**Experiment 1: Sensor and temporal resolution[[1]](#footnote-1)**

*1A: Sensor resolution*

A straightforward option to increase the period the sensors can be used in the field is to adjust the resolution at which the sensors make their observations. The iButton Hygrochron (DS1923)[[2]](#footnote-2) can make temperature measurements at 8- or 11-bit resolution (0.5°C or 0.0625°C), and humidity measurements at 8- or 12-bit resolution, which is either 0.6% or 0.04% RH. The readings, however, are either stored in 8- or 16-bit, with the total memory for this sensor being 8192 bytes. According to the specifications, the temperature accuracy is better than ±0.5°C for most of the range (-10°C to +65°C), while the accuracy of the humidity measurements is ±5% RH. By using a lower resolution sensor for both temperature and humidity, the sensor can store double the number of observation at the same time interval as would be possible at the higher resolution (*see table 1*). An experiment has been done with a total of 30 sensors, making observations at 5-minute interval during one week. The sensors have been placed in a certified weather station, which also houses an official observation instrument, to make sure the conditions would be the same for all sensors. Half the sensors have been set at the highest temperature and humidity resolution, while the other half has been set at the lower resolution. The sensors have been paired throughout the weather station and linear interpolation has been used to get the value at 1-minute intervals. The resulting plots from this one-week experiment can be found in *Annex 1*. From the visual analysis it is clear that the higher resolution temperature data provides a much smoother graph, while the difference for the humidity data is less clear.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of observations | Time covered at 2-hour interval | Time covered at 1-hour interval | Time covered at 30-min interval | Time covered at 15-min interval |
| 1 unit at low-res | **8192** | 683 days | 341 days | 171 days | 85 days |
| 1 unit at high-res or |  |  |  |  |  |
| 2 units at low-res | **4096** | 341 days | 171 days | 85 days | 43 days |
| 1 unit at low-res & 1 unit at high-res | **2560[[3]](#footnote-3)** | 213 days | 107 days | 53 days | 27 days |
| 2 units at high-res | **2048** | 171 days | 85 days | 43 days | 21 days |

Analysing the sensors provides more detailed information; the temperature range for the low-res sensors is 18.11 – 30.12°C, while this is 18.54 – 30.02°C for the high-res sensor; the range is 0.54°C larger for the low-res sensors. For humidity the range is 64.33 – 99.94% RH for the low-res and 63.86 – 100.23% RH for the high-res sensors[[4]](#footnote-4), making the range 0.75% higher for the high-res sensors. The mean temperature and humidity for the low-res sensors are 22.37°C and 90.45% RH; for the high-res sensors these values are 22.35°C and 90.93% RH, which is all within a very small range. Another important issue to take into account is the variability between sensors; a visual analysis shows that this is especially an issue for the low-res temperature observations. The standard deviation has been calculated for every row of the data-set (which is 1 minute), which contains 15 sensors and 10.800 observations for each sensors. After this, the mean has been taken from all the standard deviations at the row-level. The standard deviation for the low-res temperature sensors is 0.26°C, while this is < 0.05°C for the high-res sensors. For % RH, the standard deviation is 1.13 for the low-res sensors, whereas this is surprisingly higher (1.20) for the high-res sensors.

A preliminary conclusion which can be drawn from this experiment would indicate that working at a higher-resolution would not make a lot of sense for both temperature and humidity, as the reduced interval at which observations can be made will reduce the quality of the data-set more than this can be increased by the higher accuracy at higher resolution. Especially for relative humidity a high-resolution would not contribute anything. The range for the observations, as well as the standard deviation between the sensors is higher than is the case for the low-resolution. In this case, the low-resolution sensor seems to be better. For temperature, the mean is very similar (within 0.02°C), while there is a large difference between the standard deviation in both datasets. This is not necessarily a big problem, as the maximum standard deviation at any moment is less than 0.4°C. While it is already quite clear that measurements at low-res would be fine for relative humidity, the impact of the low-resolution on temperature would require study. In the case that measurements would only be made at a low-interval (1 or 2 hours), the differences might be exaggerated when using interpolation techniques to estimate the temperature at lower (e.g. 15 min) intervals. Another option that will be studied in a different experiment will be whether using the humidity sensors on all the iButtons would be useful at all – an alternative would be to only use this in one out of every *x* sensors, to be able to make more temperature measurements, while not losing a lot of information about the humidity.

