Keynote 3

May 30, 2021 to June 03, 2021 - Valencia, Spain

DataSys 2021

The relevance of data

Is all the data equally important?

Lorena Parra

Researcher at: Universitat Politecnica de Valencia & Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario,

Spain

loparbo@doctor.upv.es

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

iMiDRA Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario

Index

- Introduction: data and its relevance
- Data variability in real conditions
- Relevance of data in case of images
- ¿How we can take an advantage of managing data variability properly?
- Real cases of application of data management algorithms

Index

- Introduction: data and its relevance
- Data variability in real conditions
- Relevance of data in case of images
- How we can take an advantage of managing data variability properly?
- Real cases of application of data management algorithms

Introduction: data and its relevance

In Wireless Sensor Networks (WSNs), Internet of Things (IoT), or other data gathering systems generate colossal volumes of data.

In many cases, data is gathered every few minutes in order to control the ongoing monitored processes or activities.

10 sensors X 1 data every 5 min = 2,800 data each day…

= 20,160 data each week…

= 84,000 data each month…

Introduction: data and its relevance

Considering the cost of data storage in terms of energy and the economic cost of renting the virtual space, we must evaluate the rentability of data storage.

There are several aspects to consider:

-Data periodicity (1 data every X minutes)

-Data resolution (X bits per data)

-Number of measurement devices (X nodes per point)

-Data sending (X message every X data gathered)

These values should be defined according to the future use of data.

Introduction: data and its relevance

Data periodicity (1 data every X minutes)

We need to define with the final user of the data the periodicity of data. These will be determined according to the monitored variable and the monitored environment.

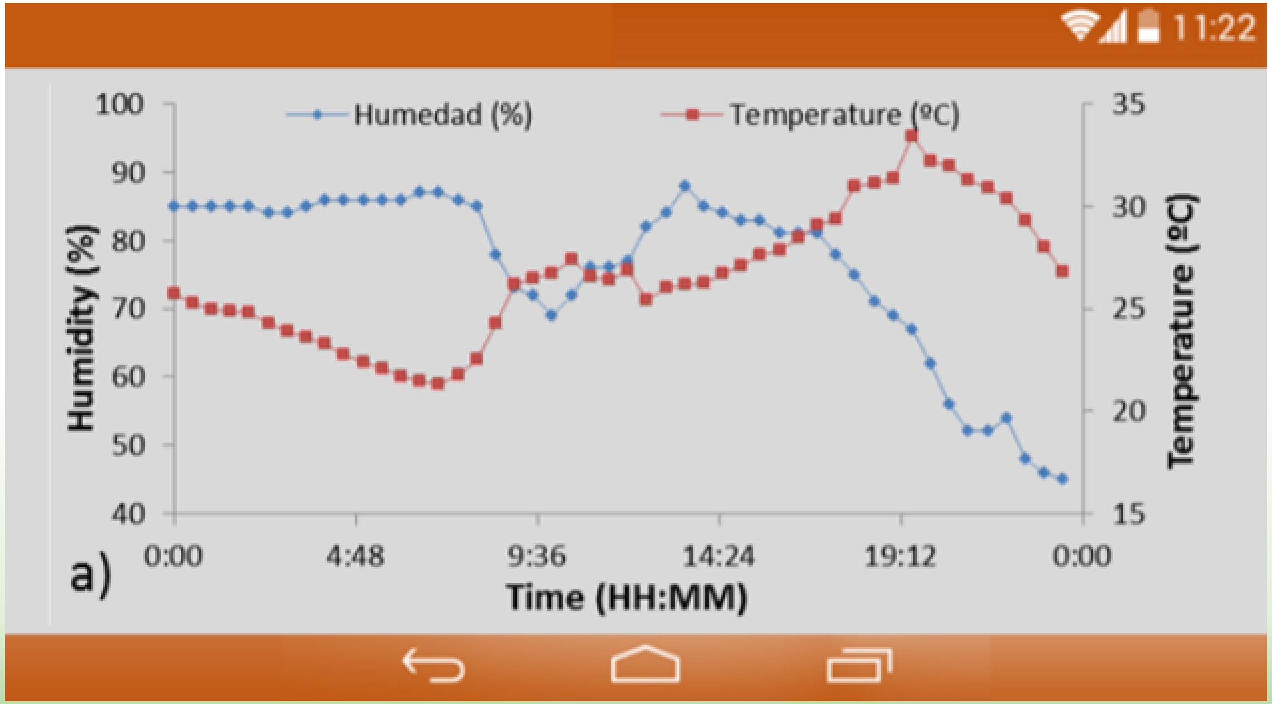
Different variables have different temporal variability

The same variable in different environments has different temporal variability

Introduction: data and its relevance

Data periodicity (1 data every X minutes)

Different variables have different temporal variability



García, L., Parra, L., Romero, O., & Lloret, J. (2019). System for monitoring the wellness state of people in domestic environments employing emoticon-based HCI. The Journal of Supercomputing, 75(4), 1869-1893.

Introduction: data and its relevance

Data periodicity (1 data every X minutes)

The same variable in different environments has different temporal variability

García, L., Parra, L., Romero, O., & Lloret, J. (2019). System for monitoring the wellness state of people in domestic environments employing emoticon-based HCI. The Journal of Supercomputing, 75(4), 1869-1893.

Introduction: data and its relevance

Data resolution (X bits per data)

Although some sensor offer very high resolutions, in most of the cases for the future use of the data this resolution is not needed.

Nonetheless, if we send and store the data with the maximum allowed resolution, we will consume energy, bandwidth, and virtual space.

Again, the required resolution will depend on the variable, on the environment and on the future use of the data.

Let us see some examples…

Introduction: data and its relevance

Data resolution (X bits per data)

Let us see some examples…

Temperature

¿Which resolution do we need for…

...human monitoring (illnesses)?

…water monitoring in oceanic observation?

…heating monitoring in heat dissipater?

Introduction: data and its relevance

Data resolution (X bits per data)

Let us see some examples…

Moisture

¿Which resolution do we need for…

…food in controlling desiccation/mixing processes?

…soil in controlling irrigation?

…monitoring chemical products stored?

Introduction: data and its relevance

Data resolution (X bits per data)

Let us see some a practical case:

When monitoring temperature of a heat dissipater in a data centre we use a:

16 bit sensor (resolution ± 0.25 ºC)

8-bit sensor (resolution ± 1 ºC)

Each sensor monitors the temperature every 5 min:

= 2240 or 4480 bits each day…

= 16,128 or 32,256 bits each week…

= 67,200 or 134,400 bits each month…

Introduction: data and its relevance

Data resolution (X bits per data)

Let us see some a practical case:

When monito… …centre we use a:

16 bit senso…

8-bit sensor …

¿Do we really need ± 0.25 ºC resolution?
We need to evaluate the cost of our elections in defining the resolution of our data

Each sensor monitors the temperature every 5 min:

= 2240 or 4480 bits each day…
= 16,128 or 32,256 bits each week…
= 67,200 or 134,400 bits each month…

Introduction: data and its relevance

Number of measurement devices (X nodes per point)

Although the cost of sensor have decreased, we have to find trade-off between be conservative and select the number of nodes that ensures the collection of all the data and rely on data extrapolation minimizing the number of monitored points.

We need to take into account the following items to define the density of the network:

Expected sensor failures

Expected heterogeneity of environment and the variable

Let us see some examples…

Introduction: data and its relevance

Number of measurement devices (X nodes per point)

Let us see some examples…

When measuring soil moisture in rainfed crops we need less density of sensors than in drip irrigation crops. For example 1 sensor per Ha in rainfed vs 1 sensor per drip line.

When measuring sound level in indoor locations for occupational risks assessment we need higher sensor density than in monitoring highroad acoustic impact. For example 1 sensor per person vs 1 sensor every 2 km.

Introduction: data and its relevance

Data sending (X message every X data gathered)

This is mostly related to the expected use of data and have a huge effect on the packet size and on the efficiency of the transmission.

If a packet is generated and sent each time that the node gathers data (considering that data can be just 16 bits) the packet will have a minimum worthy payload.

Although it does not impact on the stored data it has an enormous impact on consumed bandwidth and on the energy balance of the node.

Introduction: data and its relevance

Data sending (X message every X data gathered)

Nonetheless, in many cases, we can avoid sending the data every time that the sensor gathers it. It is much more efficient to store it in the node and sent it once the data is necessary (is going to be used).

But, how the node can knows when it is necessary?

We will see it in this keynote

Index

- Introduction: data and its relevance
- Data variability in real conditions
- Relevance of data in case of images
- How we can take an advantage of managing data variability properly?
- Real cases of application of data management algorithms

Data variability in real conditions

Before analysing how the node can decide if sensing or not the data, we need to be aware of data variability in real conditions.

As set before, each variable and each environment will have their own variability. This variability can be modified by external factors. Thus, we need to know the expected variability.

Data variability in real conditions

We need to differentiate between data and information. Not every data contains useful information.

When there is low variability in the monitored variable, a reduction on the periodicity suppose a reduction in the data without reducing the information.

We are going to see a practical case.

Data variability in real conditions

Practical case: Monitoring green areas

4 Studied variables in 12 different areas at 5 different times:

- Soil Moisture (SM)
- Soil Temperature (ST)
- Canopy Temperature (CT)
- Normalized Difference Vegetation Index (NDVI)

(a) (b)

Figure 1. Pictures of (**a**) used devices and (**b**) data gathering process.

Mauri, P. V., Parra, L., Yousfi, S., Lloret, J., & Marin, J. F. (2021). Evaluating the Effects of Environmental Conditions on Sensed Parameters for Green Areas Monitoring and Smart Irrigation Systems. Sensors, 21(6), 2255.

Data variability in real conditions

Practical case: Monitoring green areas

No effect of time on SM

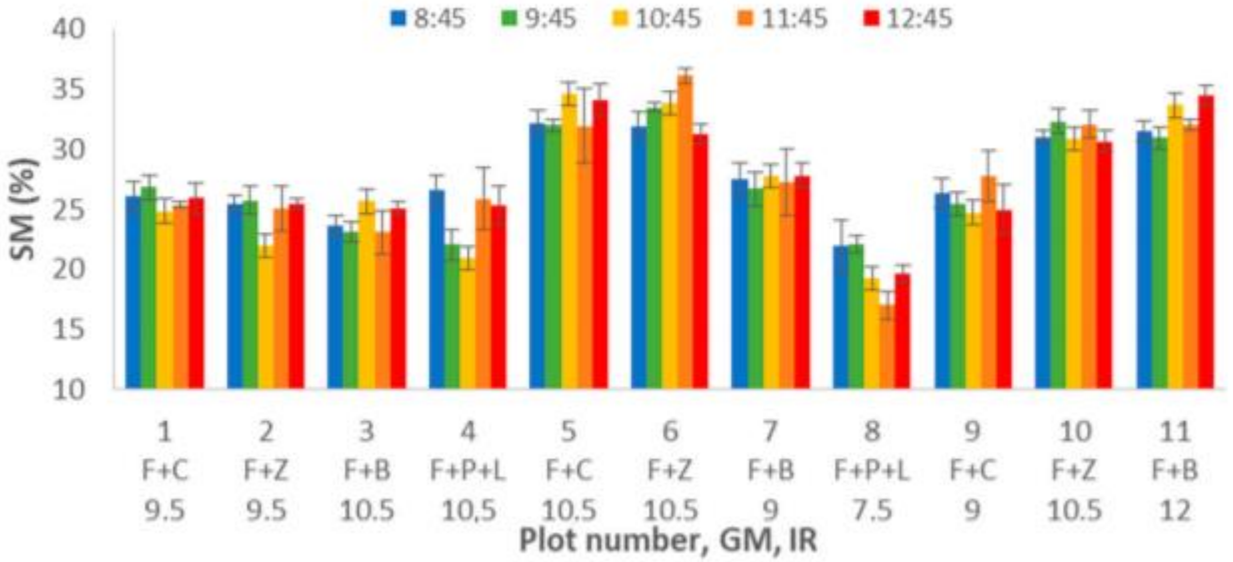


Figure 2. Mean values of SM for different plot numbers and DATs.

4 Studied variables in 12 different areas at 5 different times:

- Soil Moisture (SM)
- Soil Temperature (ST)
- Canopy Temperature (CT)
- Normalized Difference Vegetation Index (NDVI)

Mauri, P. V., Parra, L., Yousfi, S., Lloret, J., & Marin, J. F. (2021). Evaluating the Effects of Environmental Conditions on Sensed Parameters for Green Areas Monitoring and Smart Irrigation Systems. Sensors, 21(6), 2255.

