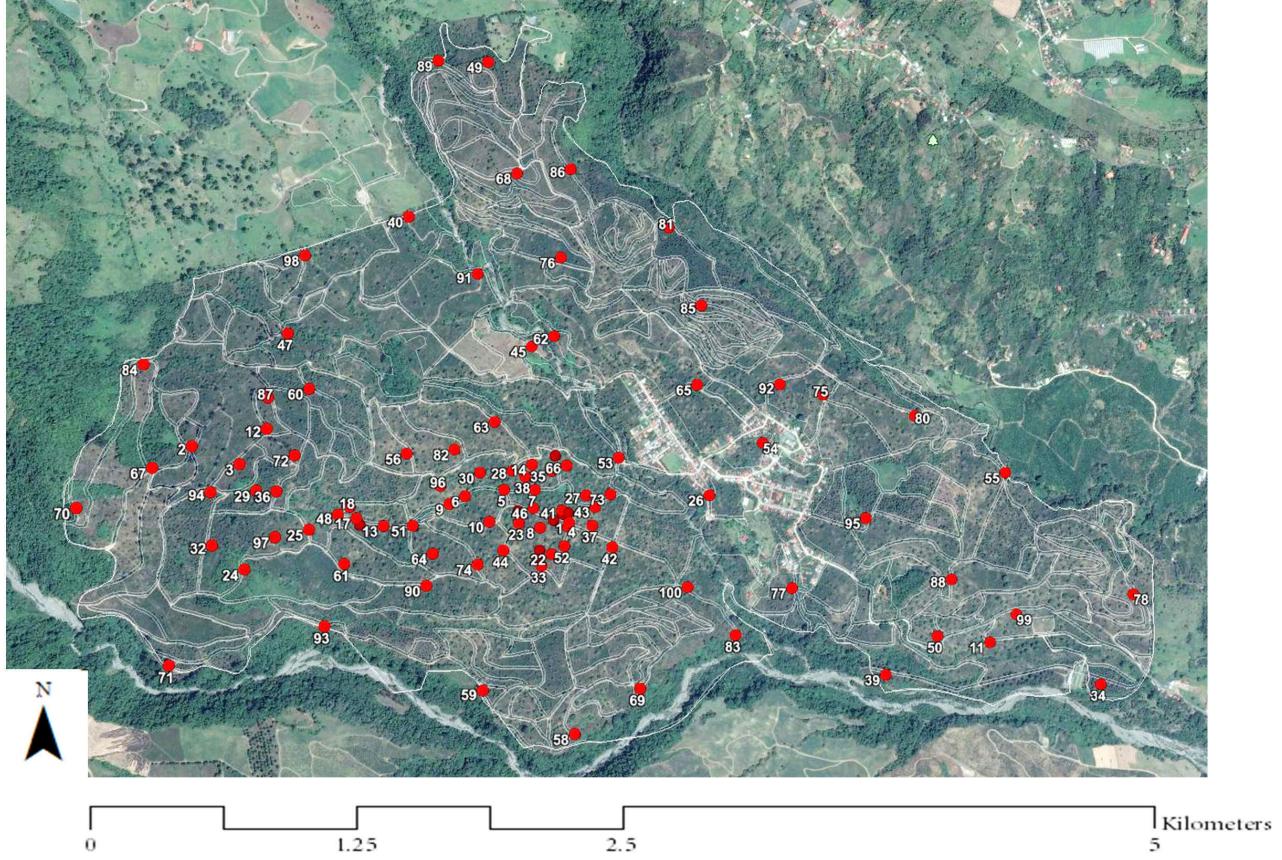
### 4.2.1 Study area and sampling

#### *Study area*

Aquiares is a large coffee estate in the Turrialba region in Costa Rica. The *Finca Aquiares* is one of the largest coffee plantations in Costa Rica, located at a latitude of  $9.95^{\circ}\text{N}$ , and longitude of  $83.72^{\circ}\text{W}$ . The estate covers around 924 hectares, of which 673 are cultivated with (shade-grown) Arabica coffee. The farm is located in between the Pacific ( $\pm 70$  km) and Atlantic Ocean ( $\pm 55$  km). The plantation is close to the - still active - Turrialba Volcano, which is  $\pm 3,340$  meters high and located around 7km from the upper parts of the farm (close to 12 km from the lower parts). The farm starts just above 800m and continues to close to 1,400m, giving a  $\pm 600$ m elevation gradient to study. The coffee with what is considered the *best quality* is located  $>1,100$  meter. This makes up around 45% of the area, but is 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 regions, 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 (*figure 14*), in which areas in the upper regions often have values that can be associated with that of forest (e.g. McWilliam et al. 1993). During the period between 4 August 2014 and 8 September 2014, one hundred iButton Hygrochron (DS1923) temperature/humidity loggers have been placed throughout Aquiares. The lowest sensors in this study has been placed at 832 meters and the highest was placed at 1,399 meters (*figure 13*), giving a large elevation range to study. After 8 September, 86 of the 100 sensors could be recovered; the remaining sensors were either missing from shields or the complete shields could not be found.

#### *Sampling strategy*

Sampling was based on a hexagonal grid inside one of the micro-watersheds in the farm ( $n = 50$ ), while the location of the remaining 50 sensors was calculated by adding two *random* 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 or broken 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. Both the bad weather and large number of people working in the farm could have changed the location of the shields. During placement of the sensors, the bad weather made it difficult to reach all selected points, which led to an assessment of other suitable locations in the vicinity of the original point. The number of points in different areas (50 in the watershed and 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 13*. When collecting the data (which was done by a local farmer), it was discovered that 7 sensors 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 stable temperature and high humidity. The data analysis was done with 80 sensors.



**Figure 13:** Boundaries of the Aquiares estate and placement of the recovered sensors

#### 4.2.2 Data correction and validation

##### *Data correction*

The raw data that was provided by the sensors has been calibrated based on the change in time ( $\Delta t$  model - with parameters) explained in the previous chapter. This means that on every day, ten hours have been adjusted based on an anchor-point (around sunrise), assuming that the hourly change in temperature in the PVC tube could be multiplied by a certain factor to get the change in temperature that could be expected at an official weather station. As the temperature range after this calibration was still higher than would be expected in the study area, the values above the 0.98 quantile ( $27.9^{\circ}\text{C}$ ) were reset to this value. The values in the 0.99 quantile ( $28.84^{\circ}\text{C}$ ) and the maximum ( $32.9^{\circ}\text{C}$ ) were considered too high and likely the result of limitations of the method of data correction. The resulting range after this correction is  $13.4 - 27.9^{\circ}\text{C}$ , with a mean of  $19.9^{\circ}\text{C}$ . The mean elevation of the sensors in the study has been 1,110 meters, which is within 50 meters of the altitude of the weather station in the study area (1,040m + 26m tower-height). The mean temperature at this weather station is very similar ( $19.8^{\circ}\text{C}$ ) to the average temperature in the study area that was derived from the sensors. Relative humidity has not been analysed in detail in this chapter. Based on a weather station in the study area, a linear model was created to predict humidity

based on temperature. The data that is derived from this correction has a range of 59.7 - 100% RH, and a mean of 87.3% RH. This is very close to the range measured at the weather station, which has provided a range of 55.1 - 99.9% and mean of 87.7%. This shows that the strong correlation between temperature and relative humidity can be used to quite accurately predict the humidity. This strong correlation will also result in similar trends (although the opposite) of temperature and humidity with the studied covariates. For this reason, humidity was excluded from this chapter

#### *Data validation*

In order to assess the accuracy of the network and how this is influenced by factors such as the distance to its neighbours and elevation, the data first has to be interpolated in the spatial dimension. As explaining different forms of interpolation is the aim of the next chapter, the accuracy assessment in this chapter will focus on one of the most basic (deterministic) spatial interpolation approaches: Inverse Distance Weighting (IDW). IDW is covered in the *gstat* R-package and does not require any complex procedure (e.g. variogram fitting and model creation) to create the output. The inverse distance weighting power (IDP) is set to 2.0, which is the default value. The formula for IDW is:

$$\hat{v} = \frac{\sum_{i=1}^n \frac{1}{d_i^p} v_i}{\sum_{i=1}^n \frac{1}{d_i^p}}$$

In this formula,  $\hat{v}$  is the value that will be estimated,  $v_i$  is a known value,  $d$  is the distance from the data point  $n$  to the point that is estimated, and  $p$  is the IDP, which - by default - is set to 2. Low values for the IDP will result in smoother estimates between different points, as the distance has a relatively low importance in the calculation. To quantify the prediction error in this method of interpolation, we calculated the RMSE for the residuals of a leave-one-out cross-validation (cf. Kilibarda et al., 2014). The RMSE has been calculated for the one-month dataset and included all hourly measurements. The RMSE is the square root of the mean of the squared residuals (*see formula*). Residuals are the differences between the predicted and known values and hence have the same unit (degrees Celsius). In each round of cross-validation, we left one sensor out, predicting temperature by applying an interpolation of the remaining  $n-1$  sensors, after which the original temperature at the excluded sensor has been compared with the predicted value based on the  $n-1$  sensors. The *gstat* R-package has a built-in IDW function, as well as a function to assess the residuals through a cross-validation that includes a certain number (*n-fold*) of validation rounds, which in this study will be set as equal to the total number of sensors (80 in this chapter). In the formula to calculate the RMSE, which is provided below,  $n$  is the number of points (sensors),  $\hat{y}_t$  is the predicted value of the point, and  $y$  is the actual measured value of this point. The RMSE calculated in this chapter will provide the baseline for more advanced interpolation approaches in the next chapter.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_t - y)^2}$$

### 4.2.3 Analysis of covariates

#### *Geostatistical concept*

Interpolation of measured points to create coverage over a wider geo-network can be done by two approaches: *deterministic* and *geospatial*. Deterministic approaches - which include IDW and Radial Basis Functions - predict values based on the similarity between points, or by other methods that do not require analysis of the statistical properties. Geostatistical approaches, which include a wide range of Kriging functions, base their predictions on the statistical properties of the original dataset and possible related covariates about which information is known. Geostatistical analysis can be based on a wide range of models, from basic models that only include the original data, to models that include both space and time dimensions. Covariates can be included in most models to strengthen the predictive potential of the interpolation by adding more points at which some related variable is measured. As costs are expected to be an important limitation in setting up sensor-networks for agricultural purposes in developing countries, covariates are considered an important factor that can improve the predictions of models and hence reduce the number of required sensors in a certain study area. This chapter does not yet focus on the different geostatistical functions, but is mainly focused on the covariates and the correlation they have with the temperature data derived from the network. Geostatistical approaches that make use of covariates to improve estimations, which includes co-kriging, make use of the correlation and generally more frequent (or cheaper) observations of these variables. Strong correlations provide a good basis for geostatistical interpolation, and will be analysed with the Pearson's  $r$ . It is recommended that at least 50 points (preferable 300) are used in geostatistical interpolation for variogram fitting, to avoid the risk of over-fitting (Hengl, 2007).

#### *Correlation*

The correlation between temperature and a set of covariates has been calculated with the Pearson product-moment correlation coefficient (Pearson's  $r$ ). This coefficient provides a value that explains the degree of (linear) correlation between different variables, ranging from -1 (total negative correlation), through 0 (no correlation), to 1 (total positive correlation). A strong correlation between certain variables in this study is especially important in the selection of covariates that can be used in co-kriging of temperature data. While it can be assumed that a stronger correlation with a covariate will improve the predictions based on the same number of sensors, there is no universally applicable 'minimum correlation' from which kriging approaches that use covariates start performing better than ordinary kriging. In this study, the strength of (absolute) correlation is classified as 0-0.25 (*insignificant*), 0.25-0.5 (*small*), 0.5-0.75 (*medium*) and 0.75-1 (*strong*). Theoretically, with the addition of covariates in co-kriging, it is not possible to do worse than in ordinary kriging, as lack of cross-correlation will result in a fall-back on autocorrelation for the principle variable (ESRI, web). Other approaches, such as universal kriging (based on a linear regression) can perform worse than ordinary kriging when basing it on a covariate that has a limited correlation with the principle variable (cf. Goovaerts, 1999). In addition to the correlation of the selected covariates with temperature at different times, the correlation of these variables with the other predictors is also provided.

#### 4.2.4 Selected covariates

##### *Covariates*

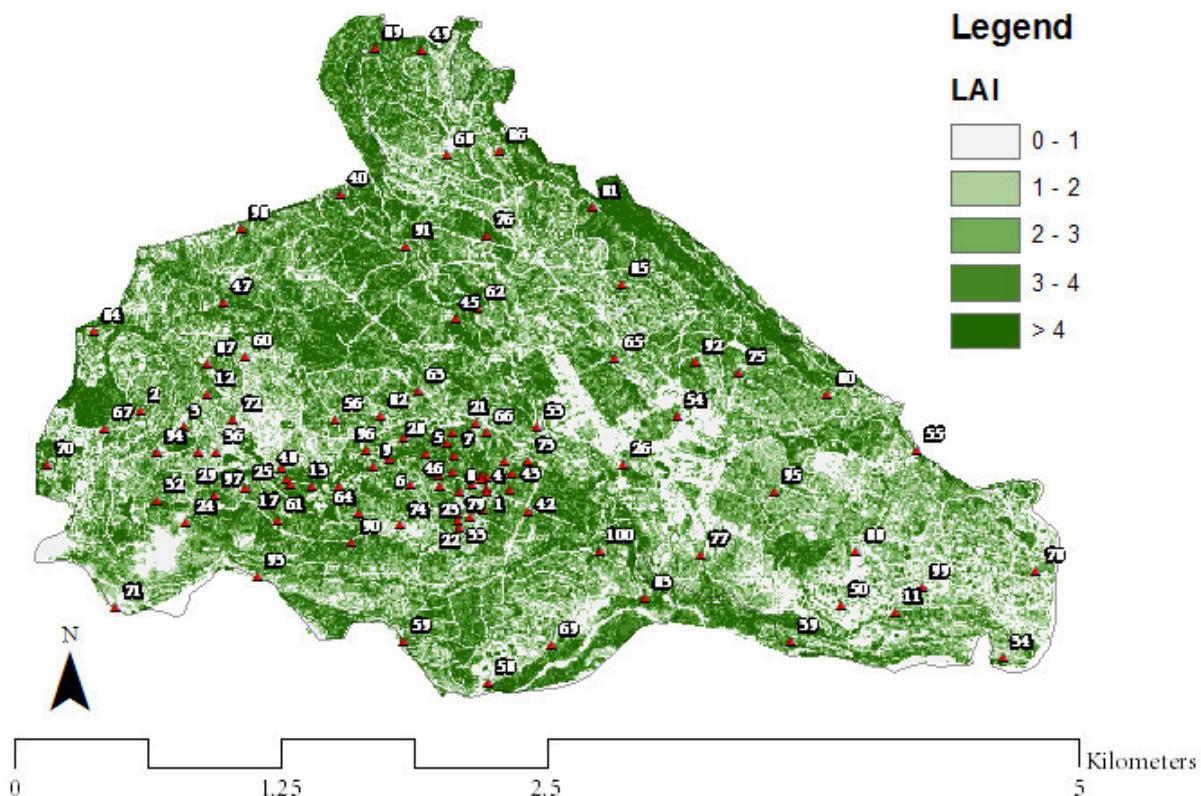
Covariates are variables that can be used to predict another variable: the dependent variable. The best known covariates that can be used to predict temperature is elevation, by taking into account the adiabatic lapse rate. The adiabatic lapse rate is '*the rate at which atmospheric temperature decreases with increasing altitude in conditions of thermal equilibrium*' (Oxford, web). While studies on the impact of certain variables on temperature have covered a wide range of potential covariates, an important issue will remain the cost to obtain recent geospatial data for this covariate. Many of the more basic covariates are freely available and show an increasingly better accuracy and resolution. A basic 90m Digital Elevation Model can be used to create a wide range of covariates. Static covariates that can be derived from a DEM include elevation, slope and aspect. With some additional calculations - such as calculations of the sun-position at a given day and hour - dynamic covariates can also be created without any costs. These covariates include the hourly changes in hill-shading and cast shadows, which together with an hourly solar constant, can be converted to hourly (direct and total) radiation, as well as total daily radiation estimates. Other covariates are generally more difficult to obtain. In this study, rasters of the leaf area index (LAI) and canopy height are used, as these were also freely available. Both can be expected to have some influence on temperature at a very small scale and are thus useful to include in studies that focus on collecting data for agricultural purposes in developing countries. Compared to canopy height (based on LiDAR images), information about the LAI is both more difficult to create and subject to faster (seasonal) changes.

##### *Elevation*

Information about the elevation of the sensors in this research is available from (1) GPS units used to store the coordinates of sensors, and (2) extraction of the sensor-coordinates in a DEM. A comparison between these approaches has resulted in very small differences, which means that the data from the DEM has been used in further analyses, as this is easier to implement in future models. The DEM has been derived from the SRTM 90m Digital Elevation Database, and has been cropped to the study area and disaggregated by a factor of 10. Incorporation of elevation in temperature interpolation is very common in GIS. Dodson and Marks (1997), for example, have tested three different lapse rates in a large mountainous region, while the WorldClim dataset has used a single lapse rate ( $6^{\circ}\text{C}/\text{km}$ ) as variable in the interpolation of weather station data (Seguinot et al., 2014). Based on the existing scientific understanding of the adiabatic lapse rate, together with the fact that the measurements have taken place during the rainy season (*the wet lapse rate is lower than the average lapse rate*) over a height difference of  $\pm 600$  meters, it can be expected that the mean temperature in the lower areas of the farm is between  $2.7$  and  $3.2^{\circ}\text{C}$  higher than in the upper areas. This is based on both the average adiabatic lapse rate ( $6.4^{\circ}\text{C}/\text{km}$ ) and wet lapse rate (for 100% RH:  $5.5^{\circ}\text{C}/\text{km}$ ). Two factors might have a smoothing impact on this difference; these are (a) the relationship between elevation and solar UV radiation (Blumthaler et al. 1997); and (b) the expected denser (more natural) vegetation at higher regions of the study area, which could increase shading and protection from wind.

### *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 index is dynamic and changes depending on season and age/management practices of the crops (Welles & Cohen 1996). For coffee in the study area, the LAI has been measured at 4.0 ( $\pm 2.9$ ) in 1982 (Ewel et al. 1982), although parts of the farm are replanted every few years. More recent studies on the LAI in Aquiares have found that the LAI varies from 2.4 - 4.4 depending on the season (Taugourdeau et al. 2014); this study also mentions that inter-annual variations can be caused by pruning and plot renovation. The currently grown variety *Caturra* is characterized by relatively limited vertical growth (<2m) and is planted in rows that are 2m apart; other trees are also planted in the farm (Dauzat et al. 2001). Above and inside canopy climate and its relationship to the LAI has been extensively studied by Goudriaan (1977), and includes impacts of canopy geometry, incoming radiation, and optical properties. The map of LAI that has been used is based on High Resolution Multi-spectral Images (MODIS) together with field verification. As sensors have been placed adjacent to coffee trees at 1.5 meters (limiting shading), the main impact on temperature would come from larger trees. It was expected that a high LAI in the study area is correlated to canopy height and negatively correlated to temperature, as it can provide shelter from wind and can reflect incoming radiation (Sellers 1985).

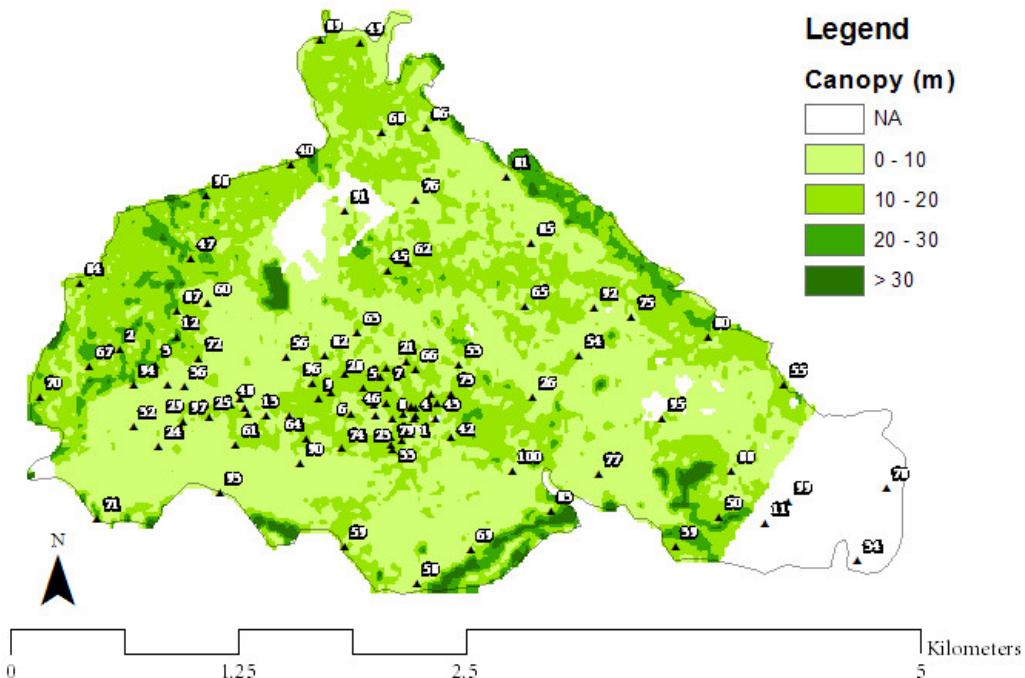


**Figure 14:** Leaf Area Index in the Aquiares estate

### *Canopy height*

The height of vegetation is another factor that can have an impact on the climate in an area, which might partly overlap with the leaf area index, as more natural areas will likely have both 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. As this data includes the absolute minimum and maximum in different cells, the values can exaggerate the canopy height, but will still show a similar trend as would be found if the median height would be assessed. The used dataset is a 2005 Costa Rica dataset (Blair et al. 2006) that covers the area around the Turrialba volcano.