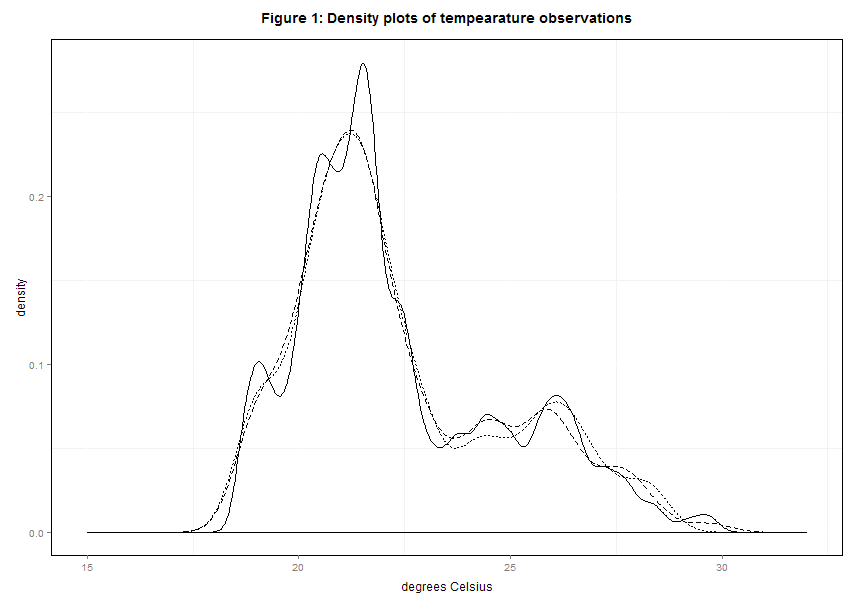
*1B: Temporal resolution*

With the sensors that are currently available, there will be certain limitations that have to be considered. The newer sensors can include *Bluetooth®* technology, but these are often developed with uses in mind that do not require a large storage capacity – in most cases the ability of having real-time data available is sufficient for the intended users. The more industrial sensors are often created with industry (and especially logistics) in mind. In this case real-time data is less important, but a large memory will help to track the conditions under which certain goods have been shipped. The sensors that have been used in this research have a quite large memory, but the conditions in the open environment will fluctuate much more than will be common in logistics. This means that hourly – or even less frequent – measurements can limit the ability to assess the extremes (which are more important than the average), if data cannot be interpolated in a proper way. When working in the field, it will be difficult to launch the iButtons at the same time (e.g. exactly at the hour), so it is likely that a project will result in a large number of datasets that all have data at hourly interval, but which will be launched at different moments throughout the hour. In order to be able to work with the data, it will be important to merge the irregular time-series into a regular one, which will require a certain process to interpolate data. The easiest will be a linear interpolation of the NA values, but this will result in values that will never be higher than the values in the relative data-set, which can cause large differences.

This experiment will initially use the high-resolution dataset that has also been used in experiment 1A, and after this the most promising interpolation techniques will also be applied to the low-resolution dataset to see whether it will provide sufficiently similar results to be able to work with only the low-resolution sensors. The original data-set (5-min interval) will be used, in which the start-moment will be chosen at random. This means no average will be taken, but only 1 value from the original data-set will be used and the rest will be ignored. Different interpolation techniques are available in the R-package {zoo}[[5]](#footnote-5), which is developed especially for irregular time-series. The techniques available here include linear (*na.approx*) and cubic spline (*na.spline*) techniques, which have limited room for adjustment. As there are many different forms of splines available, some of these options will be tested to see if they work better on the sensor data at 1-hour interval. It can be expected that linear interpolation would work find at smaller time-intervals, while there will be more smoothing at larger intervals. Splines could solve this problem, but can also increase the extreme values.

When the data is reduced to only have observations at the hour, instead of every minute, the graph continues to have the same complexity and very little data will be lost (*Annex 2 - 1*). The range for the temperature observations with the high-resolution sensor will be 18.59 – 29.85°C, while the mean will be 22.36°C. While measurements, when interpolated to 1-minute intervals surpass 30°C (which might be a common threshold), this was only the case at two observations out of the total of 151,200, which makes this negligible. The overall histogram does not change a lot, which would result in similar conclusions that will be drawn from these datasets. The standard deviation is still very small at ±0.04°C. While data at hourly interval would be sufficient for an accurate low-cost climate network, getting a good estimation of these values will require a robust interpolation technique. When only using one value for each hour, randomly selected from the original dataset (*Annex 2 – 2*), and using a linear interpolation to get the value at the exact hour, the resulting density plot will result in a smoothing of the data. The range for this dataset is 18.59 – 29.86°C, while the mean 1 remains very similar (22.36°C). The standard deviation is still only ±0.09°C. While increasing the temporal interval from 1 minute to 1 hour will result in some smoothing of the data, there does not seem to be a lot of difference between using random start-points or the closest value in the original dataset. The issue will also be analysed for the low-resolution sensors, where the standard deviation was already higher in the original dataset than it is for the high-res sensors at 1-hour interval. As visualizing all different approaches (1 vs. 2 hour, closest vs. random, and linear vs. spline), only the two ‘extremes’ have been plotted in a density plot with the original data: *1-hour, closest value, with linear interpolation* (dashed lie) and *2-hour, random value with cubic spline interpolation* (dotted line). The main difference is caused by the increase in interval.

**Figure 1: Density plots of temperature observations selected by two different approaches**



When using the low resolution, the range and mean become 18.11 - 30.11°C and 22.38°C when using the closest observation, while it will be 18.11 – 29.98°C and also 22.38°C for the random start-moment. More important is the maximum difference between any two sensors at any moment. This is lowest difference is when using the closest observation with the high-res sensor (0.65°C), followed by the low-res sensor with the closest value (1.02°C), the high-res sensor with a random start-moment (1.14°C) and finally the low-res sensor with a random moment of start (1.69°C). As all sensors are placed in the same certified weather station, values above 1°C can be considered too high. As the start moment of measuring will be very difficult to synchronize for many sensors, linear interpolation does not provide adequate results at 1-hour interval.