Data variability in real conditions

Practical case: Monitoring green areas

4 Studied variables in 12 different areas at 5 different times:

- Soil Moisture (SM)
- Soil Temperature (ST)
- Canopy Temperature (CT)
- Normalized Difference Vegetation Index (NDVI)

Effect of time on ST

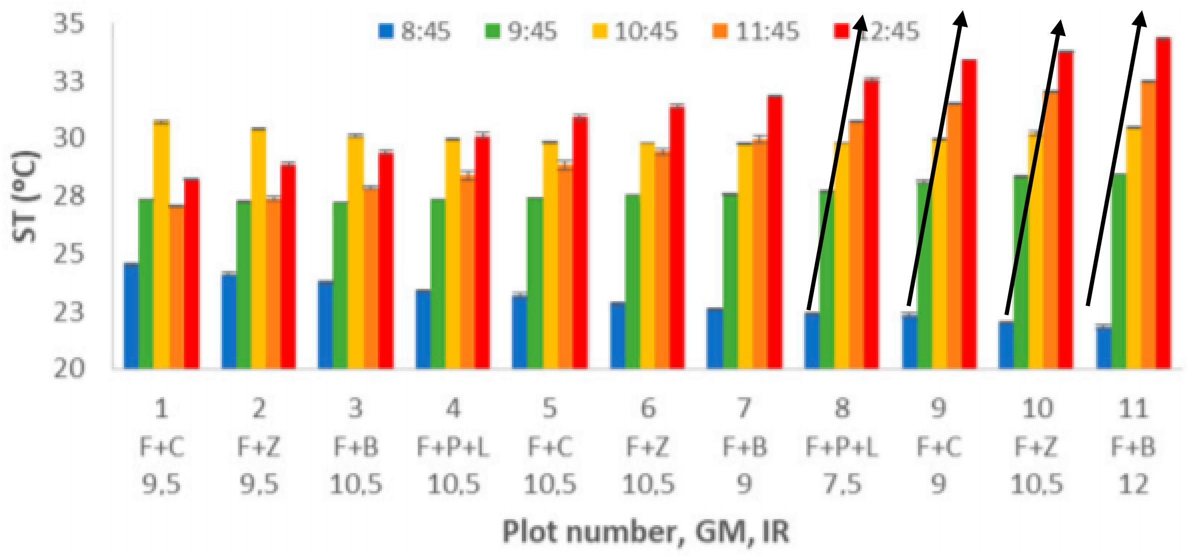


Figure 5. Mean values of ST for different plot numbers and DATs.

Mauri, P. V., Parra, L., Yousfi, S., Lloret, J., & Marin, J. F. (2021). Evaluating the Effects of Environmental Conditions on Sensed Parameters for Green Areas Monitoring and Smart Irrigation Systems. Sensors, 21(6), 2255.

Data variability in real conditions

Practical case: Monitoring green areas

4 Studied variables in 12 different areas at 5 different times:

- Soil Moisture (SM)
- Soil Temperature (ST)
- Canopy Temperature (CT)
- Normalized Difference Vegetation Index (NDVI)

Effect of time on CT

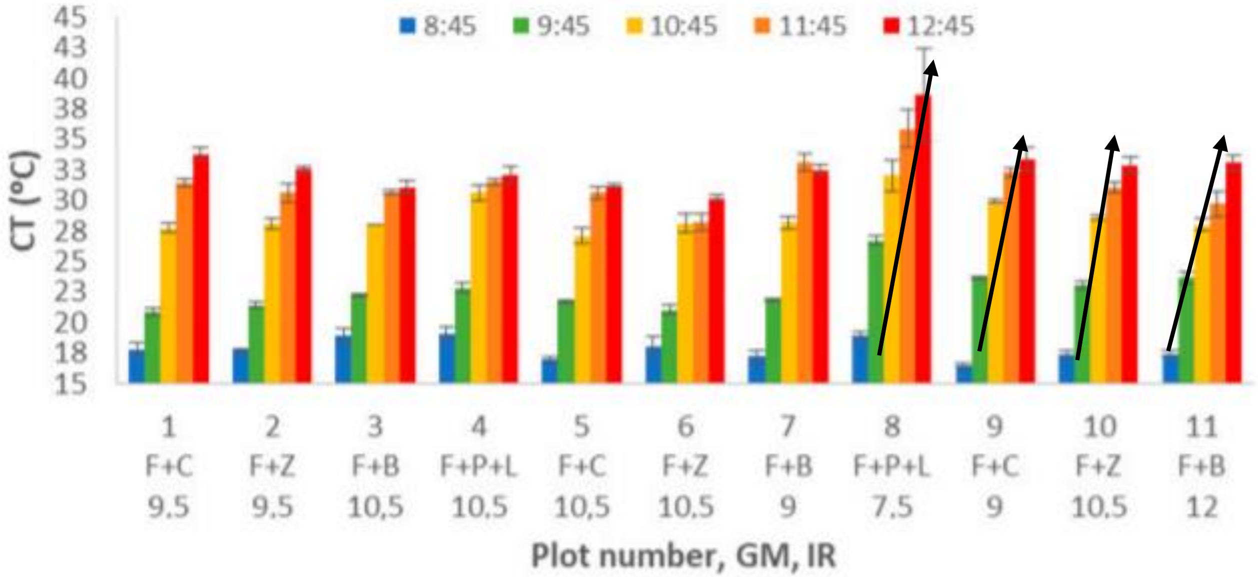


Figure 8. Mean values of CT for different plot numbers and DATs.

Mauri, P. V., Parra, L., Yousfi, S., Lloret, J., & Marin, J. F. (2021). Evaluating the Effects of Environmental Conditions on Sensed Parameters for Green Areas Monitoring and Smart Irrigation Systems. Sensors, 21(6), 2255.

Data variability in real conditions

Practical case: Monitoring green areas

4 Studied variables in 12 different areas at 5 different times:

- Soil Moisture (SM)
- Soil Temperature (ST)
- Canopy Temperature (CT)
- Normalized Difference Vegetation Index (NDVI)

No effect of time on NDVI

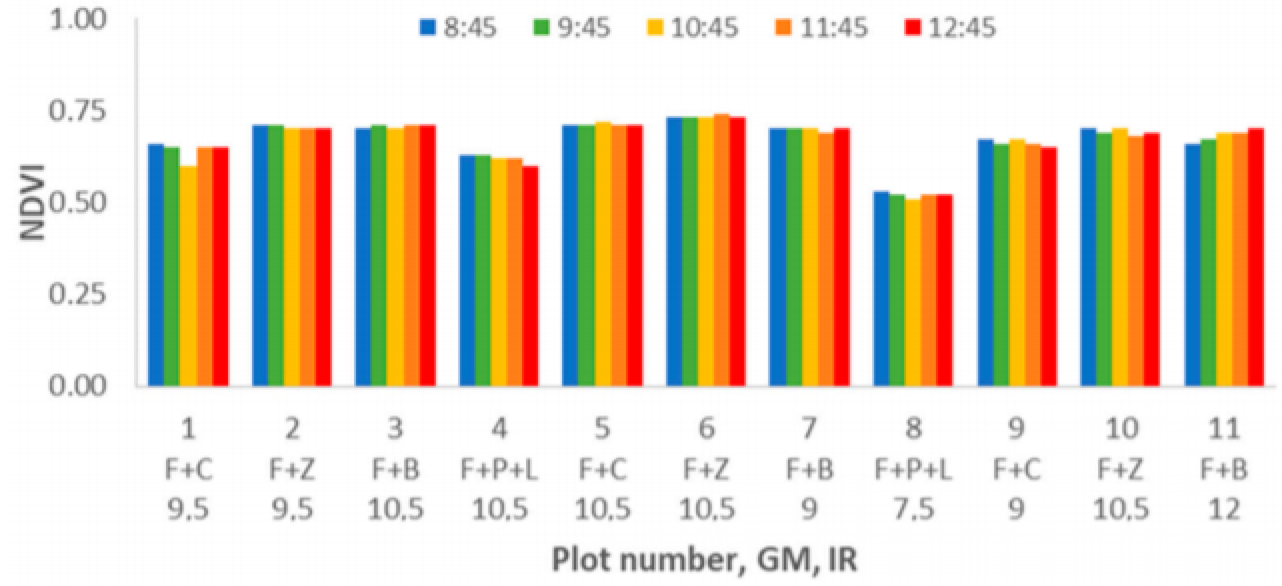


Figure 12. Mean values of NDVI for different plot numbers and DATs.

Mauri, P. V., Parra, L., Yousfi, S., Lloret, J., & Marin, J. F. (2021). Evaluating the Effects of Environmental Conditions on Sensed Parameters for Green Areas Monitoring and Smart Irrigation Systems. Sensors, 21(6), 2255.

Data variability in real conditions

Practical case: Monitoring green areas

4 Studied variables in 12 different areas at 5 different times:

We have seen how in the same environment some parameter are modified along the time and other remain constant (new data is the same than previous data, no information is added). This must be taken into account to define the periodicity. Periodicity of SM and NDVI will be higher than for ST and CT. For example:

ST and CT 1 measured every 30 min while SM 1 measures every 4h and NDVI 1 measure every 1 day.

Mauri, P. V., Parra, L., Yousfi, S., Lloret, J., & Marin, J. F. (2021). Evaluating the Effects of Environmental Conditions on Sensed Parameters for Green Areas Monitoring and Smart Irrigation Systems. Sensors, 21(6), 2255.

Index

- Introduction: data and its relevance
- Data variability in real conditions
- Relevance of data in case of images
- How we can take an advantage of managing data variability properly?
- Real cases of application of data management algorithms

Relevance of data in case of images

Images are a particular case. Opposed to the sensor data (temperature, moisture, etc…), which data consists on 8 or 16 bits, images are composed of thousands of bits. Thus, we need to be more thorough when our networks include cameras.

The drones have become an essential element in monitoring systems due to the reduction on their cost and their multiple applications. Therefore, the generation of images gathered by drones has increased.

Hopefully, the operational capacity of nodes to process these images has also grown, possibility the local processing of images.

Relevance of data in case of images

Practical case: Monitoring green areas

1º RGB pictures gathered to analyse the grass coverage.

↓

2º Pictures are decomposed into R, G, and B band.

↓

3º Histogram of green band is obtained and certain region is conjoined.

With the sum of this region we are capable of estimating the grass coverage in the picture.

Marín, J., Parra, L., Rocher, J., Sendra, S., Lloret, J., Mauri, P. V., & Masaguer, A. (2018). Urban lawn monitoring in smart city environments. Journal of Sensors, 2018.

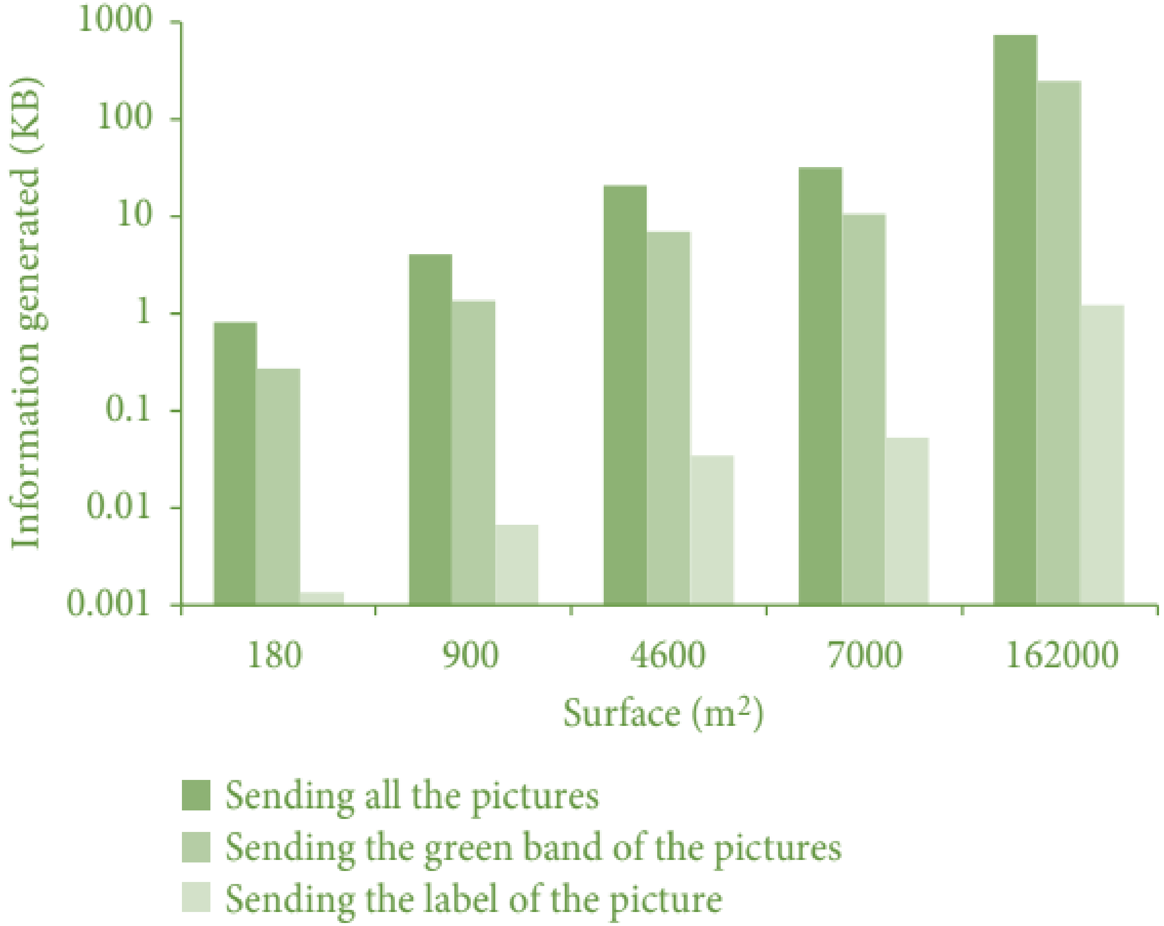
Relevance of data in case of images

Practical case: Monitoring green areas

We have the possibility of sending the whole picture (3 bands), sending just the useful band (Green band) or sending only the sum of the part of its histogram or the classification according to the histogram.