The resulting raster is shown in *figure 15*, and shows that vegetation is highest in areas around rivers (borders) and the highlands. This is likely 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 canopy height in the farm is 11 meters with a maximum of 77 meters. The vegetation is generally higher close to boundaries of the farm, which could be caused by the steepness of the terrain in these areas, or management strategy. This can create a natural farm limit, with as additional benefit that it can be considered protected area for certifications (e.g. Rainforest Alliance). It has to be noted that measurement errors can play a large role, as the map does not only show the canopy height, but also the complexity of the terrain - as the height difference within the collection size is included in the calculated *canopy* height.



**Figure 15:** Canopy height in the Aquiares estate based on LiDAR data

### *Sun-related covariates*

The sun can have an influence on the temperature in different ways; the position of the sun during the day can result in areas 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, which is *the moment when the sun crosses the meridian at the selected site*. For the mean day during the studied period (August 22<sup>nd</sup> 2014), the solar noon was at 11.38AM local time (NOAA n.d.). Solar radiation can be related to daily minimum/maximum temperature (Bristow & Campbell 1984) and to productivity in tropical ecosystems (Monteith 1972). Modelling the relationship between solar radiation and temperature requires different empirical coefficients and for that reason is different to implement in general models. Calculations of the solar radiation at a certain moment at a given latitude is relatively straightforward and could be combined with DEMs to model the insolation and shadow during different times of day. Calculating the hill-shade is a well-known functionality in most GIS software, but is generally used for aesthetic improvements of the maps. The actual meaning of the values is not often discussed, but will have significant implications for the amount of insolation at a given location. The output of the algorithm generally provides values in the range from 0 to 255, which 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. Hill-shade can be calculated with the formula below, in which  $s$  is the slope,  $z$  the zenith angle,  $az$  the azimuth angle, and  $as$  the aspect (all in radians). In a normalized calculation, values  $<0$  are reset to 0, and all other values are multiplied by 255. Normalizing is not required in this study.

$$HillShade = (\cos(s) \times \cos(z)) + (\sin(s) \times \sin(z) \times \cos(az - as))$$

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 factor at this level, but even at this scale there will be places where the cast shadows will cause some differences throughout the day. These places can be expected to be 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 at a given moment. Cast shadows have been calculated at 10-minute intervals throughout a day, taking into account objects within a ten kilometer buffer of the study area to include all the larger mountains and volcanos in the region. The insolation at a given moment can be calculated with the formula below, in which  $R_{dir}$  is the direct radiation,  $R_{dif}$  is the diffuse radiation (both  $\text{kW/m}^2$ ),  $HS$  is the non-normalized hill-shade value,  $CS$  is the cast-shadow value (in both 0 is shadow, 1 is full sun), and  $\Delta t$  is the covered time in minutes. The calculation requires input from a complex algorithm that calculates the solar position. An R-script, provided online by Josh O'Brien, has been used.

$$Insolation = ((R_{dir} \times HS \times CS) + R_{dif}) \times 60 \times \Delta t$$

#### 4.2.5 Analyses

##### *Analysis 1: Network accuracy*

The first analysis has assessed the initial network accuracy during the one month of observations. Inverse Distance Weighted interpolation was used to interpolate the data of  $n-1$  sensors at all 744 hours and the overall RMSE was calculated. This has shown how accurate the network can predict values in a complex terrain, and how errors change between the night and day. The RMSE is analysed in its trends in space and time. The variation over time has focused on the intra- and inter-day variation of errors, and includes differences between *day* and *night* and also between *hot* and *cool* days. The variation of the accuracy in space has focused on the relationship of the RMSE with issues such as distance to neighbouring sensors, and distance to the border of the study area.

##### *Analysis 2: Correlation between covariates*

A second analysis in this chapter has focused on the correlation between the different independent covariates. The correlation between elevation, slope, aspect, leaf area index, canopy height and total daily insolation is provided in a table. This shows how several of the selected predicting variables are correlated, which is important to understand before applying geostatistical analyses.

##### *Analyses 3: Average trends with covariates*

The correlation of the six static covariates with temperature at all 744 hours was calculated, and the average Pearson's  $r$  is discussed. Attention is given to the variation between days to see how the average daily correlation fluctuates in the rainy season. As the daily solar radiation can be expected to only play some sort of role on days with sun, the average daily correlation might give an indication of the different types of days during the studied period. The output of this analysis is a table of the mean and range of the Pearson's  $r$  for the different covariates. Based on the static covariates, a linear model has been created that can (partly) explain the variation in temperature. The relative importance of the different covariates in this model was assessed with the LMG method ( $R^2$  partitioned by averaging over orders). This function is part of the *relaimpo* R-package (Groemping. 2013)

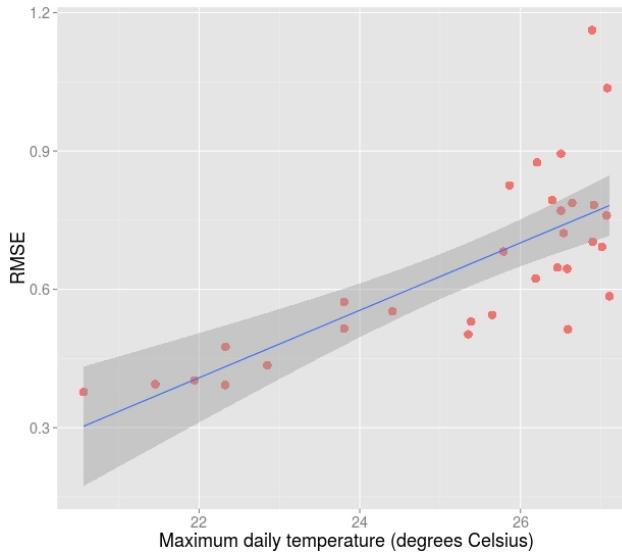
##### *Analyses 4: Daily trends in correlation*

The final analysis of this chapter has included both the static covariates and the three dynamic covariates (hourly radiation, hillshade and cast shadow). The daily trends (mean and 1<sup>st</sup>/3<sup>rd</sup> quantiles) in correlation between temperature and the covariates are provided as graphs to show the daily trends. The hourly values are derived after calculating the correlation at all 744 hours and subsequently aggregating these in hourly values. As final output of this chapter, hourly regression models have been created. The average strengths of these hourly models are provided, as well as the relative importance of different covariates in these models. The trends in model strength and covariate importance are provided as figures, and form the basis of the chapter on geostatistical interpolation.

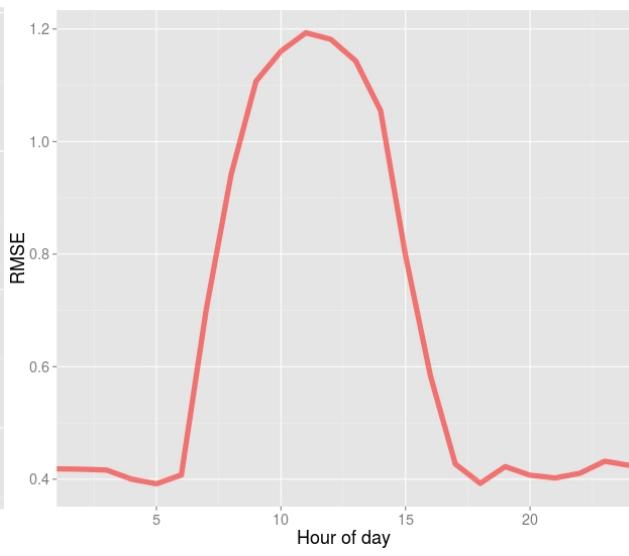
## 4.3 Results

### 4.3.1 Network accuracy

The average spatial RMSE, based on all 80 sensors and 31 days of hourly observations (IDW-interpolation) is  $0.65^{\circ}\text{C}$  with a range of  $0.18\text{-}2.62^{\circ}\text{C}$ . The mean daily RMSE stays within the same range during the month of observations and does not show any trends that would indicate that the network requires a period to ‘stabilize’. There is a clear relationship between the daily (min, mean and maximum) temperature and the RMSE. The minimum temperature has a  $r$  of -0.12, the mean temperature has a  $r$  of 0.67, while the maximum temperature has a  $r$  of 0.74 with the RMSE. The daily maximum temperature (based on the mean of all 80 sensors at any hour) and RMSE are plotted in *figure 16*. Errors show clear trends during the day (*figure 17*), and show a RMSE  $>1.0$  between 8AM and 2PM. The largest errors can be found between 11AM and noon, and can reach  $1.2^{\circ}\text{C}$ .

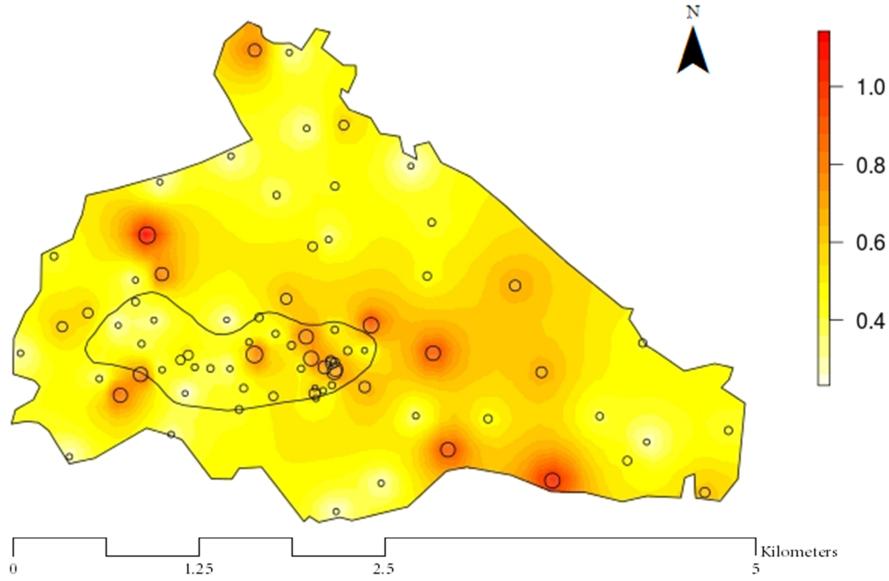


**Figure 16:** Daily  $\text{Ta}_{\max}$  vs RMSE



**Figure 17:** Daily trends in average RMSE

While the mean error of all 80 sensors at a certain moment shows a clear correlation to the type and time of day, differences between sensors show limited trends. Correlation of the errors for each sensor shows the strongest correlation with elevation ( $r = -0.19$ ) and vegetation ( $r = 0.21$ ). The RMSE in the micro-watershed, where 50 sensors were placed (35 were recovered and used) relatively close to each-other is  $0.50^{\circ}\text{C}$ , which is similar to the error for the entire farm (45 were recovered). The average error for sensors placed outside the watershed was  $0.52^{\circ}\text{C}$ . The RMSE shows an insignificant (negative) trend with the distance of each sensor to their closest three neighbours. The distance to the border of the area is also an insignificant factor for the RMSE of each sensor, which shows a small positive trend (further from the border is larger error). The average RMSE for each sensor during the month of observations is shown in *figure 18*. This map shows that the RMSE has no clear trends based on their location in the area and distance to each-other. The reason for the errors is unclear, and can include human activities as sensors placed close to the village show large errors.



**Figure 18:** Average RMSE of the 80 recovered sensors in Aquiares

#### 4.3.2 Correlation between covariates

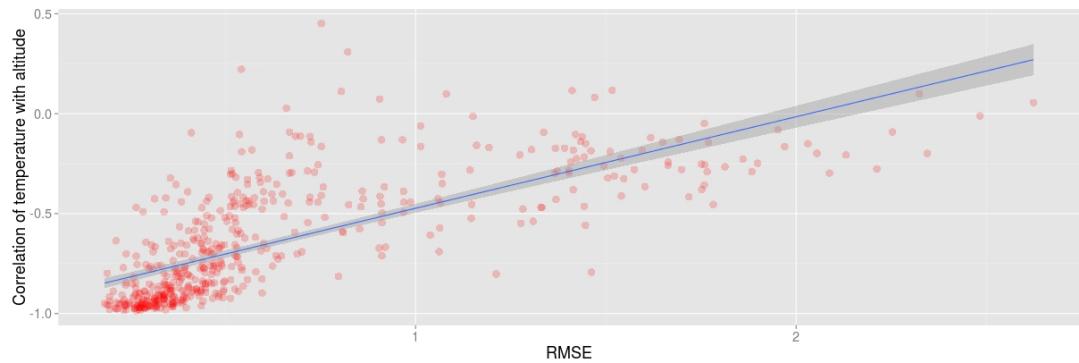
The correlation between the different covariates is provided in *table 9*. The mean temperature for each sensor is also included for a preliminary analysis. The correlation of the mean temperature with the other covariates shows that there is a strong negative correlation with the elevation, while there is a small negative correlation with canopy height. All other covariates have an insignificant ( $r < 0.25$ ) correlation with the mean temperature. Elevation, from which the covariates *slope*, *aspect*, and *daily radiation* are derived, shows a low positive correlation with slope and canopy height. Correlation of the slope is strong (negative) with daily radiation, which can be explained by the fact that these surfaces are tilted and receive less insolation during the periods with the highest sun-position (altitude). The aspect also shows a negative correlation with the daily radiation, which can be a result of both datasets being based on the same original data-source (DEM). The leaf area index has no significant correlation with the other covariates; the strongest correlation is -0.2 and is with the elevation. The only two covariates - out of the four that are created from the DEM - which are strongly correlated, are the total daily radiation and the slope.

**Table 9:** Correlation between the selected (static) covariates

	<i>Mean temperature</i>	<i>Elevation (meter)</i>	<i>Slope (radians)</i>	<i>Aspect (radians)</i>	<i>Leaf area index</i>	<i>Canopy height</i>	<i>Daily radiation</i>
<i>Mean temperature</i>	1.00	-0.81	-0.09	-0.03	0.07	-0.29	0.06
<i>Elevation (meter)</i>	-0.81	1.00	0.25	0.00	-0.20	0.30	-0.15
<i>Slope (radians)</i>	-0.09	0.25	1.00	0.09	0.01	0.00	-0.84
<i>Aspect (radians)</i>	-0.03	0.00	0.09	1.00	0.01	-0.12	-0.53
<i>Leaf area index</i>	0.07	-0.20	0.01	0.01	1.00	0.01	-0.04
<i>Canopy height</i>	-0.29	0.30	0.00	-0.12	0.01	1.00	0.00
<i>Daily radiation</i>	0.06	-0.15	-0.84	-0.53	-0.04	0.00	1.00

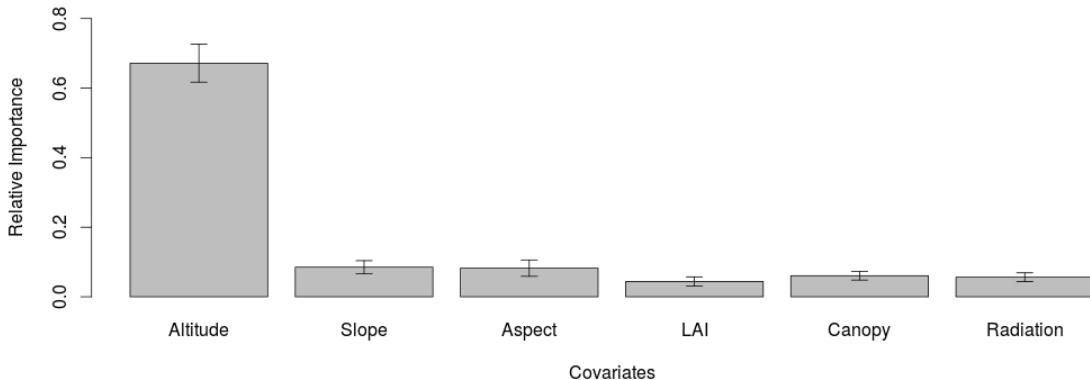
### 4.3.3 Average correlation with covariates

The correlation of *elevation* with temperature, when taking the mean of all 744 hours, is reduced to  $r = -0.64$  (-0.98 to 0.45). The daily trends differ a lot over the days, but there is a clear correlation between the hourly RMSE and the correlation of elevation with temperature at that hour ( $r = 0.73$ ). This shows that hours during which correlation of temperature with elevation is strong (generally periods with less sun), have smaller errors (figure 19). The mean correlation of slope and aspect with temperature is insignificant: *slope*: -0.02 (-0.32 to 0.47) and *aspect*: 0.04 (-0.38 to 0.46), and at no point reaches a medium ( $r > 0.5$ ) strength. *Leaf area index* shows a mean correlation of -0.03 with temperature, which stays in a -0.28 to 0.19 range. The strength of the correlation with the *canopy height* is much stronger, with a mean  $r$  of -0.18 (-0.38 to 0.29), but there is quite some overlap with elevation (higher areas have more natural area). Total (potential) *daily radiation* shows a mean  $r$  of -0.05 with temperature and has a range of -0.32 to 0.19. Correlation with radiation can only be expected to be relevant during hours with sun, and is covered in section 4.3.4.



**Figure 19:** RMSE vs. hourly correlation of Ta with elevation

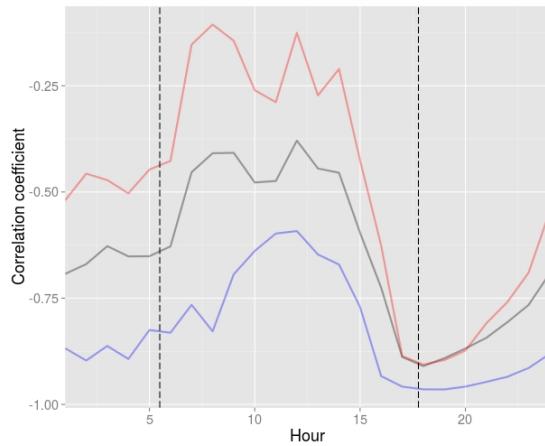
Adding all static covariates in a linear model to explain temperature at each hour result in models with a mean  $R^2$  of 0.52 (0.03 - 0.97), in which elevation is the most important variable (figure 20).



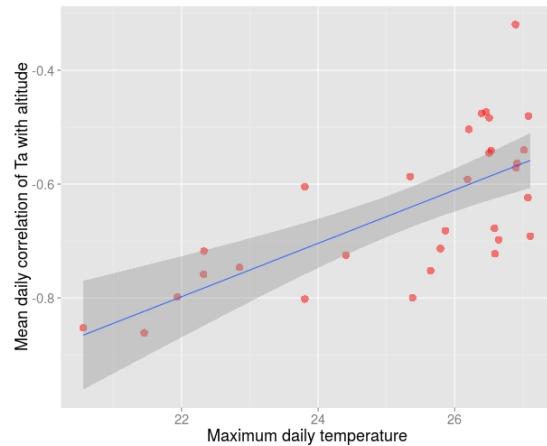
**Figure 20:** Relative importance of different covariates in explaining temperature

#### 4.3.4 Daily trends in correlation

Based on the earlier explained results, errors are largest during the day - and especially during the period around solar noon. There is also a clear relationship between the errors and the strength of correlation of elevation with temperature, which indicates that correlation varies over the day. The daily trends of correlation (mean and 1<sup>st</sup>/3<sup>rd</sup> quantiles) of *elevation* with temperature are shown in *figure 21*. This shows that the mean correlation is only strong during part of the day, while it is medium for most of the hours without sun. The period with sun, plotted between the dashed lines, shows a low correlation during the period around noon. The lines of the 1<sup>st</sup> and 3<sup>rd</sup> quantile show that similar trends can be found on most days. The mean daily strength of correlation of elevation with temperature is linearly related to the daily maximum temperature ( $R^2 = 0.46$ ) - *figure 22*.