*Introduction to splines*

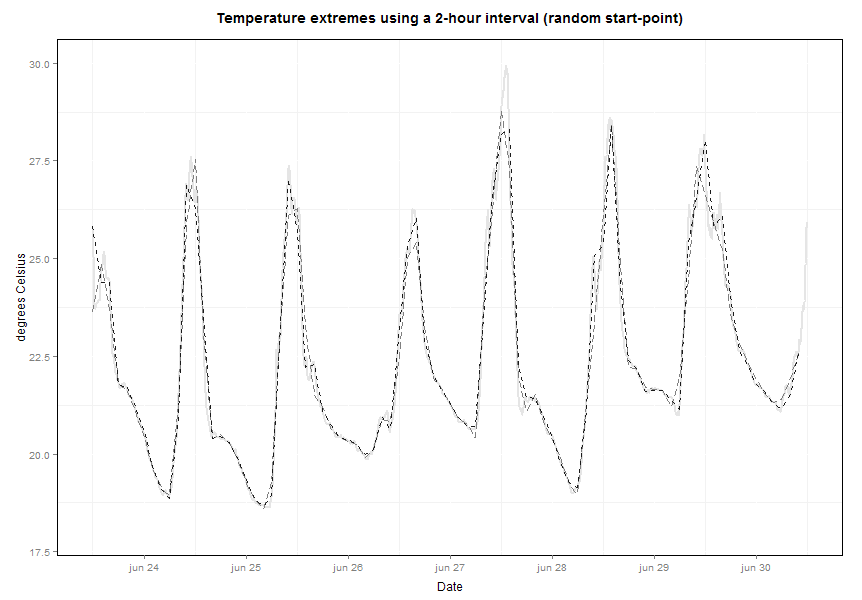
*http://www.ams.org/journals/mcom/1981-37-155/S0025-5718-1981-0616367-1/S0025-5718-1981-0616367-1.pdf*

<http://artax.karlin.mff.cuni.cz/r-help/library/pastecs/html/regspline.html>

*At this moment only the built-in cubic spline approach in the {zoo} package will be tested, but in the case more adjustment would be required (e.g. if the construction results in a certain temperature and/or humidity bias), more advanced interpolation techniques will be tried, such as can be found in the {pastecs}[[6]](#footnote-6) and {quantreg}[[7]](#footnote-7) packages.*

The graphs resulting from the built-in cubic spline function *na.spline* in the package {zoo} can be seen in *Annex 2 – 5:8*. The range when taking the closest measurement is 18.59-29.89 for the high-res sensors and 18.11-30.16 for the low-res sensors. The means is 22.36 for the high-res and 22.38 for the low-res sensors. The maximum difference at any hour is 0.34 for the high-res and 1.10 for the low-res sensors. With a random starting point, the range for the high-res sensor is 18.57-30.27°C and the mean is 22.35°C. The maximum difference at the hour, however, has gone up from 1.14 to 1.43°C. For the low-res sensor, the range becomes 18.06-30.52°C, with a mean of 22.37°C. The difference between rows stays quite similar with 1.75°C. A summary of the differences caused by these different techniques can be seen in *Annex 3*. A 2-hour interval is also included in this table to see how different techniques influence at this larger interval. The mean of al 15 sensors at high-resolution over the week of measurement for the 1-min compared to 2-hour observations (random start-point, linear and spline interpolation) is shown in *figure 2*. On some days the splines do better, while on other days the linear interpolation is closer; in all cases the maximum is lower.

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



As this is only a first step in finding the right sensor and temporal resolution – as the iButtons have been placed in a very stable environment compared to the field-situation – only some basic conclusions can be drawn at this moment. It is clear from the table that the most important factor in having values close to the original data-set is to have a fixed starting point for all sensors. This reduces the standard deviation and maximum difference at any moment. These values have been calculated from a matrix, while the other values have been calculated from a vector with all observations to find the range and distribution of the data. While some statistics are much better in the high-res dataset, especially the standard distribution and the maximum difference, the range is much smaller than is the case for the low-res dataset. The data at 2-hour interval provides similar results for much of the statistics, but for the most important information – the absolute minimum and maximum, the results are quite different. The absolute maximum is around 1°C smaller when sampling only once every two-hours, with a difference of 1.6°C when splining low-res sensors with a random starting moment. The obvious recommendation at this moment would be that creating a script that can launch all iButtons at a fixed staring moment (*a version is available for other iButtons[[8]](#footnote-8)*) would seem like a good investment – using a random starting point can provide quite useful data when using a 1-hour interval. Looking back at table 1, using temperature at high-res and humidity at low-res, the period of observation can be 107 days, while this would be 171 days when using both sensors at low-res. It has to be kept in mind that the time between launching and placing, as well as removing and reading the data have to be extracted.

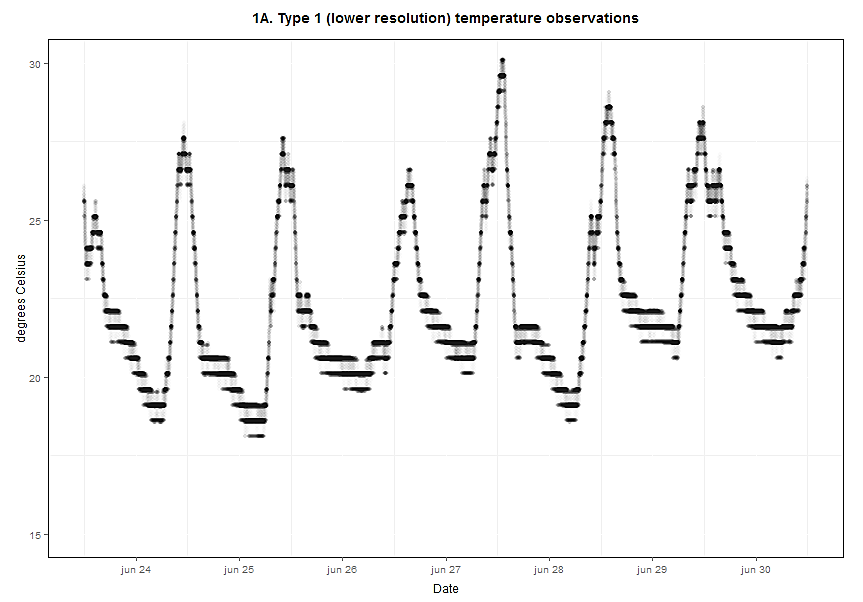
**Table 3: Minimum and maximum duration of some important food crops[[9]](#footnote-9)**