Example→

By locally processing the information and sending the label supposes a reduction of 99.8% of the data volume compared with sending the picture



Marín, J., Parra, L., Rocher, J., Sendra, S., Lloret, J., Mauri, P. V., & Masaguer, A. (2018). Urban lawn monitoring in smart city environments. Journal of Sensors, 2018.

Relevance of data in case of images

Practical case: Monitoring green areas

Nonetheless, processing images locally in the node supposes certain restrictions:

1-We need to select simple processing methods, AI is not recommended.
2-We need to use powerful nodes (which implies higher energy use).
3-If possible, we should avoid images with very high resolution.

One of the simplest methods for image processing which is useful in many cases, more useful than histograms is the band combination. This is widely used in remote sensing.

Relevance of data in case of images – Band combination

Practical case: Identification of weed species

1º RGB pictures gathered to analyse the weed species in golf course.

2º Pictures are decomposed into R, G, and B band (or B1, B2, and B3).

3º R, G, and B band are combined to generate a new band.

4º According to a threshold the pixels of the new band a classified as grass or weed.

With the classification we can identify the weed in the picture with a single band.

J. F. Marin et al., Comparison of performance in weed detection with aerial RGB and thermal images gathered at different height. In Proceedings of the Sensing Systems for Agricultural Management (SeSAM 2021), May 30 - June 03, 2021, Valencia, Spain.

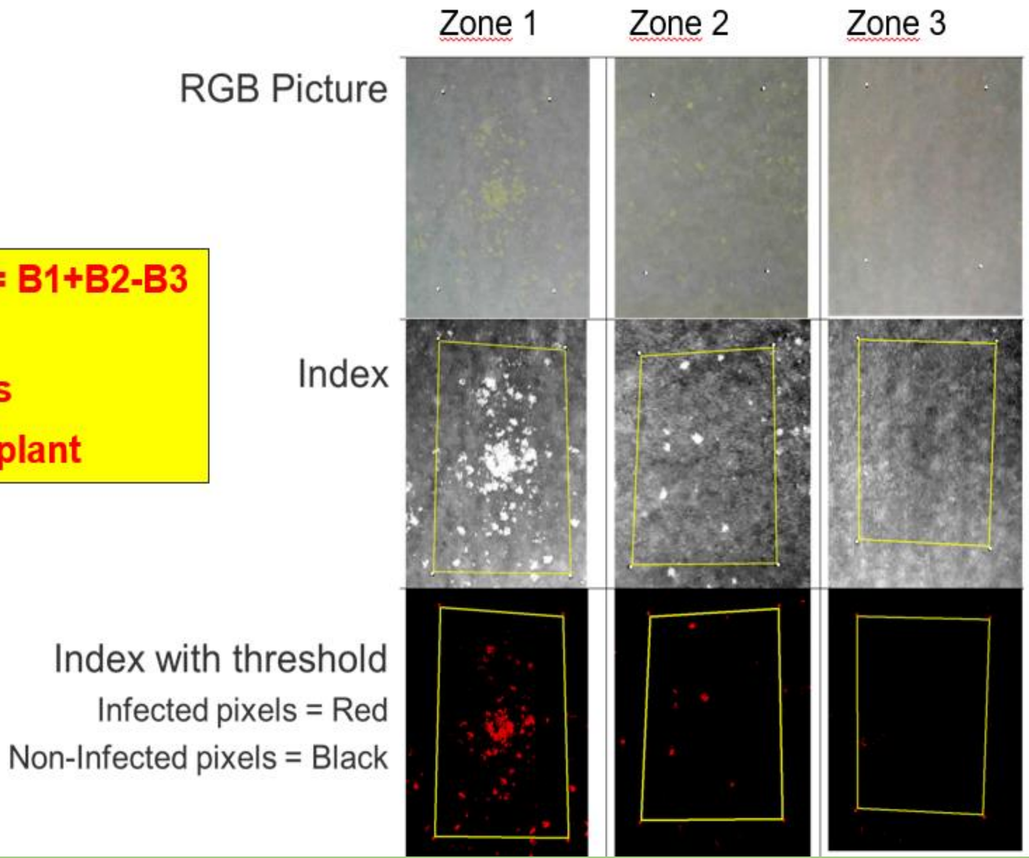
Relevance of data in case of images – Band combination

Practical case: Identification of weed species

We have the possibility of sensing the whole picture (3 bands or sending the new generated band. Or even, as in the previous case the tag of the picture (weed presence or absence)

We simplify the process of weed identification for an specific case. Standardization of indexes for other cases is still in progress.

Vegetation Index= B1+B2-B3
Theshold:
-29 to 189 = grass
190 to 260 = wild plant



J. F. Marin et al., Comparison of performance in weed detection with aerial RGB and thermal images gathered at different height. In Proceedings of the Sensing Systems for Agricultural Management (SeSAM 2021), May 30 - June 03, 2021, Valencia, Spain.

Relevance of data in case of images – Band combination

Practical case: Estimation of yield

1º RGB pictures gathered to analyse the quantity of fruit.

2º Pictures are decomposed into R, G, and B band.

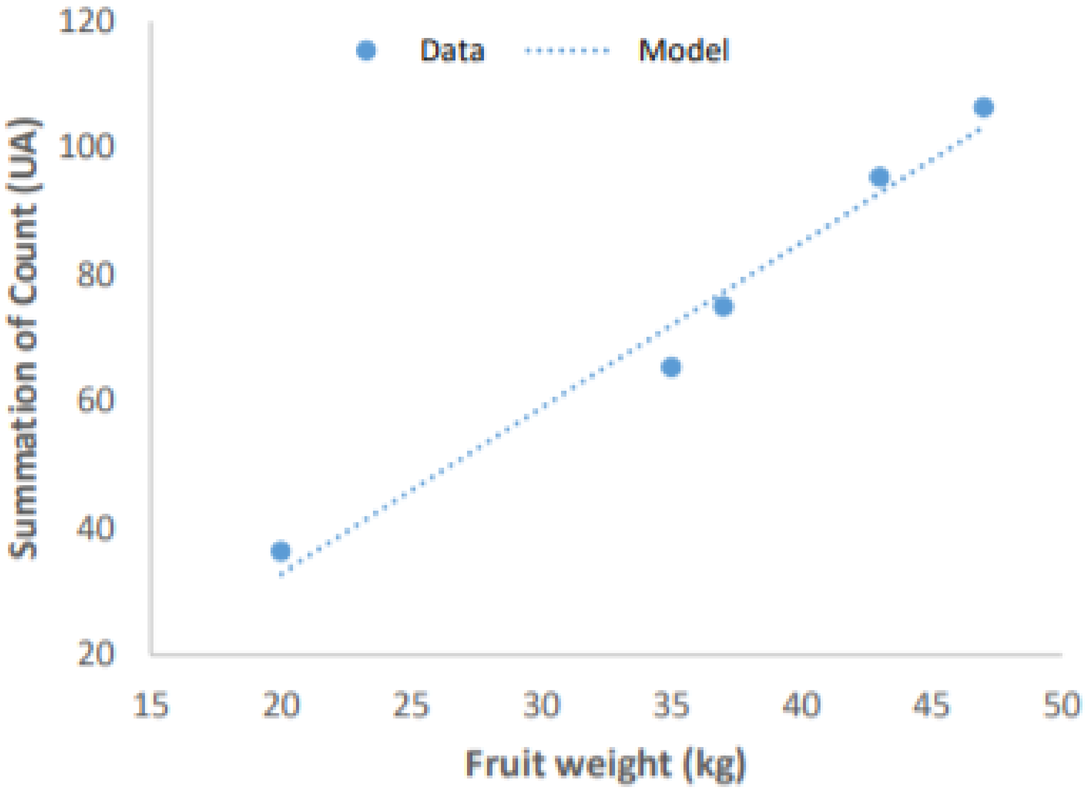
3º Histogram of the new band is obtained and certain region is conjoined.

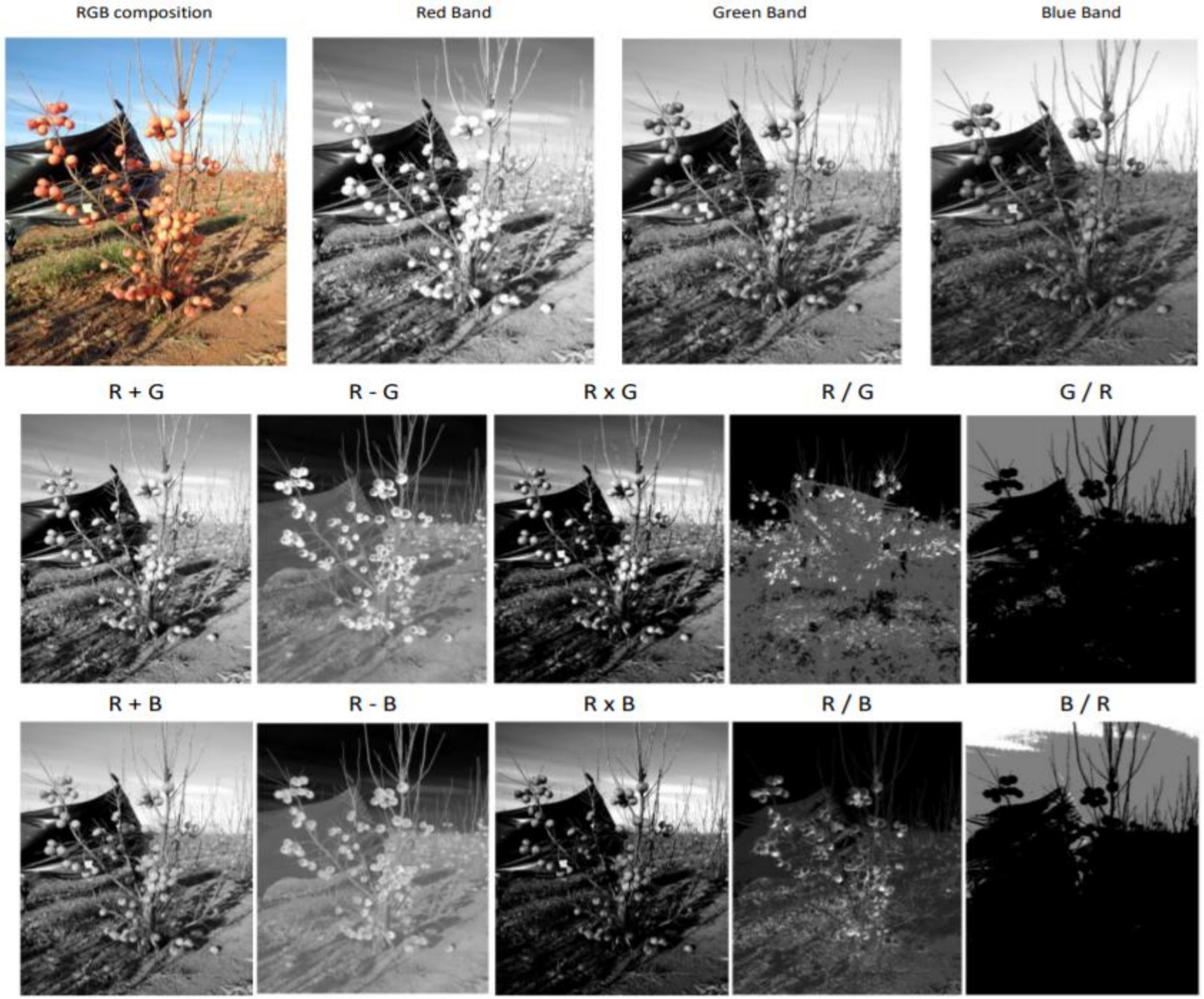
With the sum of this region we are capable of estimating the yield of this three.

Garcia, L., Parra, L., Basterrechea, D. A., Jimenez, J. M., Rocher, J., Parra, M., ... & Lorenz, P. (2019). Quantifying the production of fruit-bearing trees using image processing techniques. Proceedings of the INNOV.

Relevance of data in case of images – Band combination

Practical case: Estimation of yield





Garcia, L., Parra, L., Basterrechea, D. A., Jimenez, J. M., Rocher, J., Parra, M., ... & Lorenz, P. (2019). Quantifying the production of fruit-bearing trees using image processing techniques. Proceedings of the INNOV.