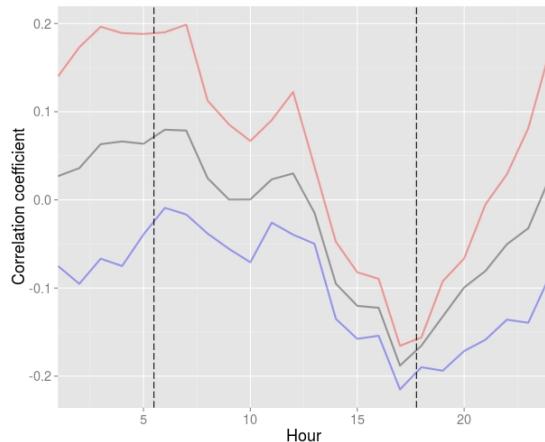


**Figure 21:** Temperature/elevation correlation

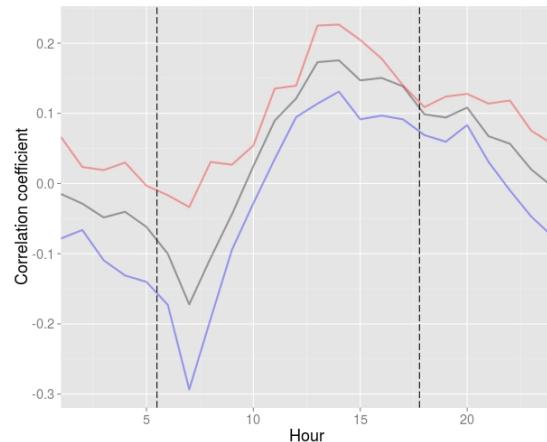


**Figure 22:** Max Ta vs. Ta/elevation correlation

The correlation of the other static covariates that are derived from the same DEM, *slope* and *aspect*, is plotted in *figures 23* and *24*. There are clear trends in correlation, but the mean strength is never stronger than 0.2. Mean correlation around noon is very weak ( $\pm 0.1$ ) for both variables.

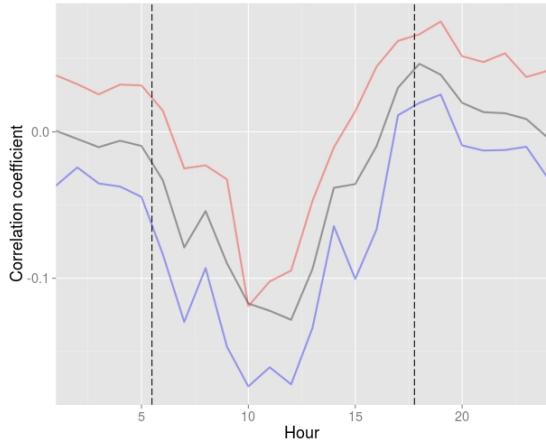


**Figure 23:** Temperature/slope correlation

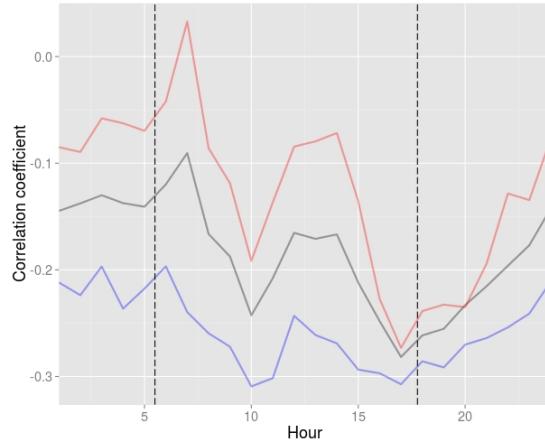


**Figure 24:** Temperature/aspect correlation

The correlation of two vegetation-related covariates - *leaf area index* and *canopy height* - with temperature are plotted in *figures 25* and *26*. The strength of the correlation of LAI with temperature is low, but clearly shows that there is a trend towards solar noon. Around solar noon, a high leaf area index has a (low) negative impact on the temperature, while the correlation is close to zero during hours without sun. The strength of the correlation of canopy height with temperature is stronger than that of LAI with temperature. The  $r$  is  $<0$  during most hours, and becomes stronger (negative) from sunrise to sunset. The strongest value is found around sunset and has a  $r$  of -0.28. Similar to LAI, the correlation is stronger during the period with sun than during the period without sun.



**Figure 25:** *Temperature/LAI* correlation

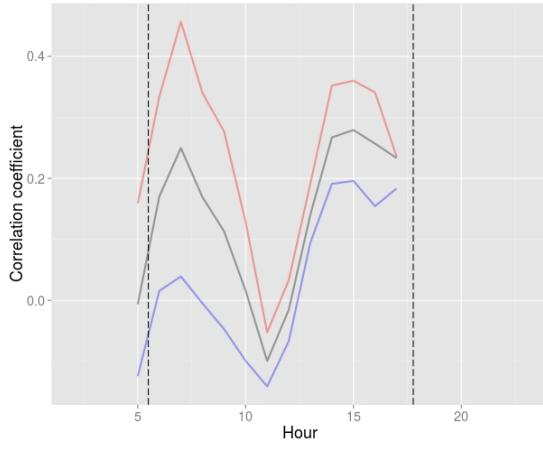


**Figure 26:** *Temperature/canopy* correlation

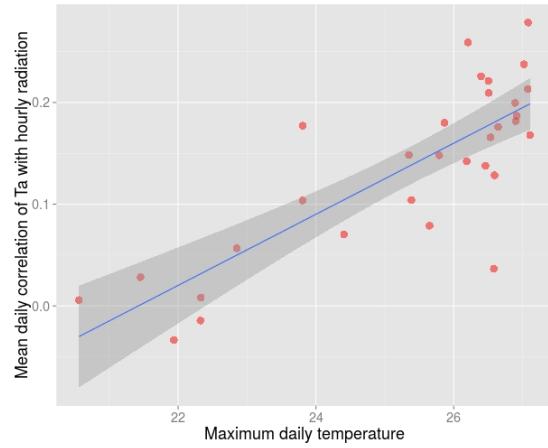
Correlation of temperature with the variables that can be derived from the position of the sun can be calculated for all hours with sun, and takes into account the changing position of the sun. These calculations can be done for the hill-shading, cast shadows, and combined (total) radiation. Correlation with *cast-shadows* is only relevant for two hours (just after sunrise and before sunset), as these are the only moments with sufficient variation between the data to assess correlation. At all other moments the value for the sensors - not necessarily for the entire area - is either 0 (no sun) or 1 (full sun). Correlation during these two hours is -0.05 (-0.09 to -0.03) in the morning and 0.18 (0.15 - 0.22) in the afternoon. This indicates that cast-shadows alone do not explain a lot of variation in this area. *Hill-shading*, without adjustment for radiation and cast-shadows, shows a correlation with temperature of 0.25 at 7AM, after which it reduces to -0.10 at 11AM. After this, it becomes positive again for most of the afternoon, with a mean correlation of  $r = 0.26$  between 2 and 5PM.

Calculating *hourly radiation* is done by multiplying hill-shade by cast-shadows and subsequently multiplying this by a radiation value. As the radiation value depends on time and is the same for all sensors at a certain moment, this multiplication will not change the hourly correlation. The correlation of hourly radiation with temperature shows a similar trend as that of hill-shade with temperature. The difference is that the trend starts one hour later (6AM), as the sun is not yet up at 5AM. A benefit that the hourly radiation calculation has over hill-shade alone is a stronger correlation between 4-5PM ( $r = 0.28$ ), which is marginally stronger than for hill-shade ( $r = 0.25$ ).

The daily variation in correlation of hourly radiation with temperature is provided in *figure 27*. The mean daily correlation ranges between -0.03 and 0.28 (mean  $r = 0.14$ ) and shows a clear linear relation ( $R^2 = 0.64$ ) with daily maximum temperature (*figure 28*). Based on trends in *figures 22* and *28*, the decreased correlation of temperature with elevation can partly be compensated by the increased correlation of temperature with radiation on days with high temperatures. The mean daily *temperature/elevation* correlation has a  $r$  of 0.77 with the *temperature/hourly radiation* correlation<sup>2</sup>.

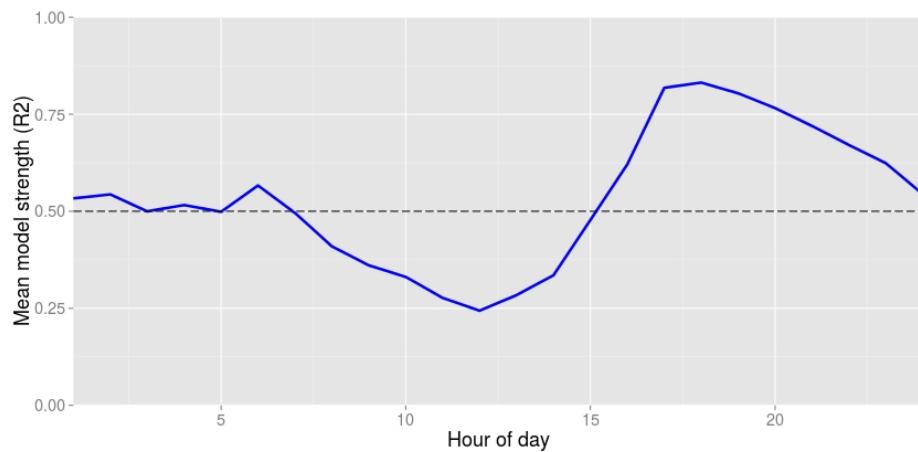


**Figure 27:** *Temperature/radiationboxiti* correlation



**Figure 28:** *Max Ta vs. mean radiation correlation*

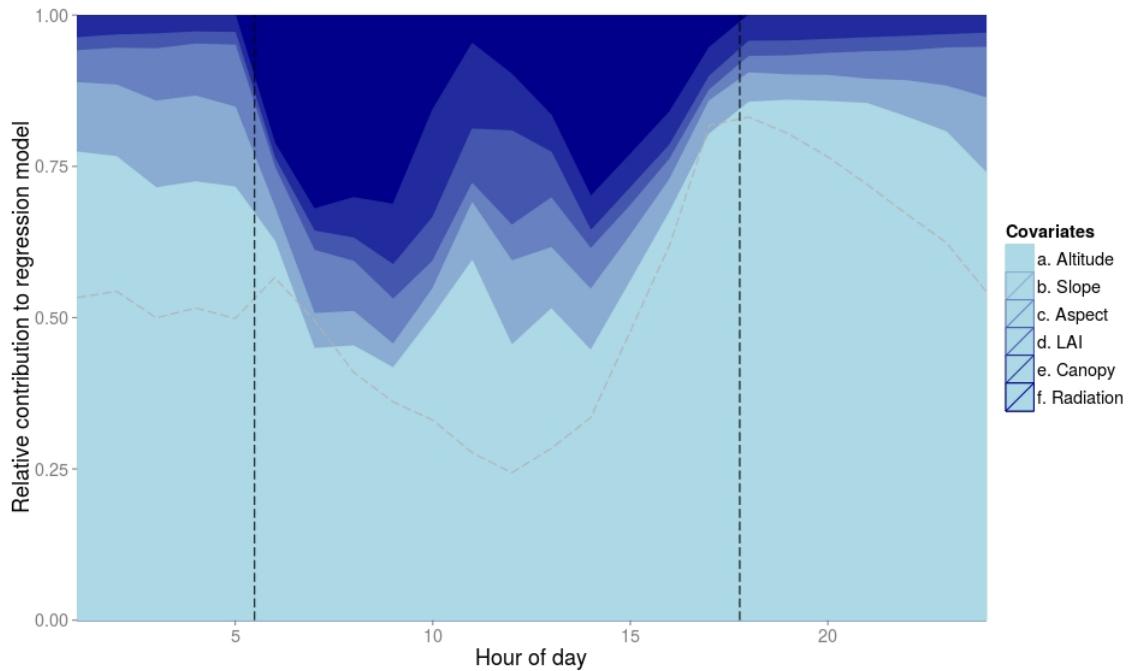
Creating a linear model at each hour, with the covariates discussed in this chapter (only including hourly radiation and not hill/cast shading), results in different strength of models depending on the time of day. The average strength of the regression model has an  $R^2$  of 0.53, which means just over half the total daily variance (dashed line) in temperature can be explained with the considered covariates. The trends in mean model strength throughout the day is plotted in *figure 29*.



**Figure 29:** Mean hourly linear regression strength ( $R^2$ ) that explains temperature variation

<sup>2</sup>The temperature/elevation correlation is negative, but is weaker when temperature/hourly radiation correlation is stronger.

The relative importance of the different covariates during the day is plotted in *figure 30*, which includes the mean strength of the resulting model for comparison (dashed line). This figure shows that *elevation* is the most important factor in explaining temperature throughout the day. During hours with sun, the regression model is strongest, and except for elevation, the *slope* and *aspect* play a relatively large role. The exact contribution and model strength at different hours is provided in *Appendix G*. After the sun is up, the relative importance of *hourly radiation* increases rapidly and contributes more than 30% to the regression model between 7 and 9AM. The contribution of radiation reduces to 5-10% between 11AM and noon, when the position of the sun is highest. The mean strength of the regression model is below 50% between 8AM and 3PM. During this period, the relatively contribution of other covariate than elevation play an important role. *Canopy height* adds more than 10% to the model between 9 and 11AM, while the *leaf area index* contributes to 16% at noon. The mean model-strength is also lowest at that moment and can only explain 24% of the temperature variance. The strongest model that has been created at every hour has an  $R^2 > 0.9$  for most hours, except for 12-2PM. Especially the models at 1 and 2 PM are weak (max  $R^2 = 0.67$ ).



**Figure 30:** Contribution of different covariates to the best model explaining temperature variation

The most important covariates, on average, are elevation (67%), hourly radiation (10%) and slope (8%). During the hours with sun, hourly radiation even contributes an average of 20%, although the average model strength only has a  $R^2$  of 0.54 during this period. The three most important covariates at every hour always include elevation. During the period without sun, the second and third most important covariates are slope and aspect. During the day, radiation is the second most important covariate during ten of the twelve hours. Canopy height is included with the three most important covariates during three hours, and leaf area index is included at only one hour.

## 4.4 Discussion

Retrieving the hundred sensors that had been placed in the study area and uploading the data, showed that approximately 20% of the sensors could not be used in further analysis. This was either due to inability to find them (in some cases they were removed from the PVC shield) or due to incorrect data. The main reason for obvious incorrect data seems to have been contact of the sensors with water. High humidity or water intrusion (which is very likely as no water-sealing tape is added to the shields), results in very high ( $>100\%$ ) relative humidity and very stable temperature throughout the day. In some cases this happened throughout the studied period, while in other cases this only occurred at a later stage and for a shorter period. Faulty data throughout the growing season can result from the timing of sensor-placement. The first one-third of sensors has been placed on a very rainy day, which means the sensors were wet when entered in the shields, and in some cases mud or other material might have stuck to the sensors. Data correction of the remaining sensors, based on the experiments in *chapter 3*, resulted in data that could be used for further analysis.

The accuracy of the network, calculated with a leaving-one-out cross validation in IDW interpolation, shows clear trends during the day. The mean RMSE during the month of observations was  $0.65^{\circ}\text{C}$ , which can reach  $1.2^{\circ}\text{C}$  around noon. The RMSE is higher during days which have a high maximum temperature, which are also the days when correlation of temperature with elevation is limited. The spatial distribution of the RMSE per sensor does not show any clear trends. The RMSE in the micro-watershed (high sensor density) and total farm (lower sensor density) is similar, and factors such as average distance to the closest three sensors or the farm-boundary have no influence.

Correlation between the static covariates and the temperature showed that elevation is the most important factor. Models on average, can explain 52% of temperature variation during the studied period. The strength of correlation between elevation and temperature, which is related to the type of day (maximum temperature), is correlated to the mean RMSE on these days. This indicates that it is more difficult to accurately monitor the temperature on hot/sunny days. As the study has been conducted during the rainy season, it is expected that the accuracy can change in other season.

The hourly correlation between *elevation* and temperature is strongest around sunset and weakest around noon. The correlation of the *slope* and *aspect* shows clear trend, but remains relatively weak for most of the day. The correlation of *LAI* with temperature is strongest around noon, which is what was expected based on the position of the sun and shading the denser canopy can provide. The correlation with *canopy height* is not as strong during the period of solar noon, but is strongest around sunset - which can be a result of the correlation the canopy height has with elevation. Adding *cast shadow* to the calculations of *hourly radiation* causes a slight increase in correlation strength around sunset, but does not impact the period around sunrise. Correlation of *hourly radiation* is strongest in the early morning and the hours before sunset. This correlation is linked to maximum temperature, and thus negatively correlated with the temperature/elevation correlation. All hourly covariates combined can explain 53% of the variance in temperature, which is a minor ( $\pm 1\%$ ) improvement over only using static covariates. Radiation contributes 20% to models during hours with sun.

## 4.5 Recommendations

A range of recommendations can be made based on this chapter that analysed the field data. Based on the number of useful sensors, it can be recommended that future projects take into account a loss of 20%. Part of this loss of useful data is caused by problems with the shield, which can be reduced by ensuring that no water can reach the sensor. Tape could be added to places where PVC-tubes are attached to each other (e.g. the PVC elbows), and better placing of the sensor could reduce potential impacts of splash water from surfaces such as leafs. Sampling can be done in different ways, but it has to be taken into account that there is a risk of losing sensors, which means that creating a number of equal area strata - each with two sensors - could be a good way to ensure that the full area can be covered, while limiting the problems when losing a sensor. A second recommendation is that soil and humidity sensors could be placed in vicinity of the temperature sensors at certain strategic locations, in order to establish the relationship between these variables. This would enable the creation of maps of soil temperature and humidity in the area - if required. In case a local weather station is not available close to the study area, it would be very useful to have one certified weather station at a protected location, in order to understand what the *certified* temperature and humidity would be in the study area, in order to correct and calculate (soil) temperature and humidity.

The accuracy assessment provides a lot of information, but the required calculations take a long time (>24 hours) on a personal computer, which means that selection of a shorter period (e.g. 1 week instead of 1 month) could be a better alternative. A better understanding of the errors, both spatial and temporal, can increase performance of similar networks in the future. The current analysis shows no relation of errors to the density of sensors or the border of the study area, which indicates differences are either caused by local climatic changes (which could explain larger errors around the village), incorrect functioning of sensors, or the inadequate correction of the raw data. Understanding the relation between errors and the type of day can be used to improve interpolation. If the change in temperature during the day can be used to understand whether the day is sunny or cloudy (or possible other classifications), the relative importance of covariates can be adjusted in co-kriging or other kriging approaches. It is also recommended to consider use of the humidity sensor for the possible classification of days, as well as the use of satellite imagery that provides information about cloudiness. Improved understanding of the type (rainy, sunny, cloudy) day is thus recommended.

Hourly models that explain temperature and include dynamic covariates ( $R^2 = 0.53$ ) are currently only slightly better than models which include static covariates ( $R^2 = 0.52$ ), which could be improved by more studies of dynamic covariates. The linear models during the day are significantly weaker than the models during the night, which means other (combinations of) covariates have to be selected and studied. Several factors that influence local climates have not been added (e.g. distance to water bodies), as this was not relevant in this study. If a model would be created to cover a wider diversity of areas, including these factors is recommended. The relationship between the *elevation/temperature* and *radiation/temperature* correlation can be better studied, to understand how this can be used in interpolation of the data. More studies in different types of terrain (e.g. flat) are also recommended.

## 5 Data interpolation

### 5.1 Introduction

The previous chapters have provided information about the resolution and shielding of the sensors that have provided sufficiently accurate results, and have also provided the correlation that temperature has with a range of covariates during the studied period. The final step has been to create an accurate map of the temperature and humidity in the area, while determining how costs can be reduced in future studies. Whereas maps that provide average temperature and humidity data are relatively easy to create and show a strong correlation with factors such as altitude, the hourly data is more complex to assess based on few sensors, as correlation with covariates is weaker (*previous chapter*). In the original study, 80 sensors have been used in  $8.7 \text{ km}^2$ , which is around one sensor per  $0.1 \text{ km}^2$ . This density will be reduced in future studies, and especially in projects where the trials are conducted in relatively flat terrain with limited local variation. The predictive ability of interpolation models with covariates has the potential to increase the accuracy of the network, which can result in less sensors providing the same level of output. The effectiveness of applying different kriging approaches to assess climatic data (e.g. temperature, precipitation, evotranspiration) in areas with complex terrain has been well studied (Goovaerts, 2000; Martínez-Cob, 1996) and has resulted in some basic recommendations. Whether these recommendations are valid for the smaller scale, as studied in this research, has been assessed. One of the results, based on similar methodology, but a different scale, is that ordinary kriging outperforms advanced kriging (based on the relationship between rainfall and elevation) when the correlation is lower than 0.75. This is a strength that is only attained by the temperature-elevation correlation (depending on time/type of day) in this research.

The limitation of most traditional approaches to temperature downscaling, is that they ignore the topographic variation at the very local scale (Benestad, 2001), which will be important for the variation in complex terrain. Models that interpolate weather station data and cover a relatively small scale, include - *except for elevation* - complex topographic indexes (Daly, Conklin and Unsworth, 2007) and geographic position (Hijmans et al., 2005). The study that - in methodology - is most similar to this study is the study by Parmentier et al. (2014), which included elevation, aspect, canopy height, percentage forest cover (expected to be comparable to LAI) and land surface temperature (derived from MODIS) to predict daily maximum temperature. Holden et al. (2011), working at a similar scale as this study, have included humidity and solar radiation, in addition to elevation, as covariates to explain time-space variations in temperature at the local level. An interpolation study by Morales et al. (2007) used distance to water bodies, slope, aspect, elevation, geographic position (latitude and longitude) as well as an indicator of land condition (NOAA mean NDVI). Different covariates have been found to be strongly related to micro-climates (and diurnal climate variation); this includes LAI (Hardwick et al., 2015), solar radiation (Bennie et al., 2008) and distance to streams (Eskelson, 2011). This chapter has combined different covariates - the relative importance and model strength of which is shown in *figure 30* - in a range of geostatistical interpolation methods to assess how accurate the network remains when a certain number of sensors is removed from the full dataset.