|  |  |  |  |
| --- | --- | --- | --- |
| Crop | Species | Minimum duration | Maximum duration |
| Barley | *Hordeum vulgare L.* | 90 | 240 |
| Cassava | *Manihot esculenta Crantz.* | 180 | 365 |
| Chickpea | *Cicer arietinum L.* | 90 | 180 |
| Common bean | *Phaseolus vulgaris L.* | 50 | 270 |
| Cowpea | *Vigna unguiculata Walp.* | 30 | 240 |
| Maize | *Zea mays L.* | 65 | 365 |
| Potato | *Solanum tuberosum L.* | 90 | 160 |
| Rice | *Oryza sativa L.* | 80 | 180 |
| Sorghum | *Sorhum bicolor Moench.* | 90 | 300 |
| Wheat | *Triticum aestivum L.* | 90 | 250 |

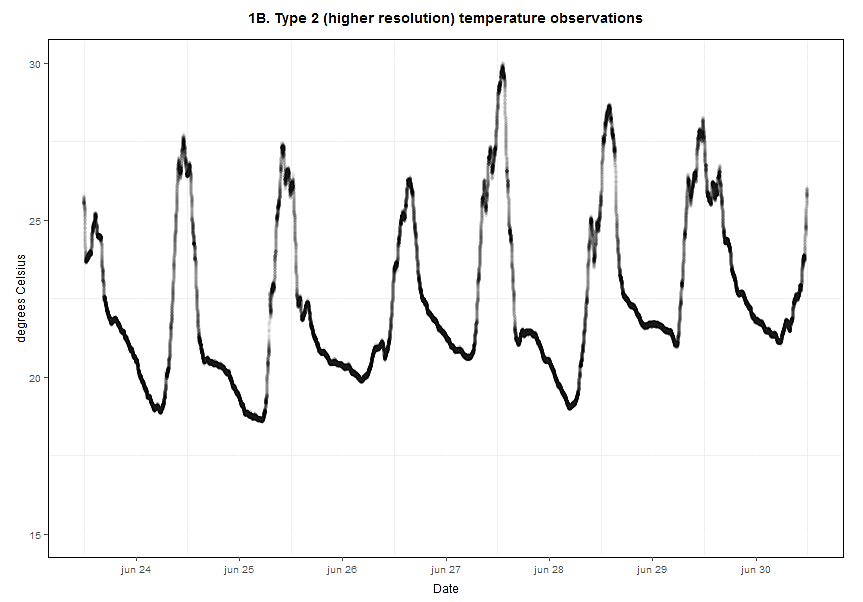
A list of some important food crops and their duration is provided in table 3. While most crops (except cassava) have a minimum duration that is shorter than the period that can be observed with the iButtons with one high- and one low-res sensor, the maximum duration is also almost always longer (except potato) than the period that would be possible with the iButtons with both sensors at low-res. As the intended projects for which a network of these sensors will be used likely include older/traditional varieties, it can be expected that the duration will be significantly larger than the minimum duration provided in the table. As using the sensors only at low-res would not be sufficient for some crops, an alternative option that will have to be studied is the usefulness of having a large number of sensors that only measure temperature at low-res, of which some will be paired to a sensor that measures humidity at low-res. This would be an option as using a 2-hour interval will result in a significant loss of information and relative humidity is often related to temperature in a certain area. The correlation between all temperature and all humidity measurements at high-resolution is -0.91, which indicates a quite strong negative correlation between the two units. Whether this correlation remains similar when the sensors will be placed over a larger area is something that will be studied as part of the analysis of the data from the sensors at the Aquiares coffee plantation in Costa Rica. Working with field-data will have several benefits; the first is that these measurements will be done with the same construction that will be used in other projects with these sensors; and the second is that measurements in the field will be done at different altitudes, which will also have an impact on temperature and humidity.