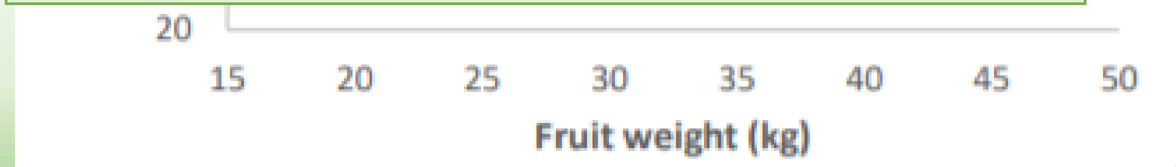
Relevance of data in case of images – Band combination

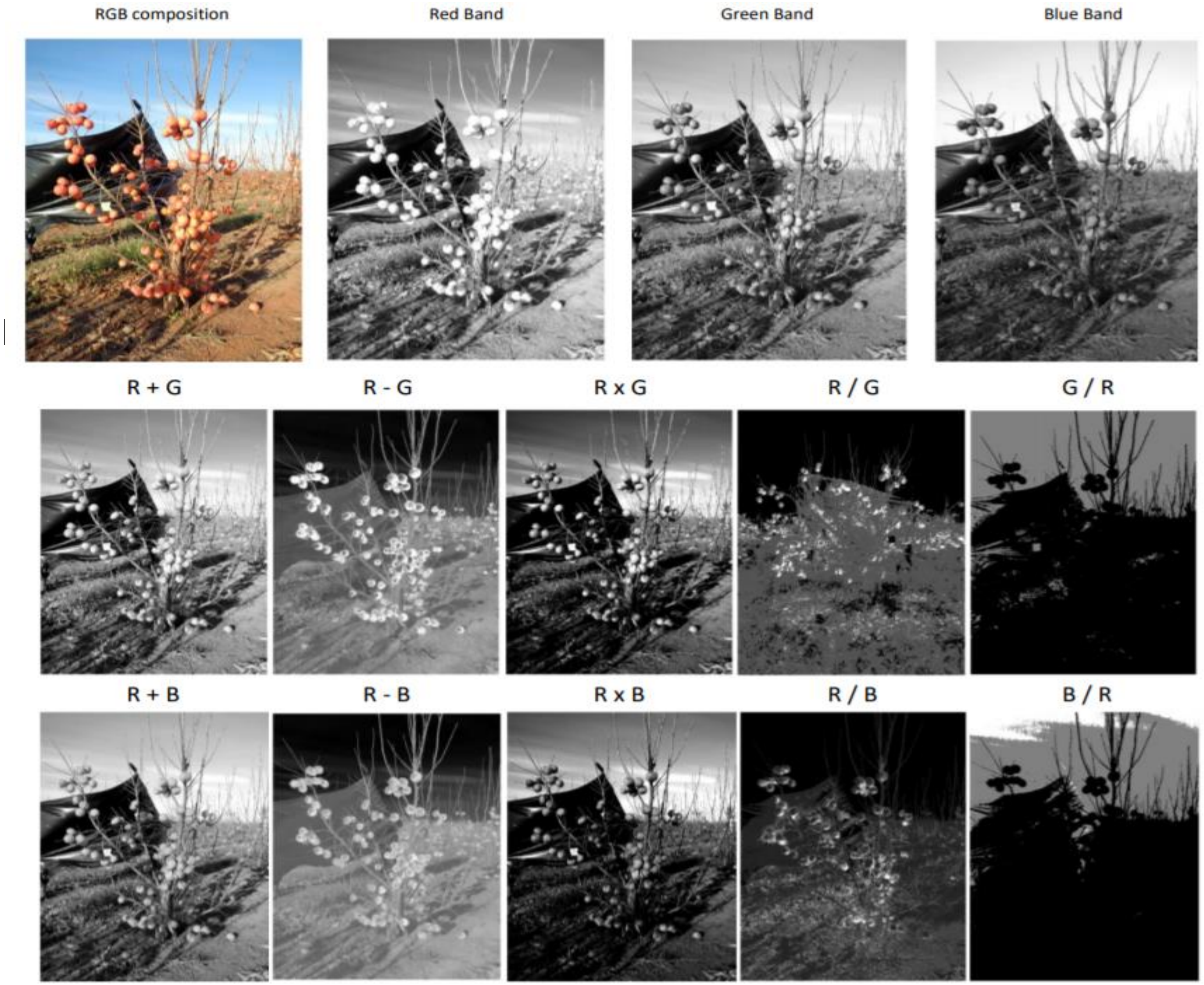
Practical case: Estimation of yield

The combination of bands allows the estimation of yield. Again we can send the whole picture, the new band or

We simplify the process of weed identification for an specific case.

Standardization of indexes for other cases is still in progress.





Garcia, L., Parra, L., Basterrechea, D. A., Jimenez, J. M., Rocher, J., Parra, M., ... & Lorenz, P. (2019). Quantifying the production of fruit-bearing trees using image processing techniques. Proceedings of the INNOV.

Index

- Introduction: data and its relevance
- Data variability in real conditions
- Relevance of data in case of images
- How we can take an advantage of managing data variability properly?
- Real cases of application of data management algorithms

How we can take an advantage of managing data variability properly?

We have seen several examples in which knowing and managing data variability is useful for reducing the generation of data without loosing information:

-Reducing and adapting the periodicity of data for each variable

-Reducing the resolution for certain variables

-Adapting the number of nodes to the variability of the environment

Nonetheless, in some cases, we are seeking to monitor the apparition of abnormal data, which cannot be related to expected events. In other cases, we cannot loss because we do not know if it can be useful in the future or not.

How we can take an advantage of managing data variability properly?

We can develop certain algorithms based on data analysis to decide if the data gathered can be considered as useful information or not.

Let us see an example:

How we can take an advantage of managing data variability properly?

Let us see an example:

21/05/2020	0:15	16,76
21/05/2020	0:35	16,76
21/05/2020	0:55	16,69
21/05/2020	1:15	16,69
21/05/2020	1:35	16,61
21/05/2020	1:55	16,61
21/05/2020	2:15	16,53
21/05/2020	2:35	16,53
21/05/2020	2:55	16,46
21/05/2020	3:15	16,46
21/05/2020	3:35	16,38
21/05/2020	3:55	16,30
21/05/2020	4:15	16,30
21/05/2020	4:35	16,30
21/05/2020	4:55	16,23
21/05/2020	5:15	16,23
21/05/2020	5:35	16,23
21/05/2020	5:55	16,07

21/05/2020	6:15	16,15
21/05/2020	6:35	16,00
21/05/2020	6:55	16,07
21/05/2020	7:15	16,23
21/05/2020	7:35	16,15
21/05/2020	7:55	16,23
21/05/2020	8:15	16,30
21/05/2020	8:35	16,23
21/05/2020	8:55	16,30
21/05/2020	9:15	16,46
21/05/2020	9:35	16,53
21/05/2020	9:55	16,53
21/05/2020	10:15	16,69
21/05/2020	10:35	16,76
21/05/2020	10:55	16,84
21/05/2020	11:15	16,99
21/05/2020	11:35	16,99
21/05/2020	11:55	17,07

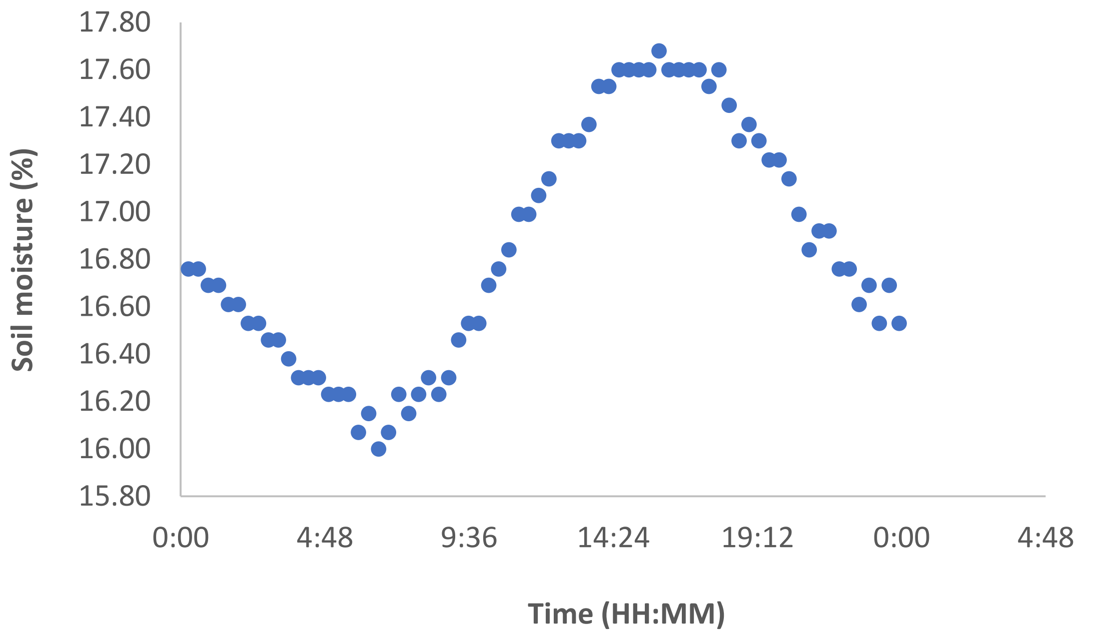
21/05/2020	12:15	17,14
21/05/2020	12:35	17,30
21/05/2020	12:55	17,30
21/05/2020	13:15	17,30
21/05/2020	13:35	17,37
21/05/2020	13:55	17,53
21/05/2020	14:15	17,53
21/05/2020	14:35	17,60
21/05/2020	14:55	17,60
21/05/2020	15:15	17,60
21/05/2020	15:35	17,60
21/05/2020	15:55	17,68
21/05/2020	16:15	17,60
21/05/2020	16:35	17,60
21/05/2020	16:55	17,60
21/05/2020	17:15	17,60
21/05/2020	17:35	17,53
21/05/2020	17:55	17,60

21/05/2020	18:15	17,45
21/05/2020	18:35	17,30
21/05/2020	18:55	17,37
21/05/2020	19:15	17,30
21/05/2020	19:35	17,22
21/05/2020	19:55	17,22
21/05/2020	20:15	17,14
21/05/2020	20:35	16,99
21/05/2020	20:55	16,84
21/05/2020	21:15	16,92
21/05/2020	21:35	16,92
21/05/2020	21:55	16,76
21/05/2020	22:15	16,76
21/05/2020	22:35	16,61
21/05/2020	22:55	16,69
21/05/2020	23:15	16,53
21/05/2020	23:35	16,69
21/05/2020	23:35	16,53

How we can take an advantage of managing data variability properly?

Let us see an example:

Soil moisture data in rainfed crop, data is gathered every 20min.