## 5.2 Methodology

### 5.2.1 Data inputs

The data that has been analysed in this chapter is the same dataset that was used in *chapter 4*. Data is derived from 80 sensors during a period of over one month (5 August to 8 September 2014) in Costa Rica. The number of sensors has been reduced to 40, 20, and 8 sensors for some of the analyses. The 40 and 20 sensors were simply the first 40 and 20 sensors in the original dataset, as this still provided a good sampling. The 8 sensors have been manually selected to provide a good coverage. The data has been corrected during the hours with sun, due to the faster heating in the used PVC shields compared to certified sensor shields. The data above the 0.98 quantile has subsequently been corrected and reset to the value at the 0.98 quantile, in order to have data that is in line with what can be expected in the area based on a local weather station. In this analysis one week of observations is selected to calculate the accuracy (RMSE). The maps that are created and used to explain differences between the methods are based on the hourly/daily mean during this week. The impact of different kriging approaches was tested on all 24 hours, which means a total of 168 layers, each with 80, 40, 20 and 8 sensors, are interpolated. An important issue that has to be taken into account with different forms of interpolation is that missing values (*NA*) are not accepted. This means all missing values in the time-series had to be removed prior to this analysis, as otherwise the steps discussed in this chapter did not work. Removing missing values is straightforward and can be done with different forms of interpolation (covered in *chapter 3*). During the data correction that takes place after this step (e.g. removing extreme values), it is important to replace these values by other values (not by *NA*'s). The loading/correction of data is mainly done with the R *zoo* package.

### 5.2.2 Prediction grid

Based on literature (*section 5.1*) and the results in *section 4.3.4*, it was expected that elevation would be the most important covariate in this study. It has also been discussed earlier in this thesis that the desired resolution of the outputs in this study is at  $\pm 1$  hectare, linked to the level at which farmers are expected to make decisions regarding varieties and other inputs. The DEM that is used in this study, and is open-access (STRM 4.1), is created at 90m resolution. As all other covariates - except for LAI and canopy height - are derived from this map, the DEM has formed the basis of the prediction grid in this study. The resolution is sufficiently high for this study (and could be disaggregated at a later stage), and this would require fewest adjustment and limited additional calculations. For different types of interpolation, different prediction grids will be required. For ordinary kriging, the DEM can be converted to a '*SpatialPixels*' object (R *sp* package), which is a blank grid with the same cell-size. In order to be able to interpolate in time, different forms of data transformation have been done. The main step has been from *raster*, to '*SpatialPixels*' and to an object of class '*STF*', which is an object that can handle spatio-temporal data. This object does not only include a data-slot for the area, but also contains a data-slot that contains the measurements for a certain period. The prediction grid for universal kriging is also based on a '*SpatialPixels*' object,

but will be filled by a selection of covariates to create a ‘*SpatialPixelsDataFrame*’. The values for the pixels can also be adjusted during calculations, which means hourly radiation values can be included. This prediction grid contains 1,768 cells, which is a rectangular grid that covers the entire study area.

### 5.2.3 Geostatistical interpolation

A short introduction to interpolation has been provided in the previous section (*methodology*), as deterministic interpolation was already attempted for a basic assessment of network accuracy. Different interpolation methods have been assessed in this study, but it has been decided that kriging provides the best range of options, as it has been successfully applied in similar studies (*section 5.1*). Kriging is an important geostatistical technique that calculates values of a variable over a larger area, based on a certain sampling of this variable. The basic technique has been introduced by Matheron (1963) and has evolved from Ordinary Kriging to a wide number of more complex methods that take into account covariates and trends in both space and time dimensions. In its most basic form, kriging is very similar to IDW, as they both weigh the nearby points to predict the value at an unknown location (*formula below*). In this formula,  $Z(s_i)$  is the known value at location  $i$ ,  $\lambda$  is a certain unknown weight at this location,  $s_0$  is the location of prediction, and  $n$  the total measured values. The main difference is that kriging adds more information to the selected *weight* ( $\lambda_i$ ) than only the distance (ESRI, web). In ordinary kriging, this weight is also dependent on the spatial relationship of the known point and a fitted model of these points (variogram). Due to the complexity of the different selected kriging algorithms, the differences are only discussed - but not provided as algorithms.

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$$

Except for the main R packages that deal with spatial objects (*raster*, *rgdal*, *rgeos* and *sp*), a number of specialized geostastical packages has been used in the analyses. These are *gstat* (Pebesma, 2004), and *spacetime* (Pebesma, 2012). For automatic fitting of variograms, three additional packages have been used: *automap* (Huemstra, 2013) *randomForest* (Liaw & Wiener, 2002) and *GSIF* (Hengl, 2015). In order to avoid errors when using different kriging approaches, the projection of the spatial objects has been set to the WGS4 Web Mercator (Auxiliary Sphere) reference system (EPSG:3857), as this will transform the map units from degrees to meters, which is required for some calculations.

### 5.2.4 Selected kriging-approaches

Based on existing studies and the results provided earlier in this study, a small selection of kriging approaches has been analysed. Ordinary kriging, which is relatively similar to IDW in that they both only look at the temperature data and distance (although Ordinary Kriging includes spatial dependence), is taken as the control method. Universal kriging (UK), also known as regression kriging, combines a regression model (with a selection of covariates) with the basic kriging of residuals of the regression. UK will be analysed with one covariate (elevation) and a combination of covariates, based on *section 4.3.4*; this has included different covariates depending on the time of day (radiation

cannot be included when it shows no variation). Two automatized approaches of UK were tested. This includes automatic fitting of the variogram (included in the *automap* R package), and a machine-learning method to improve regression: Random Forest kriging. Co-kriging and spatio-temporal kriging were analysed with different covariates. More details are provided in *table 10*.

### 5.2.5 Analyses

The analyses in this chapter mainly focused at the overall accuracy of the different interpolation approaches, but has also covered some other relevant issues. The RMSE was analysed both in the spatial and temporal domain to indicate how this changes with different sets of sensors. Maps have been provided in situations where the errors shows unexpected trends. Special attention is given to the period around solar noon where, based on section 4.3.3, the covariates cannot explain more than half of the variance in temperature. All analyses include the full 80-sensor network, as well as subset of half (40), one-quarter (20) and one-tenth (8) of the sensors. At the original density, one sensor is placed per 0.1 km<sup>2</sup>, which reduces - in different steps - to one sensor per 1km<sup>2</sup>. The errors that will be included in the analyses are the (mean, spatial and temporal) root mean square error (RMSE), as well as the mean absolute error (MAE) and the mean absolute percentage error (MAPE). RMSE was already explained in *section 4.2.2*, and is the most used measure of accuracy. The difference between the MAE is that it does not include a squaring process, which will reduce the impact of outliers. The mean absolute percentage error, is a measure of accuracy that is provided as percentage and not in the same units as the original data. While this faces some important limitations (degrees Celsius are not in a 0-100 scale), this works quite well on the covered temperature range, which is always positive (15-30°C). Detailed analyses are provided in Appendices *H* through *K*.

The analyses has included all 13,440 (*7 (days)* × *24 (hours)* × *80 (sensors)*) residuals, in order to be able to provide spatial and temporal trends. The predictive power (covered in *sections 5.3.1 - 5.3.5*) is assessed by a leave-one-out (LOO) cross-validation. The overall RMSE has been provided, based on all residuals derived from the LOO, but attention is also given to differences in space and time. The spatial RMSE is based on the residuals at all hours, and provides the mean of the hourly RMSE for the 168 hours. The temporal RMSE is based on the errors for each sensor, and calculates the mean of the RMSE for all (80, 40, 20 or 8) sensors. The accuracy in predicting the mean temperature (model-fit) - that is covered in *section 5.3.6* - is assessed by calculating the RMSE of the difference (residuals) between the mean value of the sensors that are included in the interpolation and the actually observed values. This has indicated by how much values were modified in the kriging approaches. Discussion of the size of errors acknowledges the level of sensor accuracy ( $\pm 0.5^{\circ}\text{C}$ ) and possible errors that result from the interpolation of data from the original 20-minute resolution to regular one-hour interval. As the accuracy of the sensors is  $\pm 0.5^{\circ}\text{C}$ , an RMSE  $< 0.5^{\circ}\text{C}$  is considered *low*. Errors ranging from 0.5 to  $1^{\circ}\text{C}$  are still lower than the accuracy at which most crop thresholds are provided and are thus considered *medium*, while errors (RMSE)  $> 1^{\circ}\text{C}$  are *large* and should be avoided. Accuracy assessment does not only focus at the daily mean error, but gives special attention to the range during the day - which will be important to assess extremes.

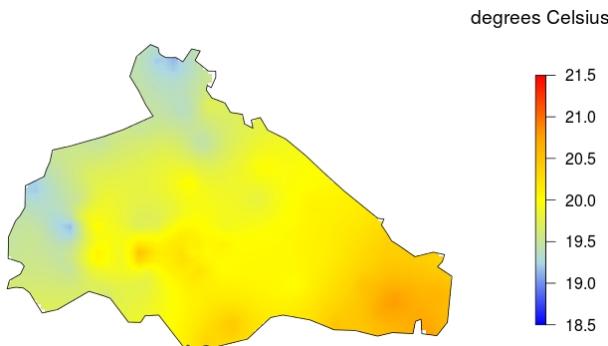
**Table 10:** Summary of selected interpolation approaches

Kriging approach	Section
<i>Ordinary kriging</i> is considered as anchor algorithm in the field of geostatistics, as it provides robust output under a wide range of conditions (Deutch and Journel in Atkinson & Lloyd, 1998). In this study, the variogram is fitted automatically - after an initial user estimation. No covariates are included in this control method. The only requirement is a spatial object that includes the location of sensors with the data ( <i>SpatialPixelsDataFrame</i> ), and a prediction grid ( <i>SpatialGrid</i> ) that covers the full geospatial network for which data will be predicted.	5.3.1
<i>Universal kriging</i> is largely similar to Ordinary Kriging, but also made use of covariates that are included in creation of the regression model. The covariates have been log-transformed as this will make the distribution more symmetric (normal distribution) and can help to create a better fitting semi-variogram model (Armstrong & Boufassa, 1988). In the first UK approach, only elevation is included - as this shows the strongest correlation and has resulted in good result in similar studies. A second approach has included the results of <i>section 4.3.3</i> and includes different covariates throughout the day. Radiation is included during the day; slope and aspect during the night.	5.3.2
<i>Automatized Universal Kriging (UK)</i> is mathematically similar to basic UK but includes aspects that make this approach easier to implement by reducing user input. One aspect that can cause difficulties for inexperienced users is variogram fitting. This has been done by the <i>automap</i> R package, to assess how this influenced the accuracy of the data and the time required (although time is not considered a serious restriction due to cloud computing options). Random-Forest kriging is a machine-learning method to spatial interpolation. The advantage of this approach is that it fits all the variograms automatically based on a set of covariates, and it not sensitive to noise or overfitting of the data (Grimm et al., 2008). This approach to kriging is used in several studies (Bradter et al., 2013; Peters et al., 2007; Wiesmeier et al., 2011).	5.3.3
<i>Co-kriging (CK)</i> is an interpolation approach that is often used when the <i>variable of interest</i> is too costly to measure, and measurements of a correlated variable are cheaper to make, or readily available (Fontin & Dale, 2005). Compared to other methods, in which a single variogram is created, different models are created for different quantiles in co-kriging. In CK, both autocorrelation for a certain variable of interest, as well as cross-correlation between this variable and covariates are used to improve prediction accuracy. Co-kriging was done with elevation and a set of covariates.	5.3.4
<i>Spatio-temporal kriging</i> can be done with most of the previously discussed forms of kriging, and can thus include covariates. Adding a temporal dimension to kriging is based on the assumption that temporally close observations have a stronger correlation than observations further apart in time. Different types of models exists, which are explained in detail in Pebesma (2012). Practical implementation is relatively similar to that of the other kriging approaches, and functions are mostly based in the <i>gstat</i> and <i>spacetime</i> packages. Autofitting of the variograms is also possible, although only one fitted variogram is used in the spatio-temporal kriging. Spatio-temporal interpolation was done with ordinary kriging, elevation, and a dynamic set of covariates.	5.3.5

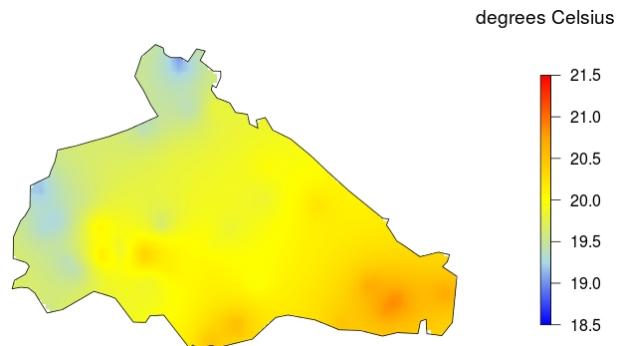
## 5.3 Results

### 5.3.1 Ordinary Kriging

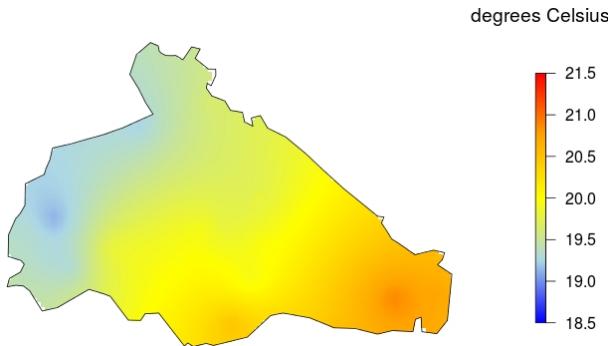
The overall RMSE for ordinary kriging with 80 sensors is  $0.83^{\circ}\text{C}$ , which becomes  $0.88^{\circ}\text{C}$  with 40 sensors,  $0.85^{\circ}\text{C}$  with 20 sensors and  $0.98^{\circ}\text{C}$  with 8 sensors. In the spatial domain the RMSE is  $0.69^{\circ}\text{C}$  (*Annex H-I*), while in the temporal domain the RMSE is  $0.83^{\circ}\text{C}$ . The mean temporal-RMSE does not change a lot when working with different numbers of sensors - as the number of hours remains the same. In the spatial domain, the trends in RMSE increases from  $0.69^{\circ}\text{C}$  with 80 sensors to  $0.95^{\circ}\text{C}$  with 8 sensors, although the range of the errors decreases when working with less sensors. The mean absolute percentage error increases from 2.5 to 3.3% in both the space and time dimension, when reducing the number of sensors from 80 to 8 sensors. The MAPE is always similar in both dimensions, and thus gives a good indication of the change in accuracy. While an increase of 0.8% in mean-absolute error does not seem large, this is an average, and can be larger during different hours. The visual impact of a reduction in sensors, without adding a covariate to improve prediction, is shown in *figures 31-34*; there is a clear decrease in complexity when reducing the number of sensors. With 8 sensors, the range remains similar, but the trends over the area become very smooth.



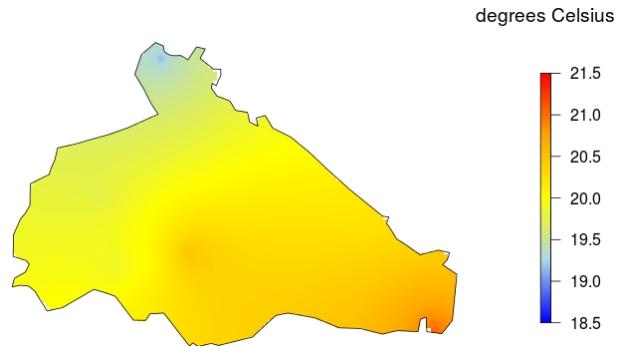
**Figure 31:** Ordinary Kriging (80 sensors)



**Figure 32:** Ordinary Kriging (40 sensors)



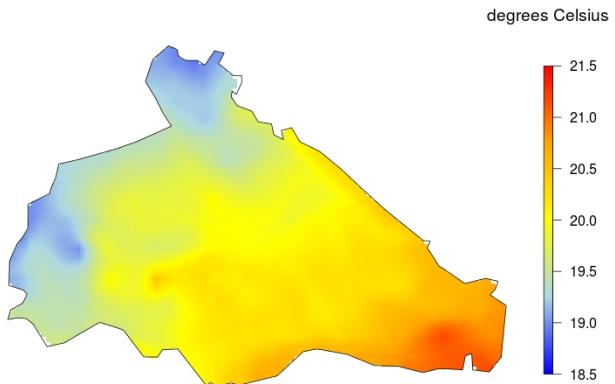
**Figure 33:** Ordinary Kriging (20 sensors)



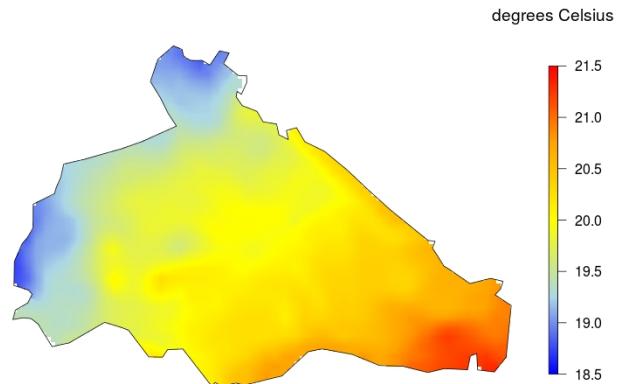
**Figure 34:** Ordinary Kriging (8 sensors)

### 5.3.2 Universal Kriging

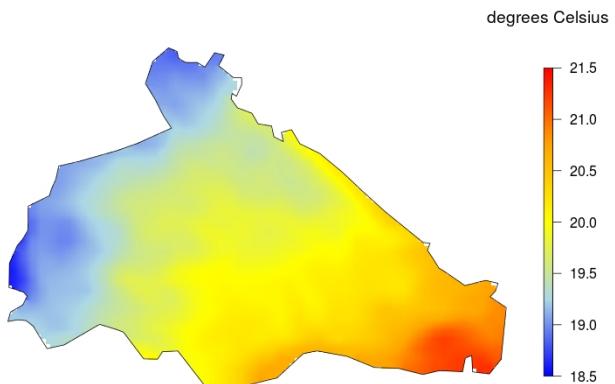
Two variations of universal kriging have been applied with elevation as covariate; the first was fully based on the basic functionality in the *gstat* package, while the second method also used automatic fitting of the variogram - without providing a sample (based on the *automap* package). Statistics of the basic universal kriging approach are provided in Annex H-II, and of automatized universal kriging in Annex H-III. Both approaches have a lower RMSE than ordinary kriging with the original dataset ( $0.79$  vs  $0.83^{\circ}\text{C}$ ). With all 80 sensors, automatized fitting provides the lowest errors of the three approaches. When reducing sensors, basic universal kriging starts performing better than automatized universal kriging, although differences remain very small. Universal kriging has a RMSE of  $0.79$  at 40 sensors,  $0.80$  at 20 sensors and  $0.88^{\circ}\text{C}$  with 8 sensors. Automatic kriging takes about twice as long for calculations as universal kriging with a single variogram as sample. For automatic kriging, the RMSE is  $0.83$  with 40 sensors,  $0.81$  with 20 sensors, and  $0.90^{\circ}\text{C}$  with 8 sensors. Errors with Universal Kriging are always lower than with Ordinary Kriging. Universal kriging - *including elevation* - shows less smoothing when reducing the number of sensors (*figures 35-38*). Differences are especially clear with the smallest (20 and 8) number of sensors. This can help in reducing the number of sensors, while still resulting in distinct local variations in temperature.