**Annex 1: Temperature & humidity observations**

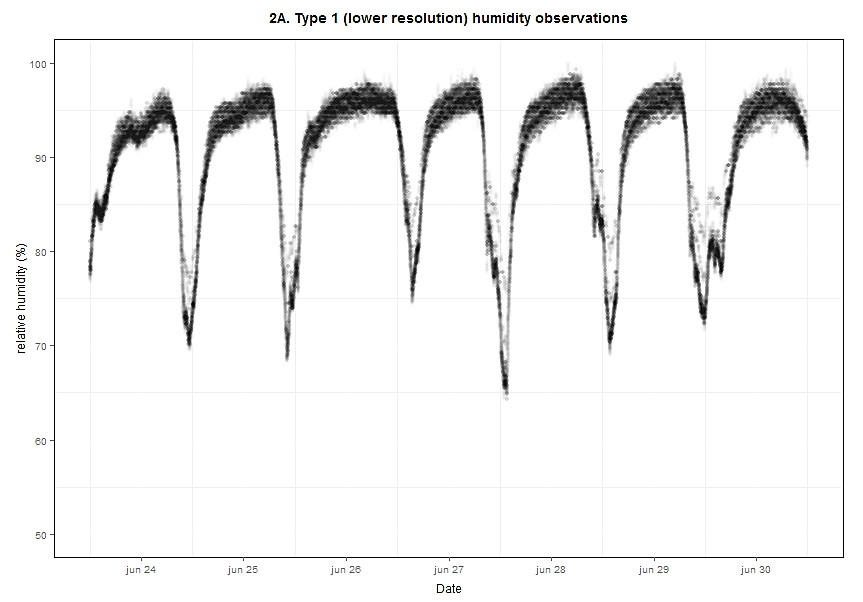
**1A. Type 1 (lower resolution) temperature observations at minute interval**



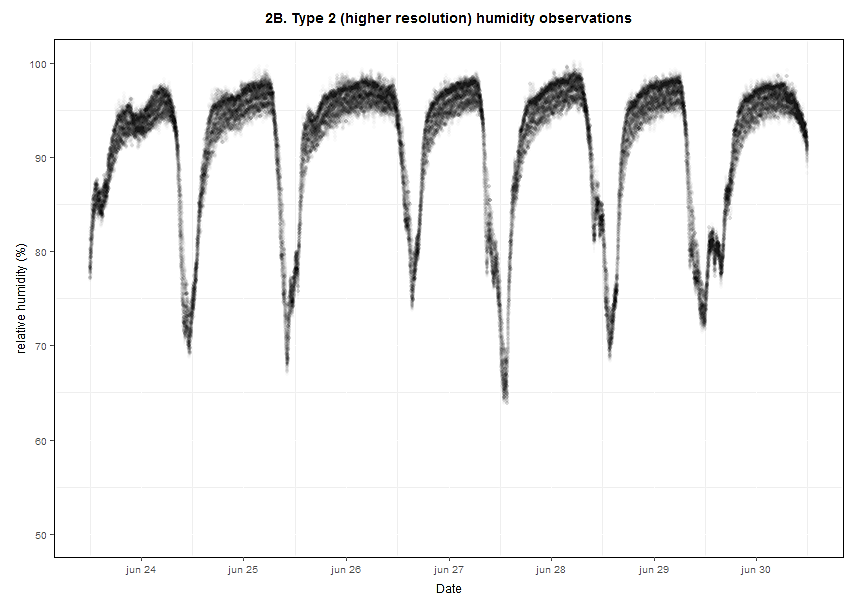
**1B. Type 2 (higher resolution) temperature observations at minute interval**