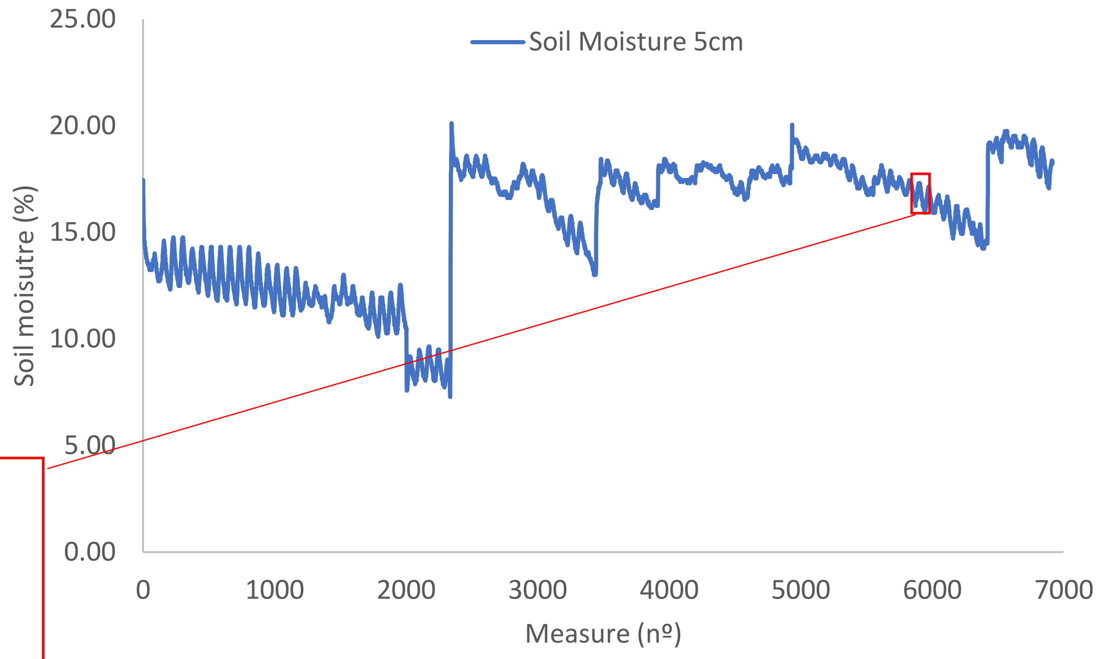


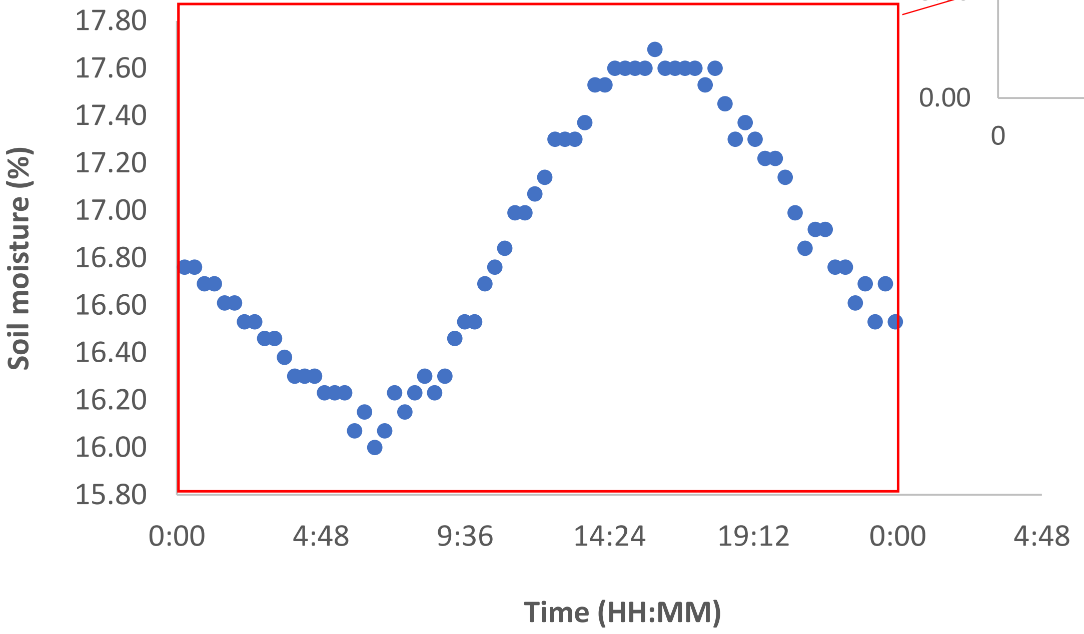
It seems that there is variability in data. But if everyday is the same pattern, we can gather data just four to five times per day.

How we can take an advantage of managing data variability properly?

Let us see an example:

Soil moisture data in rainfed crop, data is gathered every 20min.



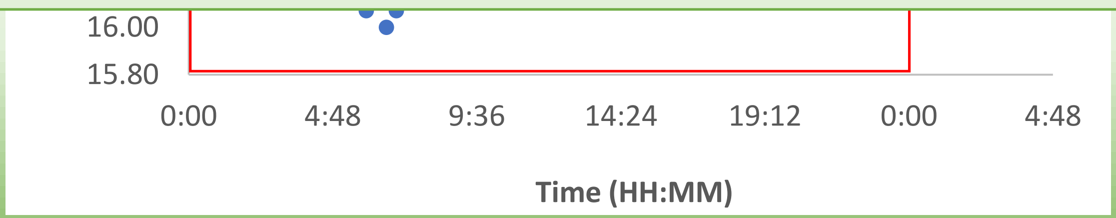


How we can take an advantage of managing data variability properly?

After checking the data with the whole data set (several month) the variability is minimum and having two measured per day is sufficient to represent the tendency.

But, what we can do if we do not have this information when designing the network?

We can compare the data gathered with the previous value and having an approximation of expected value decide if this data is relevant or not.



How we can take an advantage of managing data variability properly?

Let us see an example:

Soil moisture data in rainfed crop, data is gathered every 20 min.

We know that soil moisture goes from 0 to 100%, according to the type of soil (if we know the type of soil) we can reduce this range. It rarely overcomes the 60%. Thus, our variable can have values from 0 to 60%. We can define the minimum variation in the moisture that suppose that data is useful information. We can assume that a variation higher than 0.5% of the soil moisture is considered as an important change in the variable.

How we can take an advantage of managing data variability properly?

Let us see an example:

Comparing data to define if new data is useful information.

Note: We cannot compare the new data with the previous data, otherwise…

How we can take an advantage of managing data variability properly?

Let us see an example:

Information if

$Data_t - Data_{t-1} > 0.5$

1º data is information

No one data has a variation higher than 0.5 with the previous data

21/05/2020	0:15	**16,76**
21/05/2020	0:35	16,76
21/05/2020	0:55	16,69
21/05/2020	1:15	16,69
21/05/2020	1:35	16,61
21/05/2020	1:55	16,61
21/05/2020	2:15	16,53
21/05/2020	2:35	16,53
21/05/2020	2:55	16,46
21/05/2020	3:15	16,46
21/05/2020	3:35	16,38
21/05/2020	3:55	16,30
21/05/2020	4:15	16,30
21/05/2020	4:35	16,30
21/05/2020	4:55	16,23
21/05/2020	5:15	16,23
21/05/2020	5:35	16,23
21/05/2020	5:55	16,07

21/05/2020	6:15	16,15
21/05/2020	6:35	16,00
21/05/2020	6:55	16,07
21/05/2020	7:15	16,23
21/05/2020	7:35	16,15
21/05/2020	7:55	16,23
21/05/2020	8:15	16,30
21/05/2020	8:35	16,23
21/05/2020	8:55	16,30
21/05/2020	9:15	16,46
21/05/2020	9:35	16,53
21/05/2020	9:55	16,53
21/05/2020	10:15	16,69
21/05/2020	10:35	16,76
21/05/2020	10:55	16,84
21/05/2020	11:15	16,99
21/05/2020	11:35	16,99
21/05/2020	11:55	17,07

21/05/2020	12:15	17,14
21/05/2020	12:35	17,30
21/05/2020	12:55	17,30
21/05/2020	13:15	17,30
21/05/2020	13:35	17,37
21/05/2020	13:55	17,53
21/05/2020	14:15	17,53
21/05/2020	14:35	17,60
21/05/2020	14:55	17,60
21/05/2020	15:15	17,60
21/05/2020	15:35	17,60
21/05/2020	15:55	17,68
21/05/2020	16:15	17,60
21/05/2020	16:35	17,60
21/05/2020	16:55	17,60
21/05/2020	17:15	17,60
21/05/2020	17:35	17,53
21/05/2020	17:55	17,60

21/05/2020	18:15	17,45
21/05/2020	18:35	17,30
21/05/2020	18:55	17,37
21/05/2020	19:15	17,30
21/05/2020	19:35	17,22
21/05/2020	19:55	17,22
21/05/2020	20:15	17,14
21/05/2020	20:35	16,99
21/05/2020	20:55	16,84
21/05/2020	21:15	16,92
21/05/2020	21:35	16,92
21/05/2020	21:55	16,76
21/05/2020	22:15	16,76
21/05/2020	22:35	16,61
21/05/2020	22:55	16,69
21/05/2020	23:15	16,53
21/05/2020	23:35	16,69
21/05/2020	23:35	16,53

How we can take an advantage of managing data variability properly?

Let us see an example:

Information if

$Data_t - Data_{stored} > 0.5$

1º data is information

5 values are considered as information

21/05/2020	0:15	**16,76**
21/05/2020	0:35	16,76
21/05/2020	0:55	16,69
21/05/2020	1:15	16,69
21/05/2020	1:35	16,61
21/05/2020	1:55	16,61
21/05/2020	2:15	16,53
21/05/2020	2:35	16,53
21/05/2020	2:55	16,46
21/05/2020	3:15	16,46
21/05/2020	3:35	16,38
21/05/2020	3:55	16,30
21/05/2020	4:15	16,30
21/05/2020	4:35	16,30
21/05/2020	4:55	**16,23**
21/05/2020	5:15	16,23
21/05/2020	5:35	16,23
21/05/2020	5:55	16,07

21/05/2020	6:15	16,15
21/05/2020	6:35	16,00
21/05/2020	6:55	16,07
21/05/2020	7:15	16,23
21/05/2020	7:35	16,15
21/05/2020	7:55	16,23
21/05/2020	8:15	16,30
21/05/2020	8:35	16,23
21/05/2020	8:55	16,30
21/05/2020	9:15	16,46
21/05/2020	9:35	16,53
21/05/2020	9:55	16,53
21/05/2020	10:15	16,69
21/05/2020	10:35	**16,76**
21/05/2020	10:55	16,84
21/05/2020	11:15	16,99
21/05/2020	11:35	16,99
21/05/2020	11:55	17,07

21/05/2020	12:15	17,14
21/05/2020	12:35	**17,30**
21/05/2020	12:55	17,30
21/05/2020	13:15	17,30
21/05/2020	13:35	17,37
21/05/2020	13:55	17,53
21/05/2020	14:15	17,53
21/05/2020	14:35	17,60
21/05/2020	14:55	17,60
21/05/2020	15:15	17,60
21/05/2020	15:35	17,60
21/05/2020	15:55	17,68
21/05/2020	16:15	17,60
21/05/2020	16:35	17,60
21/05/2020	16:55	17,60
21/05/2020	17:15	17,60
21/05/2020	17:35	17,53
21/05/2020	17:55	17,60

21/05/2020	18:15	17,45
21/05/2020	18:35	17,30
21/05/2020	18:55	17,37
21/05/2020	19:15	17,30
21/05/2020	19:35	17,22
21/05/2020	19:55	17,22
21/05/2020	20:15	17,14
21/05/2020	20:35	16,99
21/05/2020	20:55	16,84
21/05/2020	21:15	16,92
21/05/2020	21:35	16,92
21/05/2020	21:55	**16,76**
21/05/2020	22:15	16,76
21/05/2020	22:35	16,61
21/05/2020	22:55	16,69
21/05/2020	23:15	16,53
21/05/2020	23:35	16,69
21/05/2020	23:35	16,53

How we can take an advantage of managing data variability properly?

Last considerations:

In some cases, we cannot not have any preliminary ideas or experiences to set a threshold. Thus, we can decide to store all the gathered data and send data only when the change in detected. We will not save virtual space for data storage, but we can save energy in the nodes and bandwidth.

Index

- Introduction: data and its relevance
- Data variability in real conditions
- Relevance of data in case of images
- How we can take an advantage of managing data variability properly?
- Real cases of application of data management algorithms

Real cases of application of data management algorithms

Finally, we are going to focus on some examples in outdoor and indoor monitoring networks in which we will se the improvement in the network when nodes can decide if sending or note the data.

Real cases of application of data management algorithms

First example: Indoor monitoring

Two variables directly measured (temperature and humidity), while for other two (noise and luminosity) average values are used. → Preliminary information about data variability.

Thresholds set according to preliminary information.

Since we have a lot of knowledge about data variability and data is not critical (pollution, health records, etc…) we only send or delete. There is no data storage.

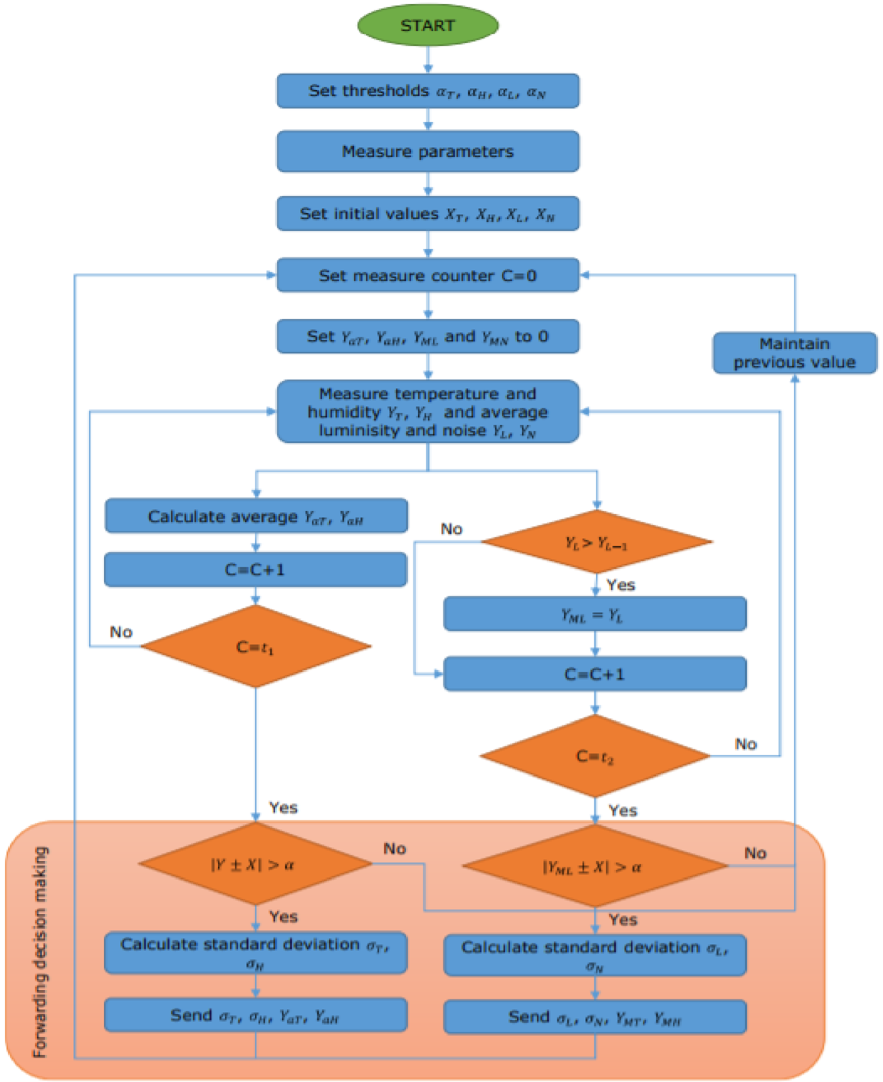


Fig. 3. Algorithm for measuring and forwarding decision making.