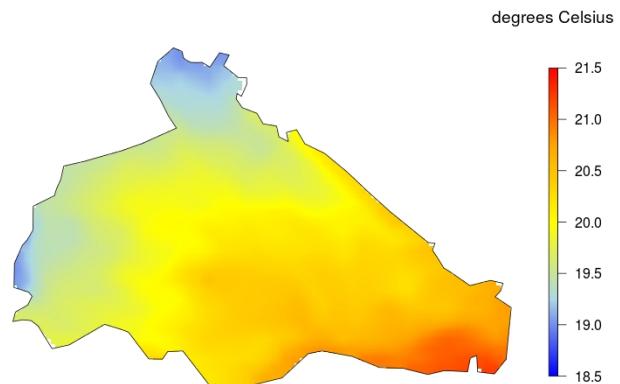
**Figure 35:** Universal Kriging (80 sensors)



**Figure 36:** Universal Kriging (40 sensors)

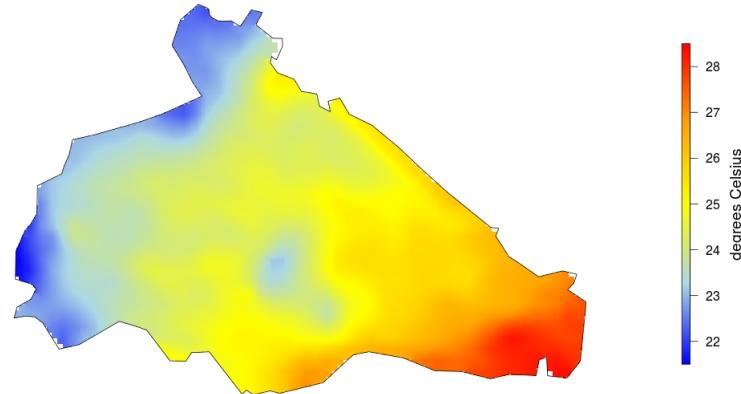


**Figure 37:** Universal Kriging (20 sensors)

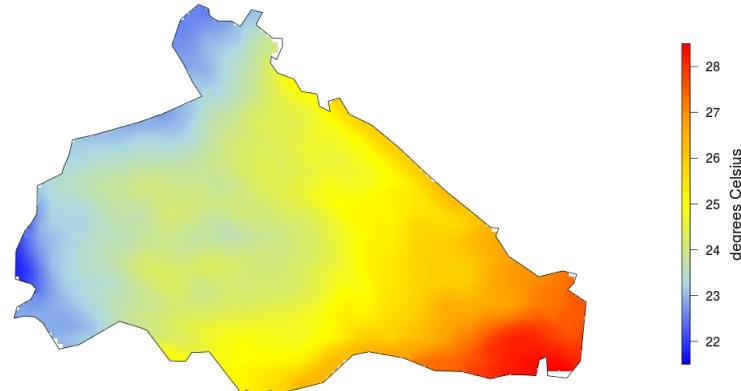


**Figure 38:** Universal Kriging (8 sensors)

Although the differences between Universal Kriging and automatized Universal Kriging are very small, the two approaches resulted in different maps at hours with sun (*figures 39 and 40*). These plots, which show the most problematic hour (11AM), indicate that - while the range and trends are similar - the local variation is different between the two approaches. Auto-kriging shows more smoothing in the middle parts of the farm. These differences are the result of automatic variogram-fitting; providing a single variogram for all hours can result in problems that result from a lack of spatial dependence. This is a problem that only occurs in non-automatized kriging and especially during the hours when correlation of elevation with temperature is not very strong (hours with high temperature). In these cases, the differences are relatively large between universal kriging and automatic kriging. Which of the two approaches will be preferred will depend on the reliability of sensors; more local variation can provide more detailed information about micro-climates, while it can also increase the impact of incorrect functioning sensors in the network. The main advantage of automatic kriging is the ease of implementation (no user-input/estimation is required).



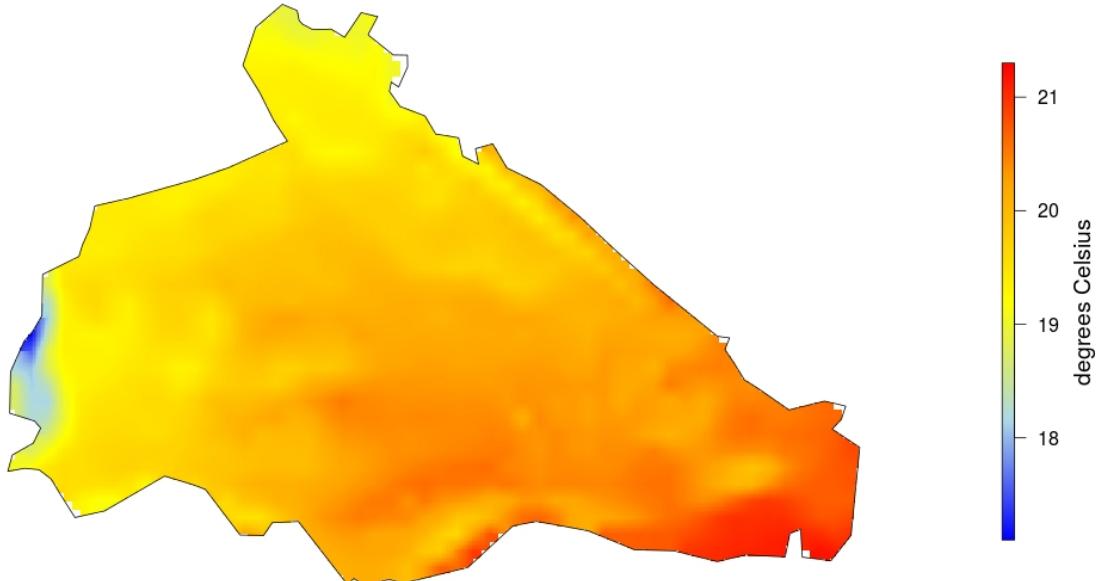
**Figure 39:** Temperature at 11AM on the first day of measurements with 20 sensors (universal kriging)



**Figure 40:** Temperature at 11AM on the first day of measurements with 20 sensors (automatized kriging)

### 5.3.3 Dynamic universal kriging

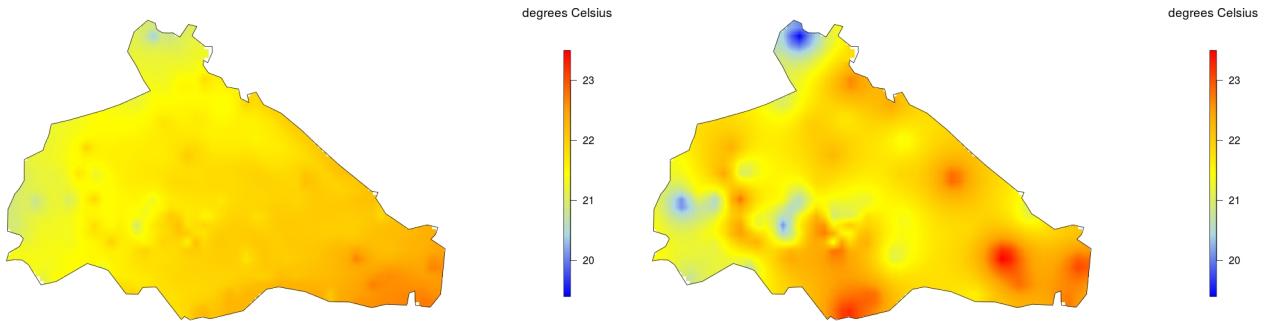
Dynamic kriging, which changes the set of covariates during the day (based on *section 4.3.3*) has been done with three different approaches. The approaches used in the previous section (basic universal kriging and auto-kriging) have been applied, as well as RandomForest kriging. The limitation of RandomForest kriging is that it requires at least 50 points for validation and does not continue with predictions once a certain hour cannot be predicted based on the covariates. The statistics that resulted from the different approaches are provided in Annex I. Basic dynamic kriging has the following RMSE for the different sensor-sets: 0.77 (80), 0.81 (40), 0.88 (20), and  $1.44^{\circ}\text{C}$  (8). For automatized dynamic kriging the errors are: 0.78 (80), 0.83 (40), 0.81 (20) and  $0.90^{\circ}\text{C}$  (8). For random forest kriging, the RMSE for all 80 sensors is  $0.78^{\circ}\text{C}$ . With the full 80-sensor dataset, errors are lowest for basic dynamic kriging. With a reduction in sensors, the errors become much larger, and especially with 8 sensors, the errors are problematic. With automatic dynamic kriging, the errors are almost the same as with elevation only. The errors, however, are significantly lower than for non-automatized dynamic kriging at 20 and 8 sensors. Random Forest kriging only works with the 80-sensor dataset and, while the RMSE at the full sensor-set is quite good, this method will be difficult to implement in smaller sensor-networks. The main difference between dynamic kriging - including hourly radiation - and universal kriging with elevation only, is the lower estimate of temperature at areas where cast shadow plays a role. This is especially problematic in areas with few sensors (left part and the areas close to borders in *figure 41*). This can be expected to be caused by overfitting, which is the result of using a very complex model to fit a small set of samples. The overall errors remain similar at the locations of sensors, but at areas where sensors are not found, prediction results in temperatures that cannot be validated, and might be unrealistic and a result of over-estimation of the impact of hourly radiation.



**Figure 41:** Predicted mean temperature based on dynamic kriging and 8 sensors

### 5.3.4 Co-kriging approaches

Co-kriging has been done with (1) elevation, (2) elevation and radiation, and (3) elevation, radiation slope and aspect. The RMSE for co-kriging with elevation with different sets of sensors is 0.86 (80), 0.90 (40), 0.91 (20) and  $1.21^{\circ}\text{C}$  (8). Including radiation results in RMSE's of 0.83 (80), 0.87 (40), 0.88 (20) and  $1.21^{\circ}\text{C}$  (8). Adding slope and aspect does not further reduce errors. This shows that adding covariates improves the accuracy at 80, 40 and 20 sensors, while errors remain similar with 8 sensors (*Annex J*). The benefits of co-kriging are limited, as errors are in all cases larger than for OK. Co-kriging could be improved with more advanced adjustment (parameter calibration), as in this study the same (limited) parameters are used for the different interpolation approaches. While the overall temperature trends show the same daily fluctuation, there is no clear gradient and the sensors that provide outliers are emphasized. The differences throughout the network (based on 80 sensor) between co-kriging and universal kriging (*elevation only*) at 9AM (mean day during the month of observations) are provided in *figure 42 and 43*. Universal kriging shows a smooth trends (range:  $20.4 - 22.7^{\circ}\text{C}$ ), while the outliers are emphasized by co-kriging (range:  $19.4 - 23.5^{\circ}\text{C}$ ). The difference between the two approaches increases during the periods with high temperatures.



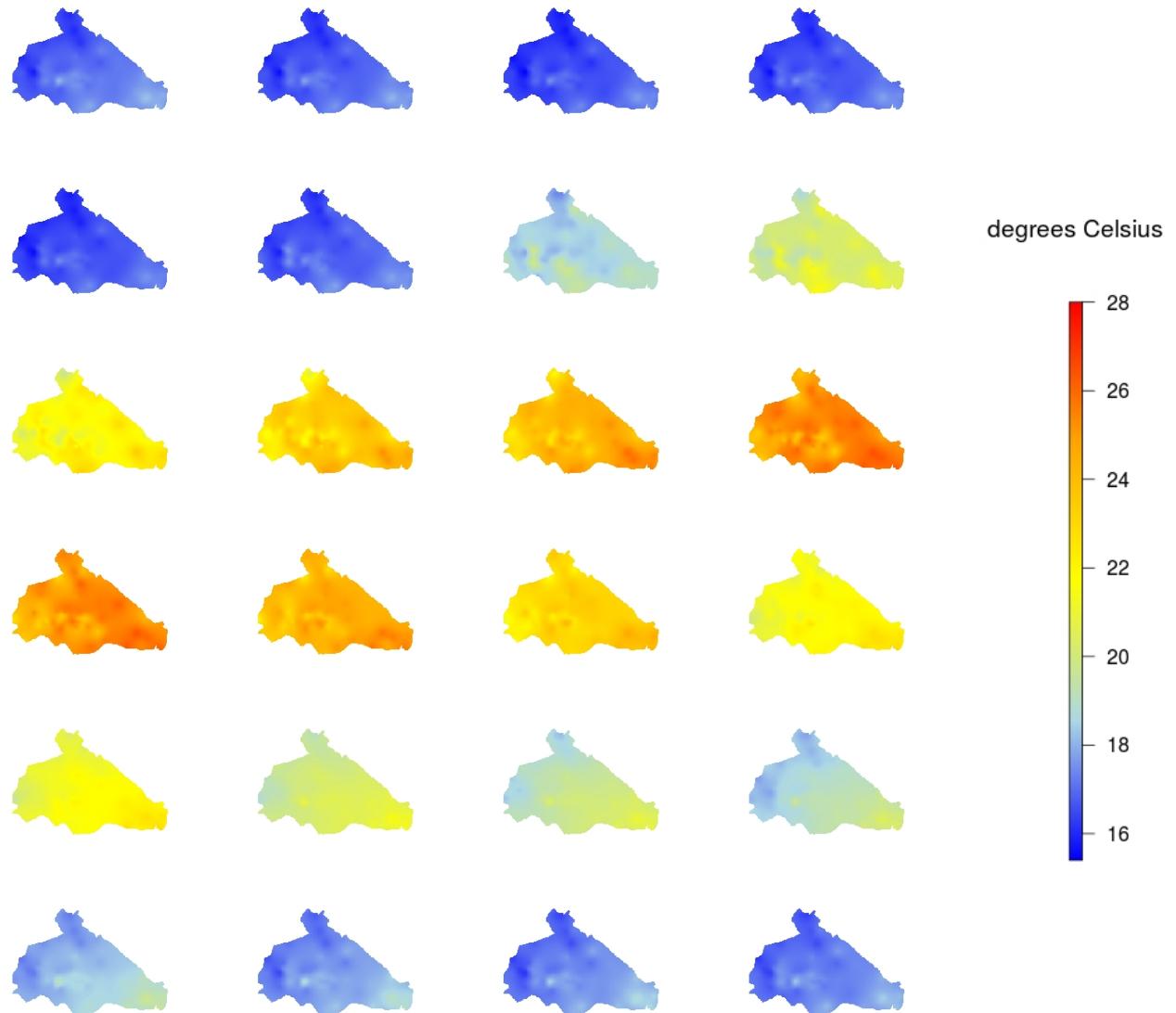
**Figure 42:** Mean temperature with Universal Kriging (elevation) at 9AM (80 sensors)

**Figure 43:** Mean temperature with Co-Kriging (elevation) at 9AM (80 sensors)

### 5.3.5 Spatio-temporal kriging

Spatio-temporal kriging was tested with the time dimension only (ST-ordinary kriging - *STOK*), with elevation (ST-universal kriging - *STUK*), and with dynamic covariates (ST-dynamic kriging - *STDK*). The RMSE of the approaches is relatively high: the RMSE with 80 sensors (*STOK*) is  $0.96^{\circ}\text{C}$ , when including elevation this remains  $0.96^{\circ}\text{C}$ , while dynamic covariates increase the RMSE to  $0.99^{\circ}\text{C}$ . At 40 sensors, the error remain similar for the three approaches, while with 20 sensors, the errors increase slightly to 0.96 (*STOK* and *STUK*) and  $1.00^{\circ}\text{C}$  (*STDK*). The 8-sensor dataset results in a RMSE of 1.07 (*STOK*), 1.18 (*STUK*), and  $1.20^{\circ}\text{C}$  (*STDK*). While the errors are larger than for any of the previously tested approaches, there is a move from the trend that RMSE *temporal > spatial*, to a RMSE that is smaller in the temporal domain (*Annex K*). The additional interpolation in the time-domain does, however, result in larger errors at the separate hours and also the largest errors

of all tested approaches. The inclusion of the time-dimension in ST interpolation is visually attractive when animating the temperature trends, and could provide benefits when reducing the interval of observations, but with the current interval the approach does not perform well. Another limitation of using the temporal interval is that temperature can change quickly under different conditions of cast shadows (position of the sun, clouds) and rain/wind. Assuming a strong correlation in time is thus not necessary when being able to sample frequently, as can be seen from the relatively large errors of this approach. A full day of observations using spatio-temporal interpolation is shown in *figure 44*. Hourly variation is limited when placing it in the context of daily variation. The sensors that provide a cooler temperature during the day generally also provide a warmer temperature during the night. This ‘stability’ can be both caused by micro-climatic differences as sensor-problems.



**Figure 44:** Daily temperature variation based on Spatio-Temporal OK (80 sensors)

### 5.3.6 Accuracy in predicting the mean temperature

As not all studies require hourly temperature - which is mainly interesting for detailed studies regarding phenological phases or pest-risk - the error of the interpolation approaches in predicting the mean temperature has also been assessed for one week. This was done by comparing the observed daily mean with values extracted from seven rasters with the daily mean temperature. The RMSE in predicting the daily mean with all tested interpolation sets and 80, 40, 20 and 8 sensors is provided in *table 11*. As this accuracy does not include the leaving-out of sensors, it mainly indicates the strength of interpolation; a higher RMSE means the actual observed values are adjusted in the interpolation due to the provided covariates. Table 11 shows that co-kriging and spatio-temporal kriging - which did not perform well during cross-validation - provide predictions closest to the observed values. While in co-kriging the additional values do not change the accuracy, adding elevation improves the accuracy in ST kriging, while the dynamic set provides the worst results.

	<b>80 Sensors</b>	<b>40 Sensors</b>	<b>20 Sensors</b>	<b>8 Sensors</b>
<i>Ordinary Kriging</i>	0.29	0.23	0.26	0.53
<i>Universal Kriging</i>	0.26	0.20	0.23	0.49
<i>Auto- Universal Kriging</i>	0.26	0.19	0.23	0.47
<i>Dynamic Kriging</i>	0.26	0.20	0.23	0.62
<i>Auto- Dynamic Kriging</i>	0.40	0.19	0.20	0.64
<i>Co-Kriging (elevation)</i>	0.14	0.10	0.10	0.62
<i>Co-Kriging (elevation and radiation)</i>	0.14	0.10	0.10	0.62
<i>Co-Kriging (elevation, slope, aspect and radiation)</i>	0.14	0.10	0.10	0.62
<i>Spatio-Temporal Kriging</i>	0.14	0.09	0.07	0.06
<i>Spatio-Temporal Kriging with elevation</i>	0.14	0.08	0.07	0.06
<i>Spatio-Temporal Kriging with dynamic covariates</i>	0.35	0.78	0.82	5.01

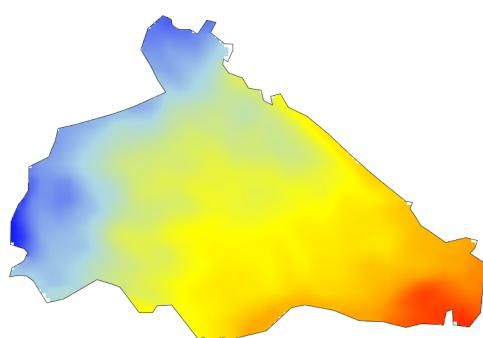
**Table 11:** RMSE (degrees Celsius) of daily mean *observed* and *predicted* mean temperature

## 5.4 Discussion

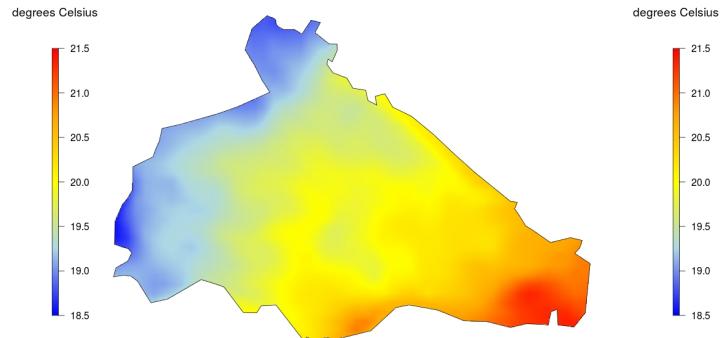
The analyses of the different spatial (and temporal) interpolation approaches can result in some general conclusions, although most of the differences will have (dis)advantages depending on the number of sensors and objective of the network. With the full 80-sensor dataset, the overall RMSE is lowest for the basic dynamic kriging ( $0.77^{\circ}\text{C}$ ). Universal kriging and dynamic kriging (basic, automatized and RandomForest) all provide a RMSE  $<0.8^{\circ}\text{C}$ , although with 40 sensors, this remains

only the case for basic UK (including elevation). At 20 sensors, the error is still  $0.8^{\circ}\text{C}$  for universal kriging, while it is  $0.9^{\circ}\text{C}$  for co-kriging with elevation. It can be expected that 20 sensors (25 when assuming a 20% loss) is the ideal number for similar projects in the future, due to relatively low cost and the ability to use geostatistical interpolation to create a geonetwork. Dynamic kriging, which was expected to improve the accuracy, provides the worst results at 8 sensors ( $\text{RMSE} = 1.44^{\circ}\text{C}$ ). Co-kriging does not provide the same level of accuracy as the universal kriging approaches, although the risk of reducing the accuracy when adding covariates is reduced. All of the approaches that do not include the temporal domain have a lower accuracy in time (mean RMSE of the sensors at all hours) than in space (mean RMSE at all hours for the sensors). This is different for spatio-temporal kriging, which provides a higher accuracy in time than in space. This temporal-accuracy is slightly better than for the other approaches, while the overall errors (at 80, 40 and 20 sensors) are higher.

While the network accuracy, based on a cross-validation, is relatively low for spatio-temporal and co-kriging, the measurements of sensors used in interpolation are least modified. Interpolating with all sensors and comparing the value at the location of sensors with the actual observed values results in the lowest errors for these two approaches. Especially spatio-temporal kriging provides values close to the observed values. This is a clear benefit when the sensors are reliable and accurate, but with the current state of sensors and shields, the smoothing impact of the other kriging approaches could be preferred. The difference in mean temperature (20 sensors, 1 week) when using universal kriging and spatio-temporal universal kriging (both with elevation) is provided in *figure 45* and *46*. While the differences are small, the variation in ST kriging remains higher and shows less 'smoothing' due to the correlation with elevation. In situation where the value at the location of the sensor is more important than the other values in the network (e.g. when all farmers participating in field-trials have a sensor close to their field), a low sensor RMSE will be preferred over a low network RMSE. Having predicted values close to the observed values can help in identifying local micro-climates if sufficient (and correctly functioning) sensors are used. A downside of a limited adjustment of actual observed values is that sensors that provide incorrect data will not be 'corrected' by the covariates that have a known impact on the temperature (especially elevation). Which approach will be preferred depends on the reliability of sensors and purpose of the network.