**2A. Type 1 (lower resolution) humidity observations at minute interval**

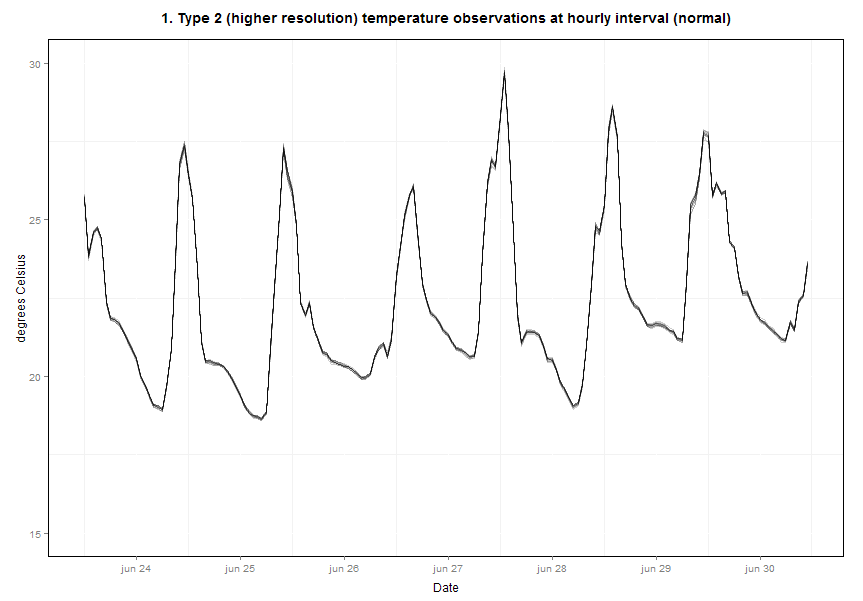


**2B. Type 2 (higher resolution) humidity observations at minute interval**

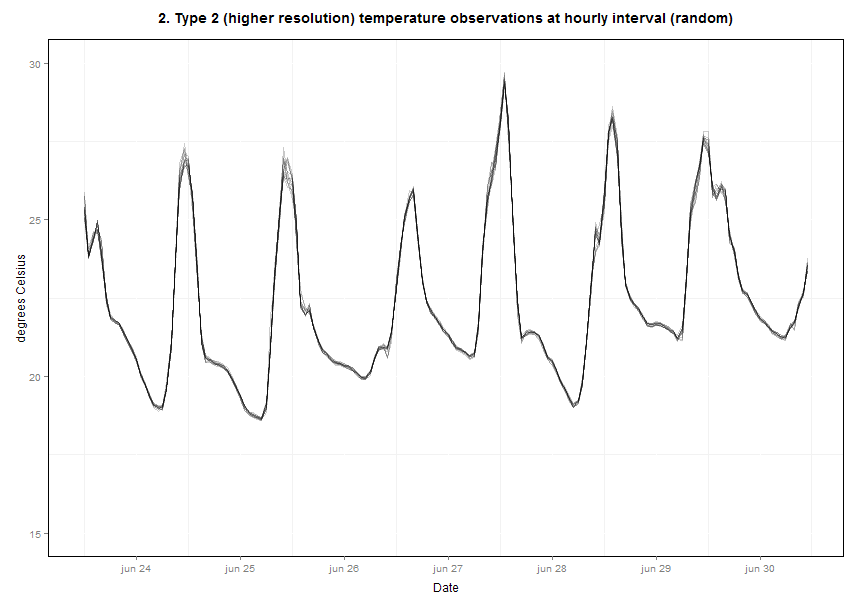


**Annex 2: Different interpolation &time-selection techniques**

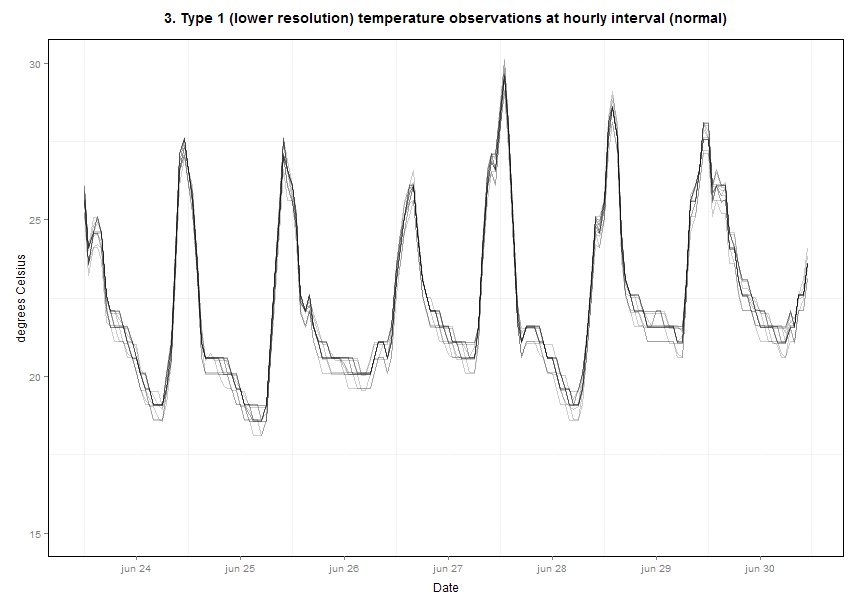
**1. Type 2 (higher resolution) temperature observations at hourly interval (normal)**



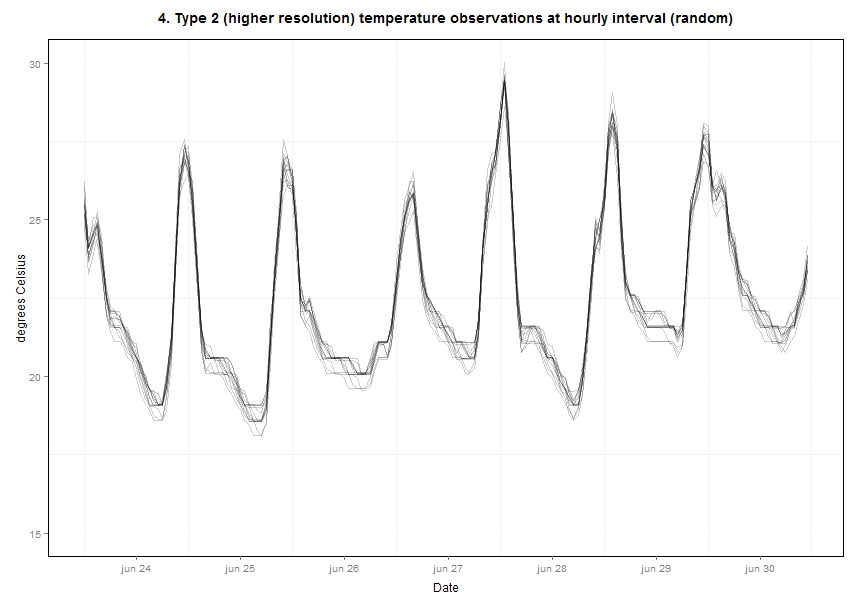
**2. Type 2 (higher resolution) temperature observations at hourly interval (random)**



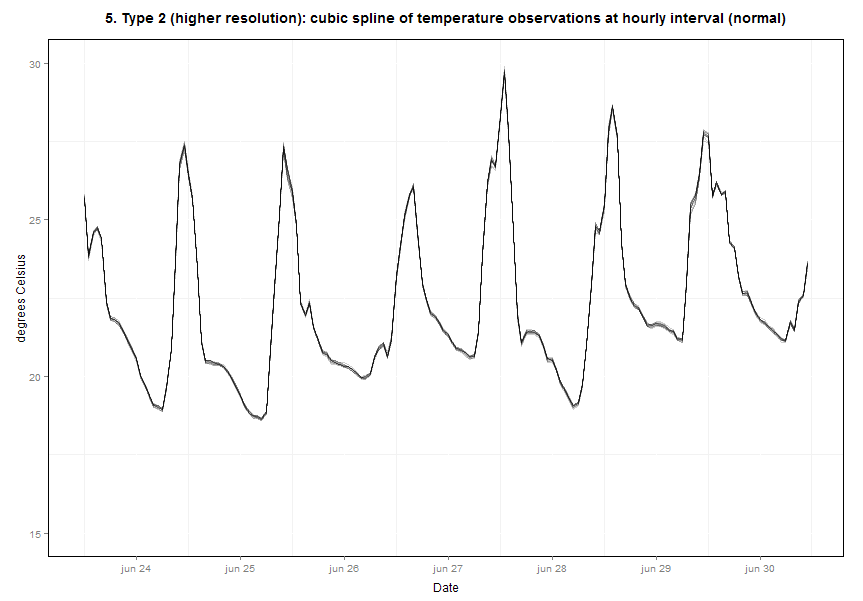
**3. Type 1 (lower resolution) temperature observations at hourly interval (normal)**