García, L., Parra, L., Romero, O., & Lloret, J. (2019). System for monitoring the wellness state of people in domestic environments employing emoticon-based HCI. The Journal of Supercomputing, 75(4), 1869-1893.

Real cases of application of data management algorithms

First example: Indoor monitoring

Time intervals to compare data: N.A. (send all data), 1, 5 or 10s.

Data is gathered during 1 min in 4 different rooms.

Thresholds are 5%, 0.5ºC., 10 lux, and 5dB for humidity, temp., luminosity and sound.

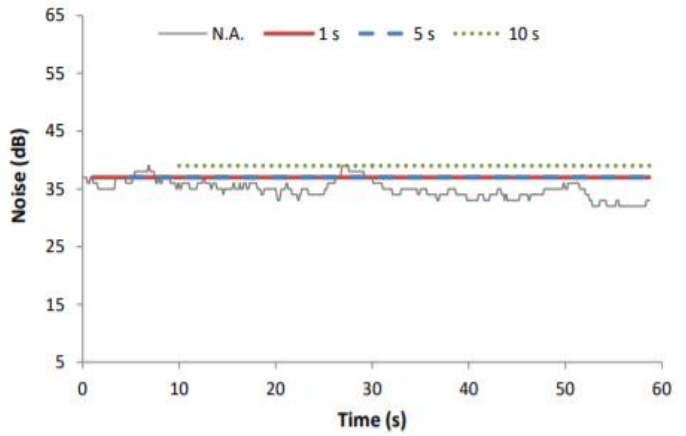


Fig. 21. Results of applying the algorithm with real data from kitchen.

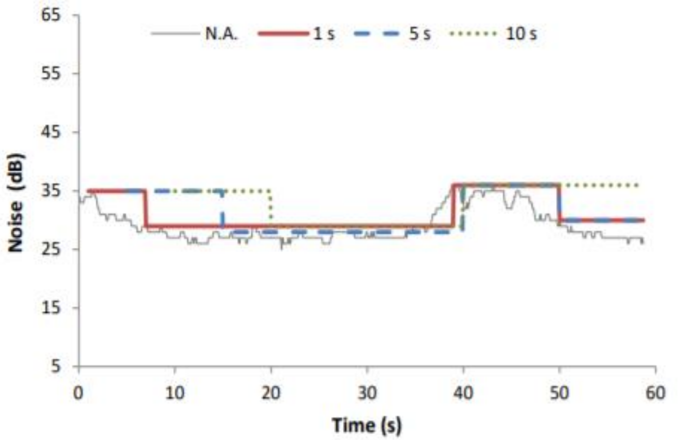


Fig. 22. Results of applying the algorithm with real data from bathroom.

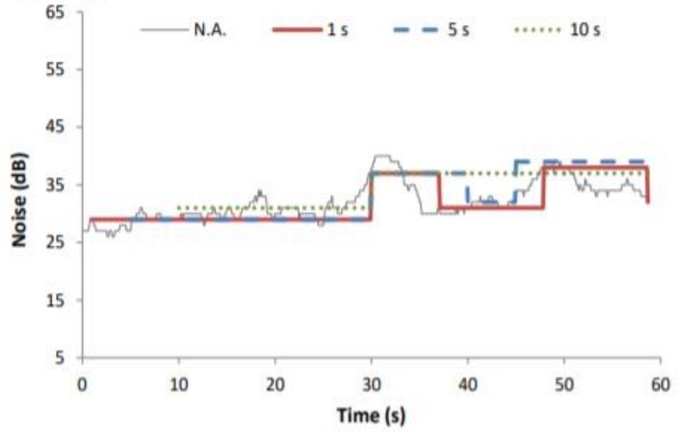


Fig. 23. Results of applying the algorithm with real data from bedroom.

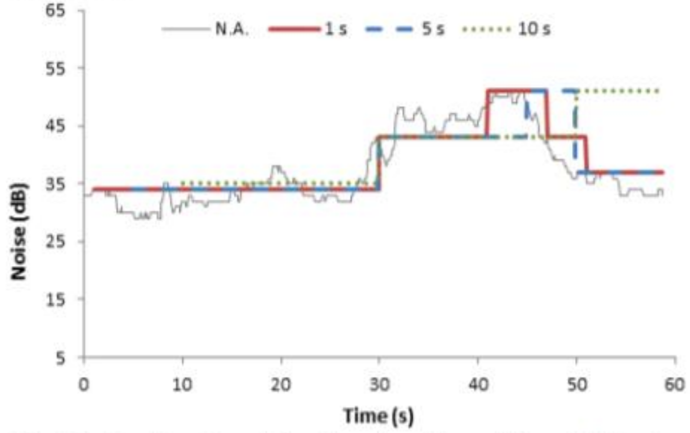


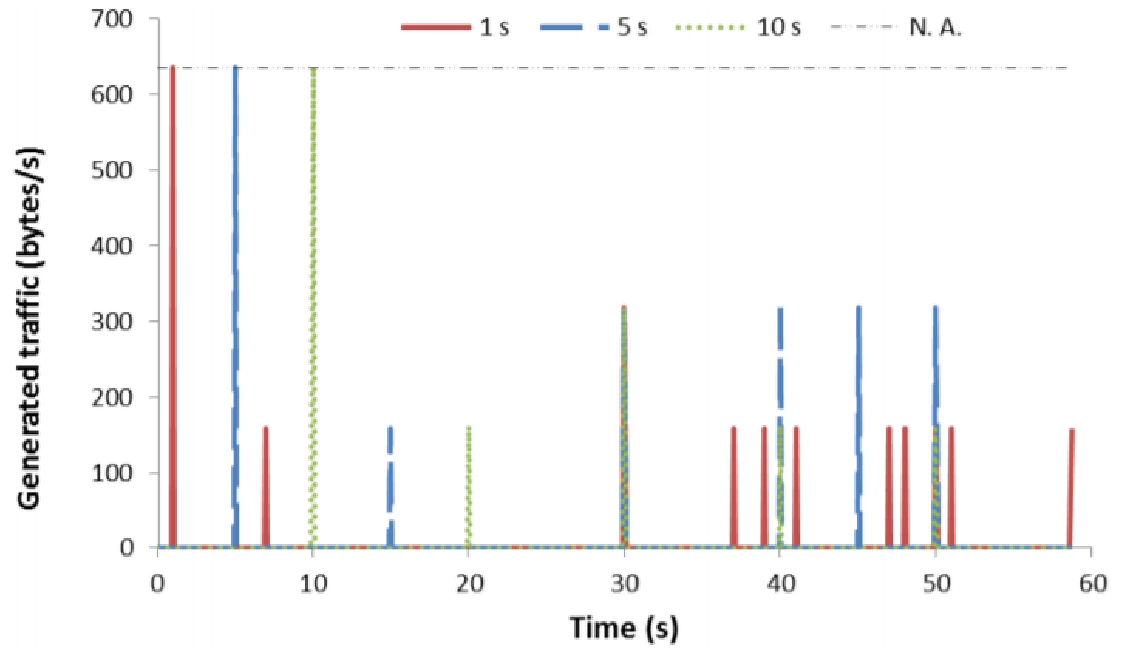
Fig. 24. Results of applying the algorithm with real data from living room.

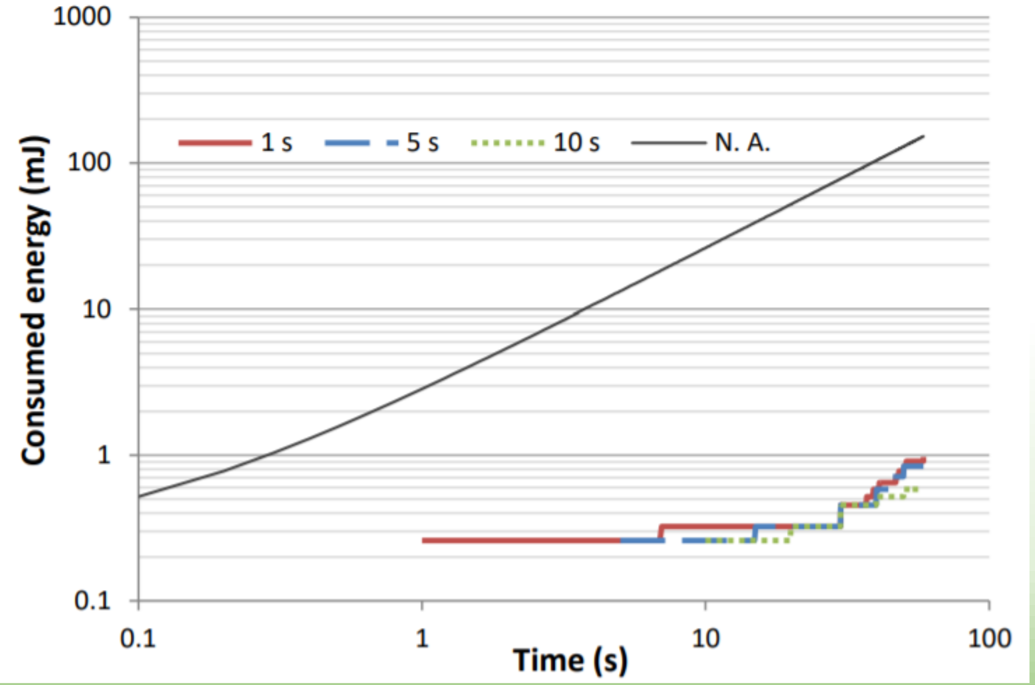
García, L., Parra, L., Romero, O., & Lloret, J. (2019). System for monitoring the wellness state of people in domestic environments employing emoticon-based HCI. The Journal of Supercomputing, 75(4), 1869-1893.

Real cases of application of data management algorithms

First example: Indoor monitoring

Effect on generated traffic and on consume energy:



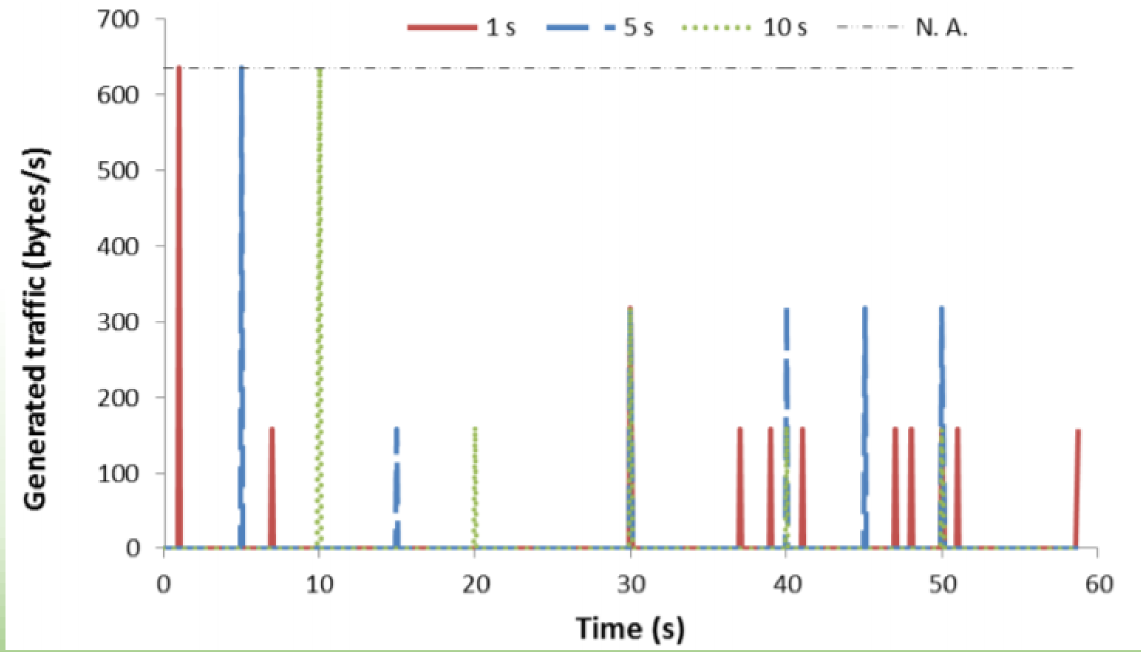


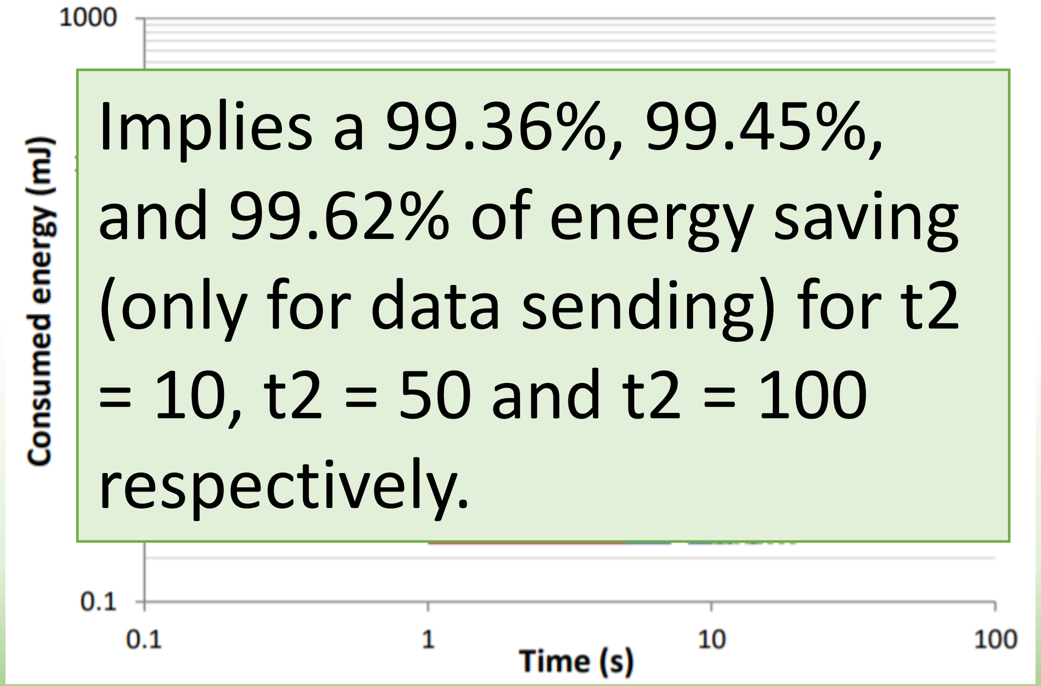
García, L., Parra, L., Romero, O., & Lloret, J. (2019). System for monitoring the wellness state of people in domestic environments employing emoticon-based HCI. The Journal of Supercomputing, 75(4), 1869-1893.

Real cases of application of data management algorithms

First example: Indoor monitoring

Effect on generated traffic and on consume energy:





Implies a 99.36%, 99.45%, and 99.62% of energy saving (only for data sending) for t2 = 10, t2 = 50 and t2 = 100 respectively.

García, L., Parra, L., Romero, O., & Lloret, J. (2019). System for monitoring the wellness state of people in domestic environments employing emoticon-based HCI. The Journal of Supercomputing, 75(4), 1869-1893.

Real cases of application of data management algorithms

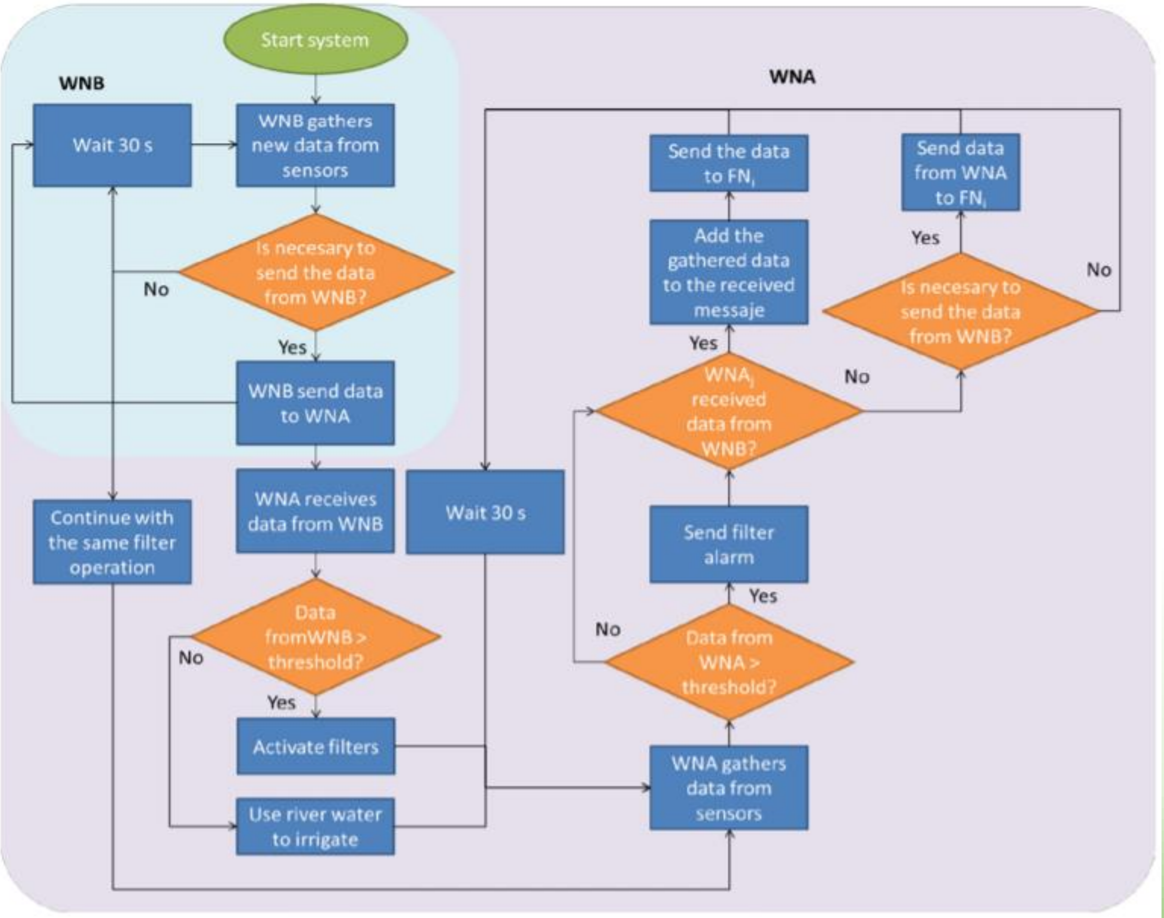
Second example: Outdoor monitoring – multi-hop network

Two areas: WNA and WNB + the crops

Operation algorithm of WNA and WNB

WNA and WNB water quality sensor

Crops soil moisture sensor



Parra-Boronat, L., Rocher-Morant, J., García-García, L., Lloret, J., Tomás Gironés, J., Romero Martínez, J. O., ... & Roig-Merino, B. (2018). Design of a WSN for smart irrigation in citrus plots with fault-tolerance and energy-saving algorithms. Network Protocols and Algorithms, 10(2), 95-115.

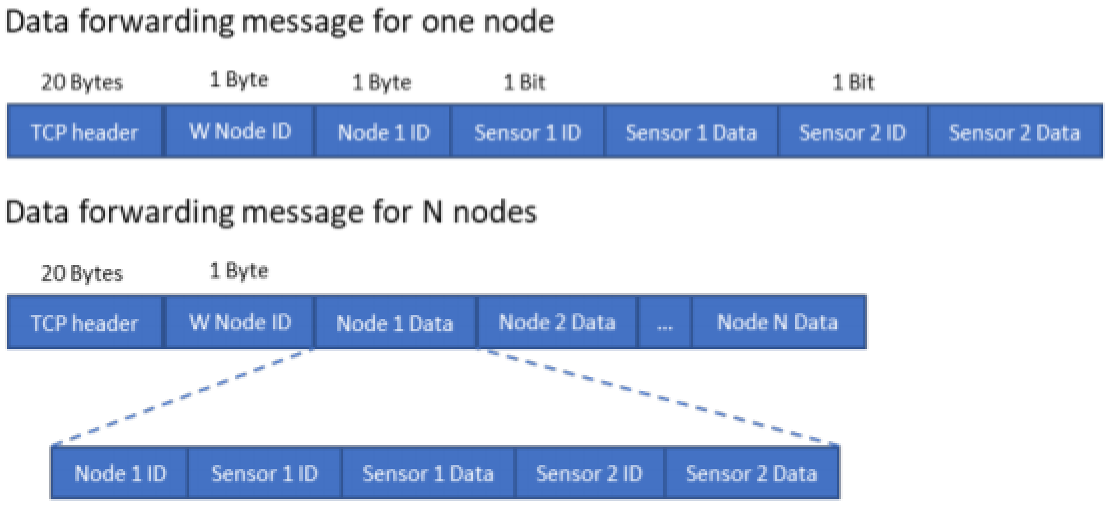
Real cases of application of data management algorithms

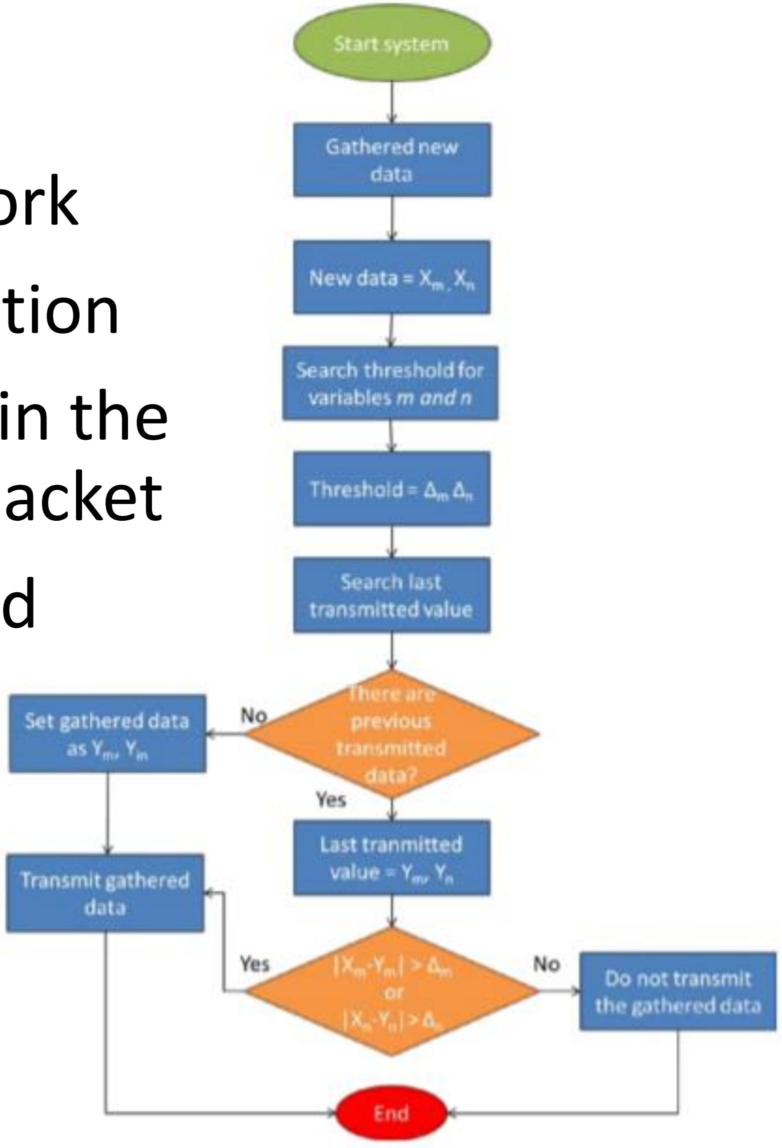
Second example: Outdoor monitoring – multi-hop network

Energy saving algorithm → $Data_t - Data_{stored} > Th$ = Information

New change: If a node forwards info the following node in the multi-hop network adds his information and sends the packet