**Figure 45:**  $T_{\text{mean}}$  derived by Universal Kriging



**Figure 46:**  $T_{\text{mean}}$  derived by Spatio-Temporal UK

## 5.5 Recommendations

The analyses of different interpolation methods has shown that some approaches have clear benefits over others as revealed by the different RMSEs in a leave-one-out cross validation. As this chapter has used some basic parameters to find main differences, it is important to further adjust the parameters once a well-suited interpolation approach has been selected. Co-kriging and spatio-temporal kriging are two of the approaches that can benefit from more study/adjustment, as they are more complex than the other (ordinary and universal) kriging approaches. Regarding covariates, it is clear that the dynamic set has not resulted in the desired improvement. Only with the full dataset, there is a small improvement in accuracy over other approaches. More important is the large decrease in accuracy when working with smaller number of sensors. Selecting a larger study area, with more impact of cast shadows (larger objects, variation of aspects) can help to improve the understanding of adding hourly radiation as covariate and possibly increase accuracy in future networks.

This chapter has shown that some interpolation approaches have a better ability to include trends and use this to predict temperature at a left-out location, while other approaches are better in maintaining the actual data and thus maintain local variation. While most interpolation can be adjusted to do one of these things, it is important to understand these differences (compare *figures 44 and 45*) and assess which will be better for the objective of the study. In studies where a general indication of temperature is required in a larger area, universal kriging can provide good results - even if it also adjusts the data at the location of sensors. The advantage is that relatively few sensors can provide a good indication of the local climate. If the sensors will be used for more detailed local-specific studies (e.g. estimating periods of pest risk and exact temperature at the location of certain trials), it will be better to work with more sensors and use an interpolation approach that limits the correction at the location of the sensors. With the current reliability of sensors and large corrections that have taken place, this local variation can also be the result of a sensor being in contact with water. For this reason, improving shielding of the sensor will also be a main recommendation here.

Analysing the accuracy with cross-validation is a crucial, but time consuming approach. Especially with more advanced approaches, such as co-kriging and spatio-temporal kriging with covariates, estimating the errors at a sufficiently large number of hours to get accurate results, can take a very long time. While a fully automatized approach, such as automatic kriging with the *automap* package, can provide very good results, a one-time indication of the errors and the location of these can improve the interpretation of the results. This means parrallelization is required, as calculations might be too *heavy* for computers and mobile devices in developing countries. For this reason, opportunities for parrallelization (e.g. cloud-computing) have to be studied to find alternatives for these complex calculations. Although accuracy differs between the tested approaches, only in few cases the RMSE becomes greater than  $1^{\circ}\text{C}$ . In most cases the errors are relatively small, especially in the context of daily trends (*figure 43*). A final recommendation is to compare the interpolation approaches and assess how they influence the size of extremes throughout the area (e.g. % of time above  $30^{\circ}\text{C}$ ). This will be especially important in more detailed agronomic/ecological studies.

## 6 Summary

After analysing the different aspect of the network, it is clear that most of the challenges that are faced when setting up a weather sensor network for agricultural purposes with low-cost materials can be overcome with the right sensor-adjustment and modelling. The main challenge remains, as expected, the creation of sensor-shielding that provides reliable output. Existing *certified* structures are expensive, and it has become clear that many of the more expensive components (e.g. active aeration, large housing) provide benefits that will be difficult to replicate with small-diameter PVC tubes with passive aeration. Previous studies that focused on low-cost shielding have either been conducted under a canopy, with electronic components, or in a different (cooler) climate. This study has shown some benefits of adding insulating foil, but the impact of aeration was limited - and in some situations negative due to increased risk of sensor-contact with water. Two easy adjustments that were not tested in this study include: (1) different placing of the sensor (in this study there is a risk of reflecting radiation); and (2) addition of an *U-shape* tube (see Clark et al. 2006) to the main tube. The latter can avoid an impact of heating of the main tube (to which no foil was attached), although the connection between main tube and the horizontal tube was closed with tape.

As the errors of using low - in stead of high - sensor resolution are very small compared to errors during (temporal/spatial) interpolation, the sensors can measure at this resolution and double the number of observations that can be stored. Correction of temperature can be done based on a change-in-time ( $\Delta t$ ) correction, as the heating/cooling inside the PVC tube is 40-52% faster than in a certified (Stevenson) shield. The strength of these models is high ( $R^2 > 0.85$ ), and the ratio of change in the PVC versus the Stevenson is lowest when using insulating foil. Humidity has a clear and strong correlation with temperature and, as humidity is expected to be much less important in agronomic studies (excluding pest/disease risk analyses), humidity can be predicted from the temperature. This can be done by measuring both units at certain locations and creating a linear model between temperature and humidity. Correcting humidity based on  $\Delta t$  is more difficult than for temperature as the model is much weaker (max  $R^2 = 0.69$ ). One reason for this is that correction of humidity is done for more data, as it surpasses 100% on most days. A second reason is that water can enter the shields as these are not water-sealed. An important advantage of not measuring humidity is that sensors can be cheaper and/or measure a longer interval. A sensor that only measures the temperature at a low resolution at 1-hour interval can make observations during 341 days.

Of the hundred sensors that were placed in Aquires - a large coffee plantation with  $\pm 600$  meters elevation gradient - 86 could be retrieved. A further six sensors did not provide (visually) correct data, which means 80 sensors could be used in spatial interpolation. This 20% loss will have to be included when assessing the number of sensors required in future projects. The accuracy of the network has been assessed in space and time. There was a clear daily trend in errors, which are lowest during periods without sun and peak around solar noon. The mean RMSE is below  $1^{\circ}\text{C}$  ( $0.65^{\circ}\text{C}$  during one month of observations and IDW-interpolation), but has reached  $> 1^{\circ}\text{C}$  during several hours each day - peaking at  $1.2^{\circ}\text{C}$ . There is a clear ( $r = 0.74$ ) correlation between the accuracy

(mean daily RMSE) and daily maximum temperature, indicating that on warm/sunny days the errors become larger. At hours when correlation of elevation and temperature is high, the errors are low. During days with a high maximum temperature, the correlation of daily radiation with temperature becomes stronger. This indicates that on days with high temperatures, an increase of the correlation of temperature with the hourly radiation can partly compensate for the reduction of correlation with elevation. Overall, the static covariates (elevation, slope, aspect, canopy height, leaf-area-index and daily radiation) can explain just over half (52%) of variation in temperature, ranging from 3 to 97%. Elevation has - with an importance between 60 and 75% - the highest relative importance in these models. All other factors contribute 10% or less to this model.

The different covariates show distinct trends in correlation during the day, although in many cases of limited strength. The strongest correlation with temperature is elevation, which is strongest during the night and weakest around noon. Slope and aspect have both a positive and negative correlation with temperature during the day, although never stronger than  $r = 0.3$ . Leaf-area-index has a strongest correlation around solar noon, although this remains very weak. Canopy height has a strongest correlation with temperature around sunset, but is also correlated with elevation which can explain this trend. Hourly radiation (including hill-shade, cast shadows and a solar constant) has a strongest correlation at both 7AM and 3PM, which is two hours after sunrise and before sunset. The correlation is stronger than for most of the other covariates except for elevation and can be larger than  $r = 0.4$ . A linear model with all these covariates, that explains temperature variation, has a mean  $R^2$  of 0.53, which is a small improvement over using static covariates. The model is strongest around sunset and has a  $R^2 > 0.5$  from 3PM to 7AM. The weakest models can be found around noon and only have a strength of  $R^2 = 0.25$ . Elevation is the most important covariate in these models (contributing 67%); hourly radiation contributes 10% on average, and 20% during hours with sun.

Interpolation has been done with 80 (1 sensor per  $\pm 0.1\text{km}^2$ ), 40, 20 and 8 (1 sensor per  $\pm 1\text{km}^2$ ) sensors. The geostatistical approaches included *ordinary*, *universal*, *dynamic*, *co-* and *spatio-temporal* kriging. For the more basic approaches, it is clear that applying universal kriging resulted in smaller errors than ordinary kriging in this area. Dynamic kriging only outperforms universal kriging in the full 80-sensor network (all based on a leave-one-out cross-validation). At 20 sensors, an amount that is expected to be most in line with future networks (starting with 25 sensors), the RMSE with universal kriging (including elevation) was still only 0.8. Co-kriging and spatio-temporal kriging did not perform as good as the other approaches, although they resulted in the smallest differences with the original data for the sensors that were included in the interpolation

While the network in this study cost  $\pm \text{US\$ } 8,700$  (*Appendix A*), a network of 25 sensors, can be created with a much smaller budget: 20 low-res (temperature) iButton sensors (*DS1922L-F5*), 5 high-res (temperature and humidity) iButton sensors (*DS1923-F5*), 50m thin white PVC, 50 PVC elbows,  $1\text{m}^2$  insulating foil, a small amount of fibre-glass mesh, and labor for construction (drilling holes and assembling). The cost for this weather network - which can store 341 days of 1-hour resolution data - will be approximately  $\text{US\$ } 1,450$ . A 50-sensor network would still cost  $<\text{US\$ } 3,000$ .

## 7 Conclusions & Recommendations

This study has generated information that can assist in decision-making about network configuration, and identified several opportunities for further improvements when creating a low-cost weather network. In the following section, the main findings of this study will be synthesized and translated into considerations for the design of (iButton) sensor-networks for agricultural applications.

### ***Key information for decisions on network set-up***

The optimal sensor resolution that allows the longest period of observations is the lower resolution (for both temperature and relative humidity). The loss in accuracy of the actual observations is negligible compared to losses during other phases (temporal and spatial interpolation). Measuring at 1-hour interval - and with measurements close to the hour - provided better results than longer intervals and a random start-moment. Shorter intervals could further improve accuracy, but will not be able to cover a sufficiently long period to monitor most crops. The memory of iButtons is limited (especially for the cheaper ones), which reduces the benefits of measurements at < 1 hour. While insulation foil provided the best *raw* data (especially on sunny days), no shield-modification provides data of sufficient quality without some form of correction. Whereas polynomial data-correction performed well, it was found that a model that corrects data based on the rate of heating/cooling over time ( $\Delta t$ -calibration, based on the *thermal inertia* properties) provided the best result.

Of the covariates, elevations showed the strongest (negative) correlation with temperature. Temperature differences are most difficult to predict during sunny periods (hours and days). Adding hourly radiation as covariate resulted in slightly stronger models than could be created with only static covariates. The hourly radiation contributed 20% to the models during hours with sun (during which models explained  $\pm 50\%$  of variance). Models at these hours explained less variation than during the night-period. Different geostatistical interpolation approaches were tested with 80, 40, 20, and 8 sensors. The risk with the smaller groups of sensors (8 and 20) is that of overfitting of the variogram (50 points is recommended as the minimum). Universal kriging with elevation provided the most stable results, which is related to the strong elevation/temperature correlation. Adding dynamic covariates only improved the results for the 80-sensor dataset, while errors increased when using dynamic covariates with sets with fewer sensors. Co- and spatio-temporal- kriging resulted in the least correction of the values measured by sensors that are included in the interpolation.

### ***Proposed network set-up***

The set-up of the network will depend on many factors, including the: (a) network objective, (b) available budget, and (c) type of terrain. There is a clear benefit of having 50 sensors available for spatial interpolation, as this will reduce errors related to variogram overfitting and will allow more advanced interpolation approaches (e.g. RandomForest Kriging). Three different network objectives include: (1) basic temperature monitoring in perennial crops; (2) threshold monitoring in annual crops; (3) detailed temperature/humidity monitoring to assess risk of pests and diseases. The proposed set-up of networks for these three different objectives are presented in the following.

**Long-term temperature monitoring of a perennial crop.** Basic information ( $T_{mean}$ ,  $T_{minimum}$ , and  $T_{maximum}$ ) can be derived from observations at 3-hourly interval. Only measuring temperature at low resolution would enable to cover a period of nearly three years. Few sensors (10-25) would be required in such a network, although this would obviously depend on the size of the area. As coffee is often cultivated in complex terrain, a good spatial interpolation approach would be Universal Kriging with elevation. In this case, the general trend is more important than the dynamics of a micro-climate. If the number of sensors is too low for geostatistical interpolation, a deterministic approach, such as Inverse Distance Weighting, is recommended.

**Threshold monitoring in annual crops.** Another objective of a network could be to make more detailed (temperature and humidity) observations linked to annual crops. A 1- or 2- hour interval is preferred. If the crop is of short-duration, it could be an option to use sensors that measure both temperature and humidity at low-resolution (171 days). Otherwise, it would be a good solution to add one humidity-measuring sensor to every  $n^{th}$  temperature-measuring sensor. A network would require a medium (25-50) number of sensors to maintain clear local differences. The most suitable interpolation approach will depend on the area, as a geostatistical approach is possible, although would be of limited use in a flat area with limited complexity. Adding elevation is a ‘safe’ option, whereas distance to larger water bodies and hourly radiation would also be useful to include if this would be relevant in the area.

**Detailed temperature/humidity monitoring to assess risk of pests and diseases.** A final objective that is discussed is that of pest/disease-risk monitoring. This would only be useful if the distribution within the area of the pest/disease is relevant. Such networks would be the most difficult to create, as assessing the risk of pests and diseases would require detailed input (e.g. geographic distribution) and will likely involve fuzzy modelling of different risks. This would require observations at 1-hour interval or more frequently. The required number of sensors will also be the highest, as information about micro-climatic differences are required. Interpolation should limit the correction of actually observed values, which means co-kriging would be the best option as this will improve the accuracy with every added covariate.

#### *Opportunities to generate more information for improving the set-up*

1. Develop and compare different shields with combinations of low-cost materials that can be expected to improve reflectivity and aeration, as well as limit the impact of reflecting radiation.
2. Create a script that can launch all the iButtons at the same start-moment.
3. Examine more methods of data correction, such as the used  $\Delta t$ -calibration. A model based on the physical properties (rate of heating/cooling) of the shield requires well-studied parameters.
4. Study the benefits of adding dynamic covariates (especially hourly radiation) to spatial interpolation approaches. This would require the set-up of a network in a larger (complex) area.
5. Assess the opportunities for parallel computing to enable faster (cross-validation) calculations.
6. Investigate how the different spatial interpolation approaches will have an impact on the duration and size of extremes in the area.

## 8 References

- Altieri, M. 2002. Agroecological Principles for Sustainable Agriculture. In: *Agroecological innovations: increasing food production with participatory development*. Edited by N. Uphoff. ISBN: 1-85383-857-9.
- Armstrong, M., & Boufassa, A. 1988. Comparing the robustness of ordinary kriging and lognormal kriging: Outlier resistance. *Mathematical Geology*, **20**: 447-457. doi:10.1007/BF00892988
- 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**: 77-89. doi:10.1016/j.agrformet.2013.03.008
- Atkinson, P. M., & Lloyd, C. D. 1998. Mapping Precipitation in Switzerland with Ordinary and Indicator Kriging. *Journal of Geographic Information and Decision Analysis*, **2**(2): 65-76.
- Baggio, A. 2005. Wireless sensor networks in precision agriculture. In: *Workshop on Real-World Wireless Sensor Networks*. Retrieved from <https://intranet.daiict.ac.in/~ranjan/isn2005/APP/baggio05wireless.pdf>
- Barnett, B. J., & Mahul, O. 2007. Weather index insurance for agriculture and rural areas in lower-income countries. *American Journal of Agricultural Economics*, **89**: 124-247. doi:10.1111/j.1467-8276.2007.01091.x
- Barr, S., & Orgill, M. M. 1989. Influence of External Meteorology on Nocturnal Valley Drainage Winds. *Journal of Applied Meteorology*, **28**: 497-517. doi:10.1175/1520-0450(1989)028
- Benestad, R. E. 2001. A comparison between two empirical downscaling strategies. *International Journal of Climatology*, **21**: 1645-1668. doi:10.1002/joc.703
- Bivand, R., & Rundel, C. 2013. rgeos: Interface to Geometry Engine - Open Source (GEOS). *R package version 0.3-2*. Retrieved from <http://cran.r-project.org/web/packages/rgeos/index.html>
- 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**: 115-122. doi:10.1016/S0924-2716(99)00002-7
- 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. Retrieved from <http://lvis.gsfc.nasa.gov>
- Blumthaler, M., Ambach, W., & Ellinger, R. 1997. Increase in solar UV radiation with altitude. *Journal of Photochemistry and Photobiology B: Biology*, **39**: 130-134. doi:10.1016/S1011-1344(96)00018-8
- Bradter, U., Kunin, W. E., Altringham, J. D., Thom, T. J., & Benton, T. G. 2013. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. *Methods in Ecology and Evolution*, **4**: 167-174. doi:10.1111/j.2041-210x.2012.00253.x
- 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**(2): 159-166. doi:10.1016/0168-1923(84)90017-0
- Budowski, G. 1980. The Place of Agro-Forestry in Managing Tropical Forests. Retrieved from <http://www.sidalc.net/repdoc/A5462I/A5462I.PDF>
- Butler, D. R. 1977. Coffee Leaf Temperatures in a Tropical Environment. *Acta Botanica Neerlandica*, **26**(2): 129-140.
- Chakraborty, S., Tiedemann, A. V., & Teng, P. S. 2000. Climate change: potential impact on plant diseases. *Environmental Pollution*, **108**: 317-326. doi:10.1016/S0269-7491(99)00210-9
- Charney, J. G. 1949. On a physical basis for numerical prediction of large-scale motions in the atmosphere.

- Journal of Meteorology*, **6**: 372-385. doi:10.1175/1520-0469(1949)
- Cheng, C. S., Li, G., Li, Q., & Auld, H. 2008. Statistical downscaling of hourly and daily climate scenarios for various meteorological variables in South-central Canada. *Theoretical and Applied Climatology*, **91**, 129-147. doi:10.1007/s00704-007-0302-8
- Clark, P. E., Johnson, D. E., Harris, N., & Thomas, D. R. 2006. Low-Cost Radiation Shielding for Use in Mapping the Thermal Environments of Rangeland Animals. *Rangeland Ecology & Management*, **59**(6): 674-679 doi:10.2111/05-093R2.1
- Cornell University. n.d.. *Aquiares Estate: Coffee & Community Since 1890*. Retrieved from <http://ip.cals.cornell.edu/undergrad/docs/study-abroad-prog-Aquiares-Estate.pdf>
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G., Pasteris, P. P. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**: 2031-2064. doi:10.1002/joc.1688
- 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**, 143-160. doi:10.1016/S0168-1923(01)00236-2
- DAVIS. 2014a. Remote Integrated Pest Management Station for Agriculture. Retrieved October 29, 2014, from <http://www.davisnet.com/weather/uses/agriculture-solutions/remote-ipm-station.asp>
- DAVIS. 2014b. WeatherLinkIPTM for Vantage Stations. Retrieved October 29, 2014, from [http://www.davisnet.com/weather/products/weather\\_product.asp?pnum=06555#price](http://www.davisnet.com/weather/products/weather_product.asp?pnum=06555#price)
- Deardorff, J. W. 1978. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research*, **83**(C4): 1889-1903. 1978doi:10.1029/JC083iC04p01889
- Oxford Dictionaries. 2015. Adiabatic lapse rate. Retrieved February 15, 2015, from <http://www.oxforddictionaries.com/definition/english/adiabatic-lapse-rate>
- Dodson, R., & Marks, D. 1997. Daily air temperature interpolated at high spatial resolution over a large mountainous region. *Climate Research*, **8**: 1-20. doi:10.3354/cr008001
- e-Agriculture. 2014. Retrieved November 04, 2014, from <http://www.e-agriculture.org/>
- Edwards, P. N. 2010. A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming. ISBN: 9780262013925, 552 pp. MIT Press, US.
- Erickson, C. O. 1964. Meteorological Satellites. *Science* (New York, N.Y.), **143**: 612-613. doi:10.1126/science.143.3606.612
- Eskelson, B. N. I., Anderson, P. D., Hagar, J. C., & Temesgen, H. 2011. Geostatistical modeling of riparian forest microclimate and its implications for sampling. *Canadian Journal of Forest Research*, **41**(5): 974-985. doi:10.1139/x11-015
- Ewel, J., Benedict, F., Berish, C., Brown, B., Gliessman, S., Amador, M., Price, N. 1982. Leaf area, light transmission, roots and leaf damage in nine tropical plant communities. *Agro-Ecosystems*, **7**(4): 305-326 doi:10.1016/0304-3746(82)90023-3
- Ewert, F., & Pleijel, H. 1999. Phenological development, leaf emergence, tillering and leaf area index, and duration of spring wheat across Europe in response to CO<sub>2</sub> and ozone. *European Journal of Agronomy*, **10**: 171-184. doi:10.1016/S1161-0301(99)00008-8
- Gaur, R. K., & Sharma, P. 2014. Approaches to Plant Stress and their Management. Springer India.
- Geller, T. 2007. Envisioning the wind: Meteorology graphics at weather underground. *IEEE Computer Graphics and Applications*, **27**: 92-97. doi:10.1109/MCG.2007.124