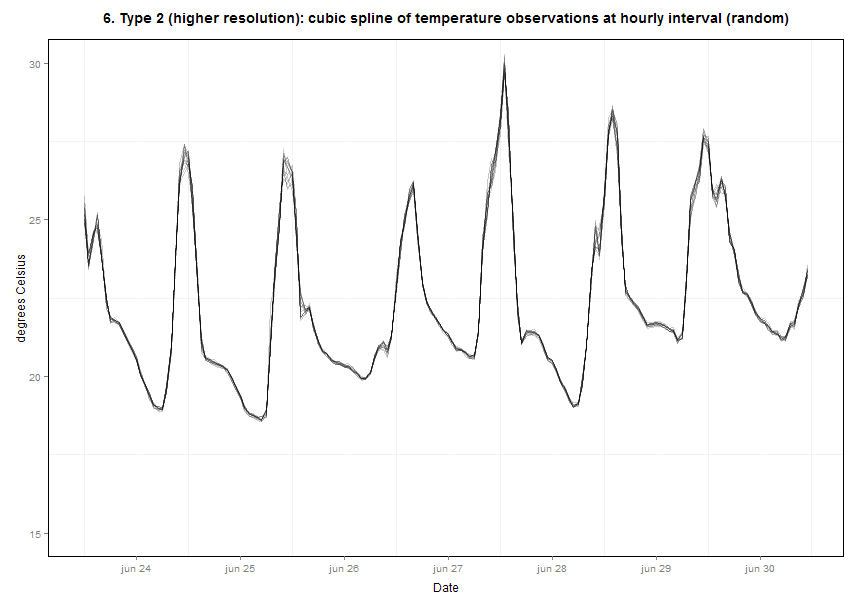
**4. Type 1 (lower resolution) temperature observations at hourly interval (random)**



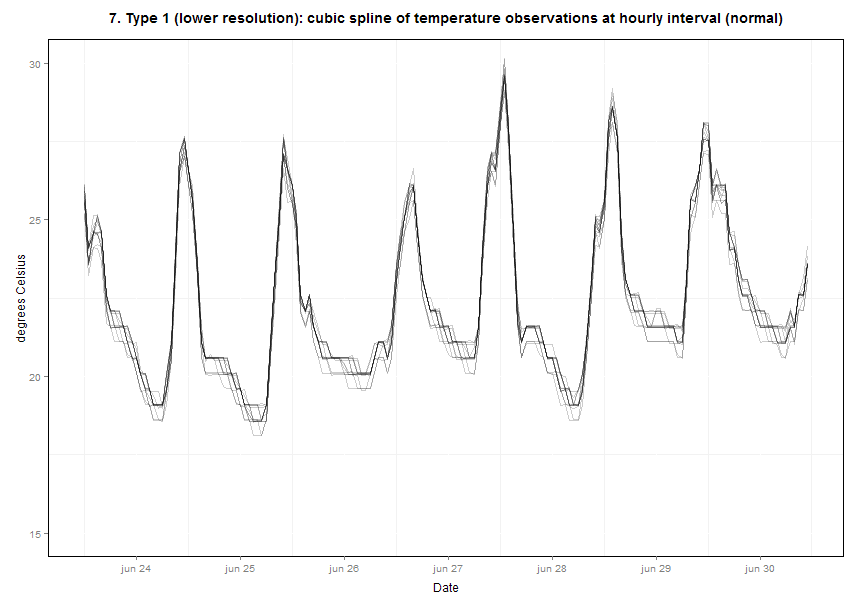
**5. Type 2 (higher resolution): cubic spline of temperature observations at hourly interval (normal)**



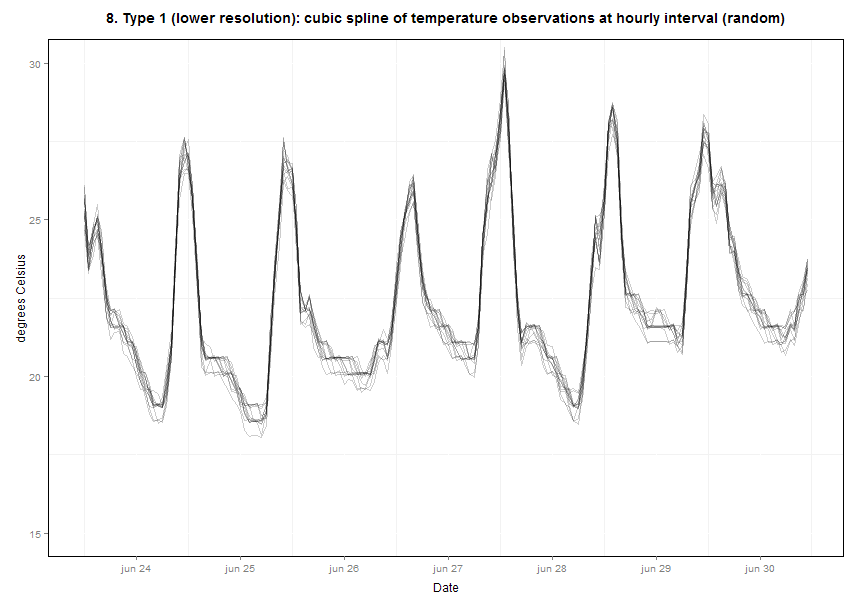
**6. Type 2 (higher resolution): cubic spline of temperature observations at hourly interval (random)**