- Goovaerts, P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, **228**: 113-129. doi:10.1016/S0022-1694(00)00144-X
- Goudriaan, J. 1977. Crop micrometeorology: a simulation study. Wageningen. Retrieved from [library.wur.nl/WebQuery/wda/104086](http://library.wur.nl/WebQuery/wda/104086)
- Grimm, R., Behrens, T., Mrker, M., & Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. *Geoderma*, **146**: 102-113. doi:10.1016/j.geoderma.2008.05.008
- Groemping, U. 2006. Relative importance for linear regression in R: the package relaimpo. *Journal Of Statistical Software*, **17**: 139-147. Retrieved from <http://www.jstatsoft.org/v17/a01/paper>
- Hacienda Aquiares. 2008. Retrieved October 01, 2014, from <http://www.cafeaquiares.com/>
- Hackett, C. 1991. PLANTGRO: a software package for coarse prediction of plant growth. pp. 242. CAB Direct.
- Hardwick, S. R., Toumi, R., Pfeifer, M., Turner, E. C., Nilus, R., & Ewers, R. M. 2015. The relationship between leaf area index and microclimate in tropical forest and oil palm plantation: Forest disturbance drives changes in microclimate. *Agricultural and Forest Meteorology*, **201**: 187-195. doi:10.1016/j.agrformet.2014.11.010
- Hellström, R. A., & Mark, B. G. 2006. An Embedded Sensor Network for Measuring Hydrometeorological Variability Within an Alpine Valley. In: *63rd Eastern Snow Conference*. Newark. Retrieved from [http://www.easternsnow.org/proceedings/2006/hellstrom\\_and\\_mark.pdf](http://www.easternsnow.org/proceedings/2006/hellstrom_and_mark.pdf)
- Hengl, T. 2007. A Practical Guide to Geostatistical Mapping of Environmental Variables. Retrieved from <http://www.geology.cz/projektvzdelavani/Pomocna/EUR22904en.pdf>
- Hengl, T. 2015. Global Soil Information Facilities. Retrieved from <http://cran.r-project.org/web/packages/GSIF/index.html>
- Hiemstra, P. 2013. automap: Automatic interpolation package. Retrieved from <http://cran.r-project.org/web/packages/automap/automap.pdf>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**: 1965-1978. doi:10.1002/joc.1276
- Hijmans, R. J., & Graham, C. H. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, **12**: 2272-2281. doi:10.1111/j.1365-2486.2006.01256.x
- Hijmans, R. J., & van Etten, J. 2010. raster: Geographic analysis and modelling with raster data. R Package Version.
- Holden, Z. A., Abatzoglou, J. T., Luce, C. H., & Baggett, L. S. 2011. Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain. *Agricultural and Forest Meteorology*, **151**: 1066-1073. doi:10.1016/j.agrformet.2011.03.011
- Holden, Z. A., Klene, A. E., F. Keefe, R., & G. Moisen, G. 2013. Design and evaluation of an inexpensive radiation shield for monitoring surface air temperatures. *Agricultural and Forest Meteorology*, **180**: 281286. doi:10.1016/j.agrformet.2013.06.011
- Intergovernmental Panel on Climate Change. 2014. Climate Change, Adaptation, and Vulnerability. *Organization & Environment*, **24**: 1-44. [http://ipcc-wg2.gov/AR5/images/uploads/IPCC\\_WG2AR5\\_SPM\\_Approved.pdf](http://ipcc-wg2.gov/AR5/images/uploads/IPCC_WG2AR5_SPM_Approved.pdf)
- 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**: 1075-1084 doi:10.1175/1520-0450(2001)040
- Jedermann, R., Behrens, C., Westphal, D., & Lang, W. 2006. Applying Autonomous Sensor Systems in Logistics - Combining Sensor Networks , RFIDs and Software Agents. *Sensors and Actuators A: Physical*, **132**: 370-375. doi:10.1016/j.sna.2006.02.008
- Jones, J. W., Tsuji, G. Y., Hoogenboom, G., Hunt, L. A., Thornton, P. K., Wilkens, P. W., Singh, U. 1998. Decision support system for agrotechnology transfer: DSSAT v3. In: *Understanding Options for Agricultural Production* (pp. 157-177).
- Jones, P. G., & Thornton, P. K. 2013. Generating downscaled weather data from a suite of climate models for agricultural modelling applications. *Agricultural Systems*, **114**: 1-5. doi:10.1016/j.agsy.2012.08.002
- Lane, A., & Jarvis, A. 2007. Changes in climate will modify the geography of crop suitability: agricultural biodiversity can help with adaptation. *SAT eJournal*, **4**: 1-12.
- Lawrence, M. G. 2005. The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, **86**: 225-233. doi:10.1175/BAMS-86-2-225
- Letten, A. D., Ashcroft, M. B., Keith, D. A., Gollan, J. R., & Ramp, D. 2013. The importance of temporal climate variability for spatial patterns in plant diversity. *Ecography*, **36**: 1341-1349. doi:10.1111/j.1600-0587.2013.00346.x
- Liaw, A., & Wiener, M. 2002. Classification and Regression by randomForest. *R News*, **2**: 1822. doi:10.1177/154405910408300516
- 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**: 179-204. doi:10.1146/annurev.environ.041008.093740
- Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R., & Anderson, J. 2002. Wireless Sensor Networks for Habitat Monitoring. In: *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications* (pp. 8897). doi:10.1145/570738.570751
- Martínez-Cob, A. 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. *Journal of Hydrology*, **174**(1-2): 19-35. doi:10.1016/0022-1694(95)02755-6
- Maxim Integrated. n.d.. Retrieved November 05, 2014, from <http://www.maximintegrated.com/en/products/comms/ibutton.html>
- McNider, R. T., & Pielke, R. A. 1981. Diurnal Boundary-Layer Development over Sloping Terrain. *Journal of the Atmospheric Sciences*, **38**: 2198-2212. doi:10.1175/1520-0469(1981)038<2198:DBLDOS>2.0.CO;2
- McWilliam, A.-L. C., Roberts, J. M., Cabral, O. M. R., Leitao, M. V. B. R., Costa, A. C. L. De, Maitelli, G. T., & Zamparoni, C. A. G. P. 1993. Leaf Area Index and Above-Ground Biomass of terra firme Rain Forest and Adjacent Clearings in Amazonia. *Functional Ecology*: **7**: 310-317. doi:10.2307/2390210
- Mendelsohn, R., Kurukulasuriya, P., Basist, A., Kogan, F., & Williams, C. 2007. Climate analysis with satellite versus weather station data. *Climatic Change*, **81**: 71-83. doi:10.1007/s10584-006-9139-x
- Menne, M. J., Williams, C. N., & Palecki, M. A. 2010. On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research: Atmospheres*, **115**. doi:10.1029/2009JD013094
- Menzel, A., Sparks, T. H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Zust, A. 2006. European phenological response to climate change matches the warming pattern. *Global Change Biology*, **12**: 1969-1976. doi:10.1111/j.1365-2486.2006.01193.x

- Merriam-Webster Dictionary. 2014. Merriam-Webster Dictionary. Retrieved from <http://www.merriam-webster.com/dictionary/>
- Miller-Rushing, A., Primack, R., & Bonney, R. 2012. The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, **10**: 285-290. doi:10.1890/110278
- Mittra, S., Van Etten, J., & Franco, T. 2013. Collecting Weather Data in the Field with High Spatial and Temporal Resolution Using iButtons (June 2013 Version). Retrieved from [http://www.bioversityinternational.org/uploads/tx\\_news/Technical\\_Guide\\_for\\_collecting\\_weather\\_data\\_in\\_the\\_field\\_using\\_iButtons\\_1628\\_2\\_.pdf](http://www.bioversityinternational.org/uploads/tx_news/Technical_Guide_for_collecting_weather_data_in_the_field_using_iButtons_1628_2_.pdf)
- Monsanto. 2014. FieldScripts. Retrieved October 29, 2014, from <http://www.monsanto.com/products/pages/fieldscripts.aspx>
- Monteith, J. L. 1972. Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, **9**: 747-766. doi:10.2307/2401901
- Morales, L., Canessa, F., Mattar, C., & Orrego, R. 2007. Comparison of stochastic and regression geo-statistics interpolation methods for detection of microclimatic areas. Retrieved from [http://www.itc.nl/ISSDQ2007/proceedings/Session\\_2\\_SpatialStatistics/PaperMorales\\_et\\_al..pdf](http://www.itc.nl/ISSDQ2007/proceedings/Session_2_SpatialStatistics/PaperMorales_et_al..pdf)
- 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**: 19680-19685. doi:10.1073/pnas.0701855104
- Nhemachena, C., & Hassan, R. M. 2007. Micro-Level Analysis of Farmers' Adaptation to Climate Change in Southern Africa. *IPFRI Discussion Paper 00714*. Retrieved from <http://www.ifpri.org/publication/micro-level-analysis-farmers-adaptation-climate-change-southern-africa-0>
- International Standards Organisation. 2007. Meteorology – Air temperature measurements – Test methods for comparing the performance of thermometer shields/screens and defining important characteristics (ISO 17714:2007). Retrieved from [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=31498](http://www.iso.org/iso/catalogue_detail.htm?csnumber=31498)
- Parmentier, B., McGill, B., Wilson, A., Regetz, J., Jetz, W., Guralnick, R., Schildhauer, M. 2014. An Assessment of Methods and Remote-Sensing Derived Covariates for Regional Predictions of 1 km Daily Maximum Air Temperature. *Remote Sensing*, **6**(9): 8639-8670. doi:10.3390/rs6098639
- Pebesma, E. 2012. spacetemp: Spatio-Temporal Data in R. *Journal of Statistical Software*, **51**: 7.
- Pebesma, E., & Bivand, R. S. 2005. S Classes and Methods for Spatial Data: the sp Package. *Economic Geography*, **50**: 1-21.
- Pebesma, E., Rowlingson, B., & Bivand, R. 2012. Package 'rgdal'. *R-CRAN*, **4**: 1-41.
- Perfecto, I., Rice, R., & Greenberg, R. 1996. Shade Coffee: A Disappearing Refuge for Biodiversity. *BioScience*, **46**: 598-608. doi:10.2307/1312989
- Peters, J., Baets, B. D., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. D., & Huybrechts, W. 2007. Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, **207**: 304-318. doi:10.1016/j.ecolmodel.2007.05.011
- Pielke, R. A., & Wilby, R. L. 2012. Regional climate downscaling: What's the point? *Eos*, **93**: 52-53. doi:10.1029/2012EO050008
- Porter, J. R., & Christensen, S. 2013. Deconstructing crop processes and models via identities. *Plant, Cell and Environment*, **36**: 1919-1925. doi:10.1111/pce.12107
- Ramirez-Villegas, J., Jarvis, A., & Laderach, P. 2013. Empirical approaches for assessing impacts of climate change on agriculture: The EcoCrop model and a case study with grain sorghum. *Agricultural and Forest Meteorology*, **170**: 67-78. doi:10.1016/j.agrformet.2011.09.005

- Ramirez-Villegas, J., Lau, C., Hooker, J., Jarvis, A., Ann-Kristin, K., Arnell, N., Tom, O. 2011. Climate Analogues: finding tomorrow's agriculture today. *CGIAR Research Program on Climate Change, Agriculture and Food Security*.
- Rérolle, V. M. C., Floquet, C. F. A., Mowlem, M. C., Connelly, D. P., Achterberg, E. P., & Bellerby, R. R. G. J. 2011. Seawater-pH measurements for ocean-acidification observations. *TrAC - Trends in Analytical Chemistry*, **40**: 146-157 doi:10.1016/j.trac.2012.07.016
- Resources, ESRI. 2005. Understanding cokriging. Retrieved February 15, 2015, from <http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?TopicName=Understanding> cokriging
- Rosenzweig, C., Iglesias, A., Yang, X., Epstein, P. R., & Chivian, E. 2001. Climate change and extreme weather events - implications for food production, plant diseases, and pests. *Change Human Health*, **2**: 90-104. doi:10.1097/JOM.0b013e31817d32da
- Rossiter, D. G. 2007. Technical Note: Co-kriging with the gstat package of the R environment for statistical computing. Retrieved from <http://spatial-analyst.net/book/node/294>
- Ruiz-Garcia, L., Lunadei, L., Barreiro, P., & Robla, J. I. 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**: 4728-4750. doi:10.3390/s90604728
- Sanchez, B., Rasmussen, A., & Porter, J. R. 2014. Temperatures and the growth and development of maize and rice: a review. *Global Change Biology*, **20**(2): 408-417. doi:10.1111/gcb.12389
- Schiermeier, Q. (2010). *The real holes in climate science*. Nature, **463**: 284-287. doi:10.1038/463284a
- Seguinot, J., Khourlev, C., Rogozina, I., Stroeve, A. P., & Zhang, Q. 2014. The effect of climate forcing on numerical simulations of the Cordilleran ice sheet at the Last Glacial Maximum. *The Cryosphere*, **8**: 1087-1103. doi:10.5194/tc-8-10871103
- Sellers, P. J. 1985. Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, **6**: 1335-1372. doi:10.1080/01431168508948283
- Selvaraju, R. 2012. Climate risk assessment and management in agriculture. In: *Building resilience for adaptation to climate change in the agriculture sector: Proceedings of a Joint FAO/OECD Workshop*. FAO. Retrieved from <http://www.fao.org/3/a-i3084e>
- Singh, S. 2012. E-Agriculture and Rural Development. *E-Agriculture and Rural Development: Global Innovations and Future Prospects* (pp. 140168). doi:10.4018/978-1-4666-2655-3
- Staver, C., Guharay, F., Monterroso, D., & Muschler, R. G. 2001. Designing pest-suppressive multistrata perennial crop systems: Shade-grown coffee in central america. *Agroforestry Systems*, **53**: 151-170. doi:10.1023/A:1013372403359
- Tarara, J. M., & Hoheisel, G. A. 2007. Low-cost shielding to minimize radiation errors of temperature sensors in the field. *HortScience*, **42**: 1372-1379.
- Taugourdeau, S., le Maire, G., Avelino, J., Jones, J. R., Ramirez, L. G., Jara Quesada, M., Roupsard, O. 2014. Leaf area index as an indicator of ecosystem services and management practices: An application for coffee agroforestry. *Agriculture, Ecosystems and Environment*, **192**: 19-37. doi:10.1016/j.agee.2014.03.042
- Thomas, C. K., & Smoot, A. R. 2013. An effective, economic, aspirated radiation shield for air temperature observations and its spatial gradients. *Journal of Atmospheric and Oceanic Technology*, **30**: 526-537. doi:10.1175/JTECH-D-12-00044.1 University, Duke. (n.d.). *Whats the bottom line? How to compare models*. Retrieved February 15, 2015, from <http://people.duke.edu/~rnau/compare.htm>
- 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**: 102-110. doi:10.1111/j.1759-5436.2011.00240.x
- Van Marken Lichtenbelt, W. D., Daanen, H. A. M., Wouters, L., Fronczek, R., Raymann, R. J. E. M., Severens, N. M. W., & Van Someren, E. J. W. 2006. Evaluation of wireless determination of skin temperature using iButtons. *Physiology and Behavior*, **88**: 489-497. doi:10.1016/j.physbeh.2006.04.026
- Viswanadham, N. n.d.. Food Security in India: A Logistics and Supply Chain Challenge. Bangalore. Retrieved from <http://drona.csa.iisc.ernet.in/~nv/37FoodSecurityinIndia.pdf>
- Von Arx, G., Dobbertin, M., & Rebetez, M. 2012. Spatio-temporal effects of forest canopy on understory microclimate in a long-term experiment in Switzerland. *Agricultural and Forest Meteorology*, **166-167**: 144-155. doi:10.1016/j.agrformet.2012.07.018
- Walvoort, D., Brus, D., & Gruijter, J. de. 2013. Package spcosa. R. Retrieved from <http://cran.r-project.org/web/packages/spcosa/spcosa.pdf>
- 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**: 1261-1267. doi:10.1016/j.cageo.2010.04.005
- Wang, E., Smith, C. J., Bond, W. J., & Verburg, K. 2004. Estimations of vapour pressure deficit and crop water demand in APSIM and their implications for prediction of crop yield, water use, and deep drainage. *Australian Journal of Agricultural Research*, **55**(12): 1227-1240.  
doi:<http://dx.doi.org/10.1071/AR03216>
- 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**: 41-76.
- Welles, J. M., & Cohen, S. 1996. Canopy structure measurement by gap fraction analysis using commercial instrumentation. *Journal of Experimental Botany*, **47**: 1335-1342. doi:10.1093/jxb/47.9.1335
- Wiesmeier, M., Barthold, F., Blank, B., & Kögel-Knabner, I. 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil*, **340**: 7-24. doi:10.1007/s11104-010-0425-z
- Wilks, D. S., & Shen, K. W. 1991. Threshold Relative Humidity Duration Forecasts for Plant Disease Prediction. *Journal of Applied Meteorology*, **30**: 463-477. doi:10.1175/1520-0450(1991)030
- Wilks, D. S., & Wilby, R. L. 1999. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, **23**(3): 329-357 doi:10.1191/030913399666525256
- Winslow, J. C., Hunt, E. R., & Piper, S. C. 2001. A globally applicable model of daily solar irradiance estimated from air temperature and precipitation data. *Ecological Modelling*, **143**: 227-243. doi:10.1016/S0304-3800(01)00341-6
- Wood, S. A., Jina, A. S., Jain, M., Kristjanson, P., & DeFries, R. S. 2014. Smallholder farmer cropping decisions related to climate variability across multiple regions. *Global Environmental Change*, **25**: 163-172. doi:10.1016/j.gloenvcha.2013.12.011
- World Meteorological Organization. 2008. Guide of Meteorological Instruments and Methods of Observation. In: *WMO* (Vol. I & II, pp. I.151 to I.151II.210 to II.211).
- Yu, K., Lu, Z., & Stander, J. 2003. Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society Series D: The Statistician*, **52**(3): 331-350. doi:10.1111/1467-9884.00363
- Zeileis, A., & Grothendieck, G. 2005. zoo: An S3 Class and Methods for Indexed Totally Ordered Observations. Retrieved from <http://cran.r-project.org/web/packages/zoo/index.html>

## 9 Appendices

### Appendix A. Cost of materials (sensor and shield)

**Table 12:** Cost of different sensors (US\$) when bought in different quantities

Type	Units	Memory	10-24	25-49	50-99	100-249	250-499	500-999
DS1921G-F5	T	2048	\$31.50	\$29.49	\$25.00	\$22.78	\$20.78	\$19.99
DS1922L-F5	T	8192	-	\$59.00	\$39.00	\$38.00	\$35.00	\$33.00
DS1922T-F5	T	8192	\$79.00	\$74.99	\$64.99	\$61.99	\$49.99	\$45.99
DS1923-F5	T+RH	8192	\$109.00	\$98.00	\$89.00	\$79.95	\$74.95	\$70.95

**Table 13:** Cost of materials for the experiments and complete network

Material	Unit	Price per unit CRc	Price per unit US Dollar	Bought units	Total cost CR colones	Total cost US Dollar
PVC pipe 50mm	6 meters	7,667.54	14.34	21 meters	26,836.39	50.20
PVC elbow 50mm	1 elbow	1,029.91	1.93	20 elbows	20,598.20	38.53
PVC pipe 25mm*	6 meters	6,414.00	12.00	210 meters	224,490.00	419.91
PVC elbow 25mm*	1 elbow	749.44	1.49	210 elbows	157,382.40	294.38
Insulating foil	1 meter	3,176.99	5.94	2 meters	6,353.98	11.89
Mosquito mesh	1 meter	366.53	0.69	2 meters	733.06	1.37
Reflective tape	1 meter	1,523.79	2.85	20 meters	30,475.80	57.01
White paint spray	1 can	1,911.50	3.58	1 can	1,911.50	3.58
Galvanized roof	1 meter	1,473.20	2.76	1 meter	1,473.20	2.76
Roofing tape	1 meter	1,551.15	2.90	10 meter	15,511.50	29.01
Miscellaneous (tape, marker, brush)	-	-	-	1 brush, 1 marker, 90 meter tape	10,687.03	19.99
<b>Total (experiments)</b>					123,027.60	230.12
<b>Total (network materials)</b>					382,405.50	715.29
<b>Total (network materials + 100 sensors)</b>					<b>4,656,663.99</b>	<b>8,710.29</b>

\* As receipts of these purchases were lost, the costs have been estimated based on similar materials available in Costa Rica.

## Appendix B. linear models of the change in temperature over time

**Table 14:** Linear model for to predict change in temperature over time (1 hour) in the PVC shields

Shield type	Number of measurements	Linear model to predict $T_{pvc}$	$R^2$ of model
<i>Control</i>	6,900	$1.52 \times T_{stevenson}$	0.87
<i>Holes</i>	5,520	$1.42 \times T_{stevenson}$	0.86
<i>Insulating foil</i>	5,520	$1.43 \times T_{stevenson}$	0.92
<i>Foil with holes</i>	5,520	$1.46 \times T_{stevenson}$	0.89
<i>Reflective tape</i>	2,760	$1.46 \times T_{stevenson}$	0.85

**Table 15:** Linear model for to predict change in relative humidity over time (1 hour)

Shield type	Number of measurements	Linear model to predict $RH_{pvc}$	$R^2$ of model
<i>Control</i>	5,520	$1.45 \times RH_{stevenson}$	0.66
<i>Holes</i>	4,140	$1.22 \times RH_{stevenson}$	0.58
<i>Insulating foil</i>	4,140	$1.47 \times RH_{stevenson}$	0.69
<i>Foil with holes</i>	4,140	$1.52 \times RH_{stevenson}$	0.69
<i>Reflective tape</i>	1,380	$1.59 \times RH_{stevenson}$	0.61

## Appendix C. Humidity statistics in the different PVC shields

**Table 16:** Humidity statistics in different (50mm) sensor shields

	Sunny day				Cloudy day			
	Mean	Min	Max	St.dev	Mean	Min	Max	St.dev
<i>Station</i>	85.56	65.97	98.15	1.29	92.08	73.82	98.11	0.80
<i>White paint</i>	83.17	48.26	101.35	1.09	95.79	62.72	101.84	1.05
<i>Holes (30%)</i>	84.18	52.19	102.21	1.10	98.57	56.98	104.91	3.06
<i>Insulating foil</i>	83.61	50.69	104.23	2.57	94.05	63.74	100.77	1.66
<i>Holes + foil</i>	82.23	50.50	100.58	1.17	95.79	62.08	105.83	2.47
<i>Reflective tape</i>	82.90	43.79	101.91	1.88	96.08	75.41	102.24	1.57

**Table 17:** Humidity statistics in different (25mm) sensor shields

	Sunny day				Cloudy day			
	Mean	Min	Max	St.dev	Mean	Min	Max	St.dev
<i>Station</i>	96.65	92.76	99.67	0.71	98.13	77.57	114.42	4.07
<i>Nothing</i>	99.72	96.49	102.19	0.52	88.66	46.67	101.91	1.07
<i>Holes (30%)</i>	99.83	96.30	103.08	0.56	89.14	50.37	102.35	0.82
<i>Insulating foil</i>	101.48	96.25	105.86	1.64	89.20	53.24	102.75	1.00
<i>Holes + foil</i>	99.92	95.86	102.62	0.53	88.79	49.36	101.90	0.79

## Appendix D. Cross-validation of different calibration options

**Table 18:** Cross-validation of change-in-time calibration

Removed day in validation	Model slope	Model $R^2$	Parameter 1 (multiplication)	Parameter 2 (day-adjust)	Parameter 3 (night-adjust)	RMSE (°C)
1	0.55	0.87	1.38	0.11	0.25	0.23
2	0.59	0.89	1.30	0.11	0.21	0.44
3	0.55	0.86	1.40	0.09	0.20	0.24
4	0.56	0.88	1.35	0.11	0.24	0.08
5	0.57	0.88	1.36	0.11	0.21	0.34
<i>None</i>	0.56	0.88	1.35	0.11	0.22	0.26

**Table 19:** Cross-validation of the polynomial temperature calibration

Removed day in validation	Model $R^2$	Coefficient 1 (Intercept)	Coefficient 2 (x)	Coefficient 3 (x2)	Residual standard error
1	0.96	-10.08	2.07	-0.03	0.45
2	0.97	-14.13	2.40	-0.03	0.33
3	0.96	-9.27	2.00	-0.03	0.46
4	0.96	-9.70	2.03	-0.03	0.47
5	0.97	-10.60	2.09	-0.03	0.39
<i>None</i>	0.96	-10.56	2.10	-0.03	0.43

## Appendix E. Calibrated temperature in different quantiles (3 sensors)

**Table 20:** Temperature in different quantiles for a range of calibration-types

Temperature in different quantile										
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
No calibration										
14.58	16.45	17.02	17.71	18.38	19.23	20.44	22.14	25.06	28.36	34.81
15.13	16.83	17.39	17.95	18.65	19.39	20.27	21.58	23.27	24.96	28.08
14.11	16.15	16.92	17.62	18.33	19.27	20.51	22.90	25.86	29.18	33.94
Linear calibration										
16.88	18.17	18.57	19.05	19.52	20.11	20.95	22.13	24.17	26.47	30.95
17.26	18.44	18.83	19.22	19.71	20.22	20.83	21.74	22.92	24.10	26.27
16.55	17.97	18.50	18.99	19.48	20.13	21.00	22.67	24.72	27.03	30.35
Exponential calibration										
14.30	16.55	17.20	17.97	18.68	19.57	20.73	22.27	24.52	26.53	28.76
14.99	16.98	17.62	18.23	18.97	19.72	20.58	21.78	23.20	24.46	26.38
13.70	16.20	17.09	17.88	18.63	19.60	20.81	22.90	25.06	26.84	28.67
Change in time - no parameters										
14.58	16.45	17.00	17.63	18.22	18.94	19.55	20.50	21.60	23.38	27.02
15.13	16.86	17.39	17.89	18.45	19.01	19.52	20.15	20.90	21.67	23.70
14.11	16.15	16.90	17.55	18.15	18.92	19.60	20.66	21.90	23.79	26.83
Change in time - with parameters										
14.80	16.66	17.20	17.88	18.49	19.25	20.13	21.31	23.01	25.66	30.58
15.35	17.05	17.55	18.11	18.78	19.39	20.09	20.94	22.00	23.15	25.80
14.33	16.37	17.12	17.79	18.40	19.19	20.24	21.92	23.56	26.28	30.13

## Appendix F. Zonal statistics for temperature at Aquiares

**Table 21:** Mean, minimum and maximum temperature at different altitude ranges in Aquiares

	Mean	Minimum	Mean
<i>WorldClim data for August (current data)</i>			
<900m	21.73	17.03	26.51
900-1,000m	21.51	16.78	26.31
1,000-1,100m	21.19	16.43	25.99
1,100-1,200m	20.70	15.91	25.50
1,200-1,300m	20.11	15.34	24.92
1,300-1,400m	19.73	14.98	24.53
1,400-1,500m	-	-	-
<i>Linear calibration</i>			
<900m	21.92	18.65	31.34
900-1,000m	21.53	17.25	30.11
1,000-1,100m	21.24	16.88	29.61
1,100-1,200m	20.92	16.56	29.53
1,200-1,300m	20.61	16.19	29.29
1,300-1,400m	20.31	16.15	28.61
1,400-1,500m	20.18	16.39	27.19
<i>Exponential calibration</i>			
<900m	21.44	18.56	28.81
900-1,000m	20.95	17.95	28.36
1,000-1,100m	20.55	17.47	28.21
1,100-1,200m	20.13	16.99	28.20
1,200-1,300m	19.70	16.51	28.08
1,300-1,400m	19.28	16.22	27.72
1,400-1,500m	19.20	16.00	27.04
<i>Change in time - no parameters</i>			
<900m	20.41	15.52	27.97
900-1,000m	19.89	15.22	26.76
1,000-1,100m	19.50	14.97	26.27
1,100-1,200m	19.05	14.82	25.90
1,200-1,300m	18.64	14.41	25.54
1,300-1,400m	18.36	14.54	25.22
1,400-1,500m	18.33	14.72	25.02
<i>Change in time - with parameters</i>			
<900m	21.23	15.74	31.44
900-1,000m	20.68	15.44	29.92
1,000-1,100m	20.28	15.19	29.32
1,100-1,200m	19.83	15.03	29.05
1,200-1,300m	19.70	14.63	28.72
1,300-1,400m	19.05	14.76	28.02
1,400-1,500m	18.95	14.94	26.78

## Appendix G. Strength of temperature-explaining models and relative contribution of covariates

**Table 22:** Model strength and relatively contribution of covariates to explaining temperature

Hour	R <sup>2</sup>	Elevation	Slope	Aspect	LAI	Canopy	Radiation
1	0.53	0.77	0.11	0.05	0.02	0.04	-
2	0.54	0.77	0.12	0.06	0.02	0.03	-
3	0.50	0.71	0.14	0.09	0.02	0.03	-
4	0.52	0.72	0.14	0.09	0.02	0.03	-
5	0.50	0.72	0.13	0.10	0.02	0.03	-
6	0.57	0.63	0.06	0.07	0.01	0.02	0.21
7	0.50	0.45	0.06	0.10	0.03	0.04	0.32
8	0.41	0.45	0.06	0.08	0.04	0.07	0.30
9	0.36	0.42	0.04	0.07	0.06	0.10	0.31
10	0.33	0.50	0.05	0.05	0.07	0.18	0.16
11	0.28	0.59	0.10	0.03	0.09	0.14	0.05
12	0.24	0.45	0.14	0.06	0.16	0.09	0.10
13	0.28	0.51	0.10	0.08	0.08	0.06	0.17
14	0.34	0.44	0.10	0.07	0.03	0.06	0.30
15	0.48	0.56	0.08	0.05	0.03	0.06	0.23
16	0.62	0.67	0.05	0.03	0.02	0.05	0.16
17	0.82	0.80	0.06	0.02	0.03	0.05	0.05
18	0.83	0.86	0.05	0.03	0.02	0.04	-
19	0.80	0.86	0.04	0.03	0.02	0.04	-
20	0.77	0.86	0.04	0.04	0.02	0.04	-
21	0.72	0.85	0.04	0.05	0.02	0.04	-
22	0.67	0.83	0.06	0.05	0.02	0.04	-
23	0.62	0.81	0.08	0.06	0.02	0.03	-
24	0.54	0.75	0.12	0.08	0.01	0.03	-
<b>Average</b>	<b>0.53</b>	<b>0.67</b>	<b>0.08</b>	<b>0.06</b>	<b>0.04</b>	<b>0.06</b>	<b>0.10</b>

## Appendix H. Errors in Ordinary and Universal Kriging of temperature

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
<i>80 sensors</i>	0.69 (0.18-2.21)	0.53 (0.41-0.67)	2.52 (0.59-8.12)	0.83 (0.57-1.12)	0.53 (0.40-0.70)	2.52 (1.97-3.27)
<i>40 sensors</i>	0.77 (0.24-2.17)	0.61 (0.42-0.87)	2.85 (1.01-7.79)	0.87 (0.44-1.35)	0.59 (0.32-0.87)	2.85 (1.60-4.18)
<i>20 sensors</i>	0.73 (0.32-1.57)	0.56 (0.38-0.77)	2.57 (1.19-5.34)	0.82 (0.40-1.65)	0.54 (0.29-0.89)	2.57 (1.50-4.33)
<i>8 sensors</i>	0.95 (0.61-1.37)	0.69 (0.28-1.26)	3.30 (2.54-4.32)	0.88 (0.25-2.07)	0.68 (0.21-1.63)	3.30 (1.03-8.32)

**Table 23:** Errors with Ordinary Kriging

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
<i>80 sensors</i>	0.64 (0.15-2.23)	0.49 (0.39-0.63)	2.35 (0.52-8.17)	0.79 (0.56-1.07)	0.49 (0.39-0.63)	2.35 (1.88-2.99)
<i>40 sensors</i>	0.70 (0.21-2.19)	0.54 (0.41-0.80)	2.51 (0.73-7.89)	0.81 (0.41-1.33)	0.52 (0.28-0.80)	2.51 (1.34-3.70)
<i>20 sensors</i>	0.70 (0.28-1.60)	0.55 (0.40-0.83)	2.42 (1.09-5.32)	0.78 (0.40-1.61)	0.51 (0.27-0.90)	2.42 (1.22-4.24)
<i>8 sensors</i>	0.82 (0.37-1.41)	0.57 (0.30-0.83)	2.47 (1.54-3.88)	0.78 (0.14-1.97)	0.53 (0.12-1.22)	2.47 (0.59-5.68)

**Table 24:** Errors with Universal Kriging (Elevation)

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
<i>80 sensors</i>	0.64 (0.14-2.21)	0.49 (0.39-0.63)	2.35 (0.50-8.22)	0.78 (0.56-1.04)	0.49 (0.39-0.63)	2.35 (1.87-3.05)
<i>40 sensors</i>	0.70 (0.22-2.24)	0.55 (0.41-0.81)	2.57 (0.73-8.04)	0.82 (0.44-1.29)	0.53 (0.31-0.81)	2.57 (1.52-3.77)
<i>20 sensors</i>	0.71 (0.29-1.60)	0.57 (0.37-0.83)	2.45 (1.07-5.23)	0.78 (0.38-1.50)	0.51 (0.25-0.88)	2.45 (1.22-4.34)
<i>8 sensors</i>	0.84 (0.37-1.44)	0.57 (0.31-0.83)	2.55 (1.53-4.08)	0.81 (0.16-2.05)	0.55 (0.11-1.32)	2.47 (0.59-5.68)

**Table 25:** Errors with automatized Universal Kriging (Elevation)

## Appendix I. Errors in different Dynamic Kriging approaches

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
<i>80 sensors</i>	0.63 (0.15-2.03)	0.48 (0.37-0.63)	2.31 (0.54-7.37)	0.77 (0.54-0.97)	0.49 (0.37-0.63)	2.31 (1.82-3.01)
<i>40 sensors</i>	0.69 (0.21-2.05)	0.53 (0.40-0.73)	2.48 (0.79-7.18)	0.80 (0.46-1.26)	0.52 (0.33-0.76)	2.48 (1.59-3.73)
<i>20 sensors</i>	0.76 (0.28-1.70)	0.55 (0.35-0.76)	2.54 (1.22-5.68)	0.84 (0.31-1.78)	0.54 (0.24-0.99)	2.54 (1.23-4.32)
<i>8 sensors</i>	1.33 (0.62-2.37)	0.91 (0.57-1.41)	3.53 (2.17-5.90)	1.19 (0.23-5.63)	0.77 (0.15-2.91)	3.53 (0.84-13.61)

**Table 26:** Errors with basic dynamic kriging

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
<i>80 sensors</i>	0.64 (0.14-2.21)	0.49 (0.39-0.63)	2.35 (0.50-8.22)	0.78 (0.56-1.04)	0.49 (0.39-0.63)	2.35 (1.87-3.05)
<i>40 sensors</i>	0.70 (0.22-2.24)	0.55 (0.41-0.82)	2.57 (0.73-8.04)	0.82 (0.44-1.29)	0.53 (0.31-0.81)	2.57 (1.52-3.77)
<i>20 sensors</i>	0.71 (0.29-1.60)	0.57 (0.37-0.83)	2.45 (1.07-5.23)	0.78 (0.38-1.50)	0.51 (0.25-0.88)	2.45 (1.12-4.35)
<i>8 sensors</i>	0.84 (0.37-1.44)	0.57 (0.31-0.83)	2.55 (1.54-4.08)	0.81 (0.16-2.05)	0.55 (0.11-1.32)	2.55 (0.54-6.18)

**Table 27:** Errors with automatized dynamic kriging

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
<i>80 sensors</i>	0.64 (0.19-2.12)	0.50 (0.42-0.58)	2.37 (0.69-7.86)	0.78 (0.60-1.00)	0.50 (0.41-0.65)	2.39 (2.00-3.04)

**Table 28:** Errors with random forest kriging

## Appendix J. Errors in different types of Co-Kriging

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
80 sensors	0.67 (0.14-2.35)	0.51 (0.37-0.68)	2.42 (0.50-8.17)	0.85 (0.57-1.17)	0.51 (0.37-0.69)	2.42 (1.77-3.21)
40 sensors	0.75 (0.21-2.31)	0.58 (0.45-0.80)	2.66 (0.80-8.15)	0.89 (0.53-1.36)	0.56 (0.33-0.84)	2.66 (1.56-4.16)
20 sensors	0.77 (0.28-1.70)	0.60 (0.42-0.90)	2.59 (1.18-5.51)	0.88 (0.42-1.70)	0.55 (0.31-1.06)	2.59 (1.55-5.19)
8 sensors	1.11 (0.46-2.05)	0.65 (0.34-1.16)	3.02 (1.74-5.16)	1.02 (0.20-3.52)	0.66 (0.16-2.07)	3.02 (0.78-9.59)

**Table 29:** Errors with co-kriging (elevation only)

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
80 sensors	0.66 (0.14-2.31)	0.50 (0.37-0.71)	2.37 (0.50-7.81)	0.83 (0.54-1.16)	0.50 (0.34-0.71)	2.37 (1.66-3.32)
40 sensors	0.73 (0.21-2.15)	0.56 (0.42-0.79)	2.60 (0.81-7.52)	0.86 (0.46-1.38)	0.55 (0.32-0.81)	2.60 (1.54-3.99)
20 sensors	0.75 (0.27-1.66)	0.59 (0.41-0.87)	2.51 (1.14-5.33)	0.85 (0.30-1.58)	0.54 (0.24-1.03)	2.51 (1.24-5.04)
8 sensors	1.14 (0.46-2.46)	0.61 (0.16-1.39)	2.96 (1.70-5.51)	0.98 (0.11-4.08)	0.63 (0.09-2.28)	2.96 (0.44-10.67)

**Table 30:** Errors of co-kriging (with elevation and radiation)

	Spatial			Temporal		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
80 sensors	0.66 (0.14-2.31)	0.49 (0.39-0.68)	2.35 (0.51-7.80)	0.82 (0.54-1.13)	0.50 (0.34-0.68)	2.35 (1.67-3.21)
40 sensors	0.72 (0.21-2.15)	0.54 (0.42-0.74)	2.52 (0.82-7.44)	0.84 (0.46-1.38)	0.53 (0.31-0.80)	2.52 (1.51-3.88)
20 sensors	0.71 (0.24-1.58)	0.54 (0.31-0.76)	2.31 (1.06-4.85)	0.80 (0.35-1.54)	0.50 (0.25-0.92)	2.31 (1.27-4.45)
8 sensors	1.05 (0.41-2.17)	0.54 (0.13-1.23)	2.63 (1.38-4.92)	0.91 (0.16-3.83)	0.58 (0.12-2.01)	2.63 (0.50-9.52)

**Table 31:** Errors with co-kriging (with elevation, radiation, slope and aspect)

## Appendix K. Errors in different types of Spatio-Temporal Kriging

	Spatial			Temporal		
	RMSE	MAE	MPE	RMSE	MAE	MPE
<i>80 sensors</i>	0.91 (0.28-1.61)	0.62 (0.14-2.43)	2.91 (0.83-6.53)	0.78 (0.22-2.92)	0.61 (0.14-2.43)	2.91 (0.76-10.50)
<i>40 sensors</i>	0.89 (0.28-1.61)	0.59 (0.14-1.64)	2.91 (0.83-6.53)	0.79 (0.20-2.82)	0.61 (0.14-2.28)	2.91 (0.75-9.87)
<i>20 sensors</i>	0.91 (0.49-1.61)	0.82 (0.21-1.66)	2.83 (1.38-4.81)	0.76 (0.17-2.90)	0.59 (0.12-2.37)	2.83 (0.62-10.76)
<i>8 sensors</i>	1.05 (0.71-1.44)	0.83 (0.65-0.90)	4.07 (2.49-5.55)	0.99 (0.41-3.01)	0.81 (0.32-2.67)	4.07 (1.47-12.06)

**Table 32:** Errors with spatio-temporal ordinary kriging

	Spatial			Temporal		
	RMSE	MAE	MPE	RMSE	MAE	MPE
<i>80 sensors</i>	0.90 (0.27-1.61)	0.61 (0.10-2.42)	2.89 (0.82-6.18)	0.77 (0.14-2.93)	0.60 (0.10-2.43)	2.89 (0.56-10.52)
<i>40 sensors</i>	0.90 (0.27-1.61)	0.58 (0.10-1.65)	2.90 (0.82-6.18)	0.78 (0.15-2.84)	0.61 (0.10-2.29)	2.90 (0.54-9.91)
<i>20 sensors</i>	0.91 (0.49-1.61)	0.83 (0.20-1.68)	2.84 (1.40-4.86)	0.76 (0.12-2.92)	0.59 (0.08-2.38)	2.84 (0.47-10.81)
<i>8 sensors</i>	1.14 (0.73-1.71)	0.91 (0.74-0.99)	4.57 (2.36-8.13)	1.11 (0.47-3.05)	0.90 (0.39-2.68)	4.57 (1.84-12.42)

**Table 33:** Errors with spatio-temporal universal kriging (elevation)

	Spatial			Temporal		
	RMSE	MAE	MPE	RMSE	MAE	MPE
<i>80 sensors</i>	0.95 (0.40-1.59)	0.64 (0.10-2.43)	3.16 (1.47-6.35)	0.82 (0.15-2.94)	0.66 (0.10-2.43)	3.16 (0.57-10.54)
<i>40 sensors</i>	0.95 (0.40-1.59)	0.64 (0.10-1.58)	3.19 (1.47-6.35)	0.83 (0.15-2.85)	0.66 (0.10-2.29)	3.19 (0.54-9.94)
<i>20 sensors</i>	0.97 (0.54-1.59)	0.92 (0.33-1.72)	3.13 (1.67-4.81)	0.81 (0.11-2.99)	0.65 (0.08-2.43)	3.13 (0.47-11.05)
<i>8 sensors</i>	1.16 (0.78-1.65)	0.98 (0.74-1.11)	4.68 (2.71-7.70)	1.13 (0.59-3.05)	0.93 (0.46-2.69)	4.68 (2.14-12.37)

**Table 34:** Errors with spatio-temporal dynamic kriging (elevation, radiation, slope and aspect)