**7. Type 1 (lower resolution): cubic spline of temperature observations at hourly interval (normal)**



**8. Type 1 (lower resolution): cubic spline of temperature observations at hourly interval (random)**



**Annex 3: Temperature statistics**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Minimum | 1st Quartile | Median | Mean | 3rd Quartile | Maximum | Standard  deviation | Maximum difference |
| **Type 1 (low-res):**  1-minute interval | 18.11 | 20.61 | 21.61 | 22.37 | 24.06 | 30.12 | 0.26 | 1.02 |
| **Closest observation** |  |  |  |  |  |  |  |  |
| 1 hour interval (linear) | 18.11 | 20.61 | 21.61 | 22.38 | 24.11 | 30.11 | 0.26 | 1.02 |
| 1 hour interval (spline) | 18.11 | 20.60 | 21.61 | 22.38 | 24.10 | 30.16 | 0.27 | 1.10 |
| 2 hour interval (linear) | 18.11 | 20.61 | 21.61 | 22.40 | 24.12 | 29.09 | 0.26 | 0.92 |
| 2 hour interval (spline) | 18.11 | 20.61 | 21.61 | 22.40 | 24.12 | 29.18 | 0.26 | 1.03 |
| **Random start-moment** |  |  |  |  |  |  |  |  |
| 1 hour interval (linear) | 18.11 | 20.61 | 21.61 | 22.37 | 24.05 | 29.85 | 0.27 | 1.49 |
| 1 hour interval (spline) | 17.99 | 20.61 | 21.61 | 22.37 | 24.00 | 30.41 | 0.29 | 1.59 |
| 2 hour interval (linear) | 18.11 | 20.61 | 21.61 | 22.38 | 24.10 | 28.65 | 0.27 | 1.52 |
| 2 hour interval (spline) | 18.07 | 20.60 | 21.63 | 22.39 | 24.10 | 28.81 | 0.29 | 1.64 |
| **Type 2 (high-res):**  1-minute interval | 18.54 | 20.58 | 21.64 | 22.35 | 23.96 | 30.02 | 0.05 | 0.65 |
| **Closest observation** |  |  |  |  |  |  |  |  |
| 1 hour interval (linear) | 18.59 | 20.59 | 21.61 | 22.36 | 24.12 | 29.85 | 0.04 | 0.35 |
| 1 hour interval (spline) | 18.59 | 20.59 | 21.61 | 22.36 | 24.13 | 29.89 | 0.05 | 0.34 |
| 2 hour interval (linear) | 18.65 | 20.59 | 21.62 | 22.38 | 24.17 | 28.67 | 0.04 | 0.35 |
| 2 hour interval (spline) | 18.65 | 20.59 | 21.62 | 22.39 | 24.17 | 28.69 | 0.05 | 0.34 |
| **Random start-moment** |  |  |  |  |  |  |  |  |
| 1 hour interval (linear) | 18.59 | 20.61 | 21.64 | 22.35 | 23.99 | 29.66 | 0.09 | 1.03 |
| 1 hour interval (spline) | 18.55 | 20.60 | 21.62 | 22.35 | 23.96 | 30.15 | 0.11 | 1.23 |
| 2 hour interval (linear) | 18.63 | 20.64 | 21.64 | 22.37 | 24.02 | 28.61 | 0.10 | 0.73 |
| 2 hour interval (spline) | 18.62 | 20.60 | 21.64 | 22.37 | 24.10 | 28.64 | 0.11 | 0.88 |

1. The script and data for this experiment can be found at <https://github.com/cornelisvd/thesis/> [↑](#footnote-ref-1)
2. <http://datasheets.maximintegrated.com/en/ds/DS1923.pdf> [↑](#footnote-ref-2)
3. This value is slightly lower than can be expected (2730 observations) with one 8 and one 16-bit observation. [↑](#footnote-ref-3)
4. It is not uncommon to have values > 100% RH for the sensors, especially on rainy days it can go over 100%. [↑](#footnote-ref-4)
5. <http://cran.r-project.org/web/packages/zoo/index.html> [↑](#footnote-ref-5)
6. http://cran.r-project.org/web/packages/pastecs/index.html [↑](#footnote-ref-6)
7. http://cran.r-project.org/web/packages/quantreg/index.html [↑](#footnote-ref-7)
8. <http://lukemiller.org/index.php/2012/03/launching-ibutton-thermochrons-with-the-help-of-r/> [↑](#footnote-ref-8)
9. Data derived from FAO’s Ecocrop database: <http://ecocrop.fao.org/ecocrop/srv/en/home> [↑](#footnote-ref-9)