Data about soil moisture and water quality was simulated



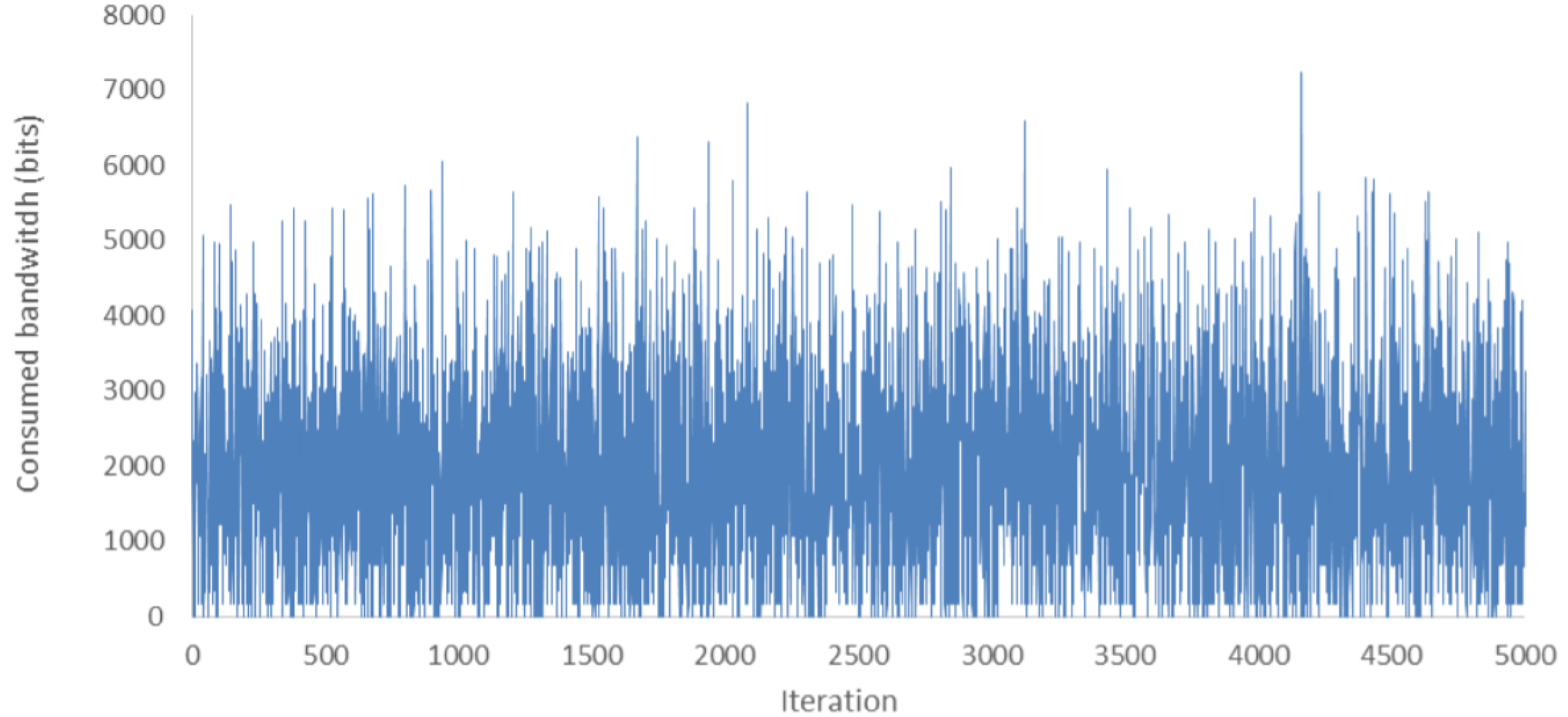


Parra-Boronat, L., Rocher-Morant, J., García-García, L., Lloret, J., Tomás Gironés, J., Romero Martínez, J. O., ... & Roig-Merino, B. (2018). Design of a WSN for smart irrigation in citrus plots with fault-tolerance and energy-saving algorithms. Network Protocols and Algorithms, 10(2), 95-115.

Real cases of application of data management algorithms

Second example: Outdoor monitoring – multi-hop network

Impact on the consumed bandwidth:

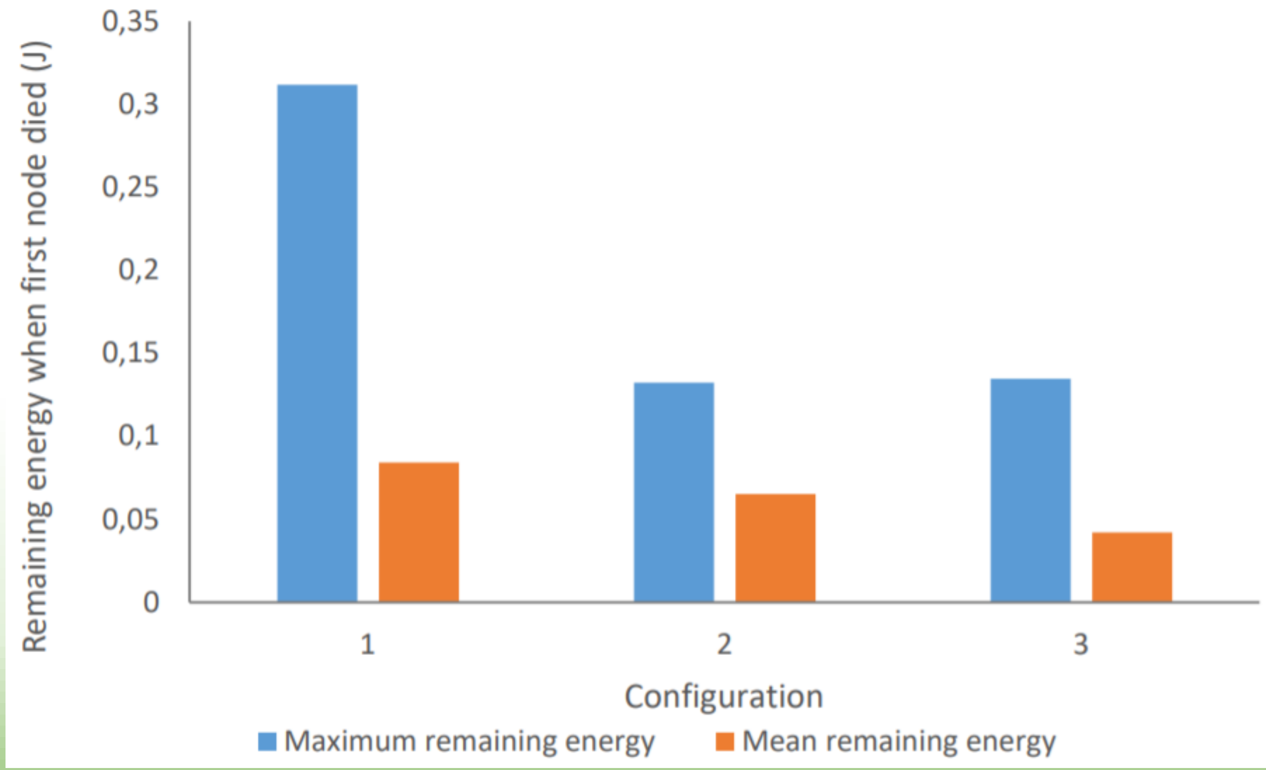


Parra-Boronat, L., Rocher-Morant, J., García-García, L., Lloret, J., Tomás Gironés, J., Romero Martínez, J. O., ... & Roig-Merino, B. (2018). Design of a WSN for smart irrigation in citrus plots with fault-tolerance and energy-saving algorithms. Network Protocols and Algorithms, 10(2), 95-115.

Real cases of application of data management algorithms

Second example: Outdoor monitoring – multi-hop network

Impact on the remaining energy per node with different configurations:

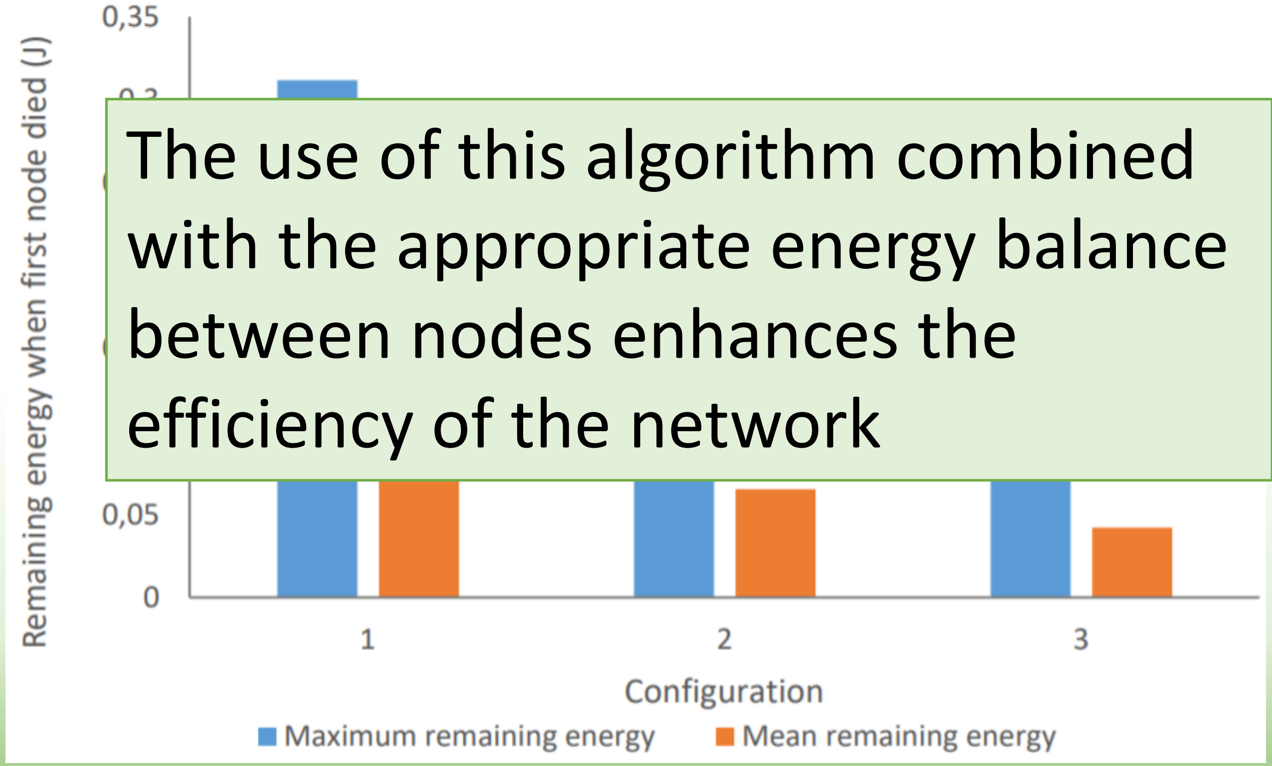


Parra-Boronat, L., Rocher-Morant, J., García-García, L., Lloret, J., Tomás Gironés, J., Romero Martínez, J. O., ... & Roig-Merino, B. (2018). Design of a WSN for smart irrigation in citrus plots with fault-tolerance and energy-saving algorithms. Network Protocols and Algorithms, 10(2), 95-115.

Real cases of application of data management algorithms

Second example: Outdoor monitoring – multi-hop network

Impact on the remaining energy per node with different configurations:



The use of this algorithm combined with the appropriate energy balance between nodes enhances the efficiency of the network

Parra-Boronat, L., Rocher-Morant, J., García-García, L., Lloret, J., Tomás Gironés, J., Romero Martínez, J. O., ... & Roig-Merino, B. (2018). Design of a WSN for smart irrigation in citrus plots with fault-tolerance and energy-saving algorithms. Network Protocols and Algorithms, 10(2), 95-115.

Real cases of application of data management algorithms

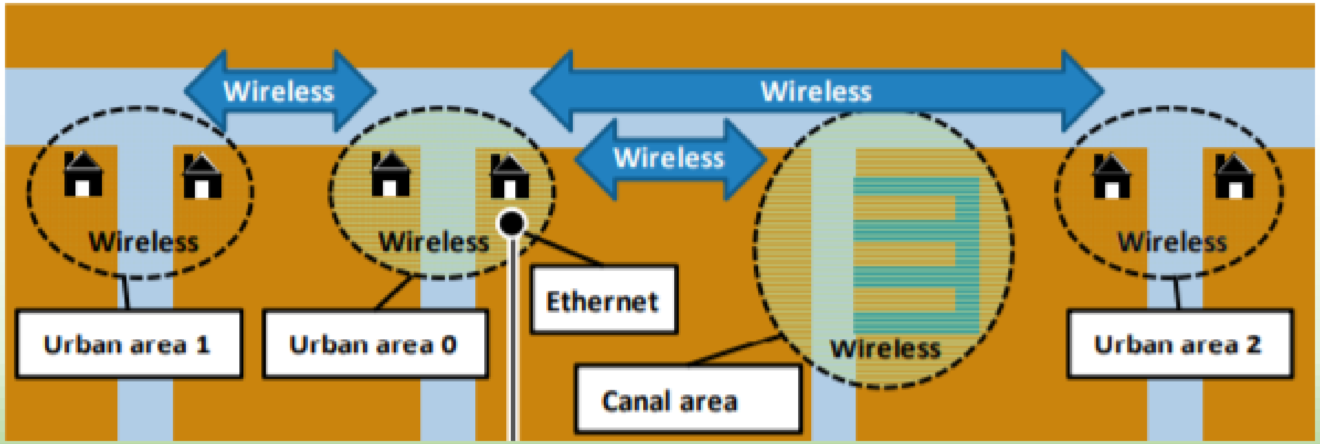
Second example: Outdoor monitoring – multi-hop network + energy harvesting

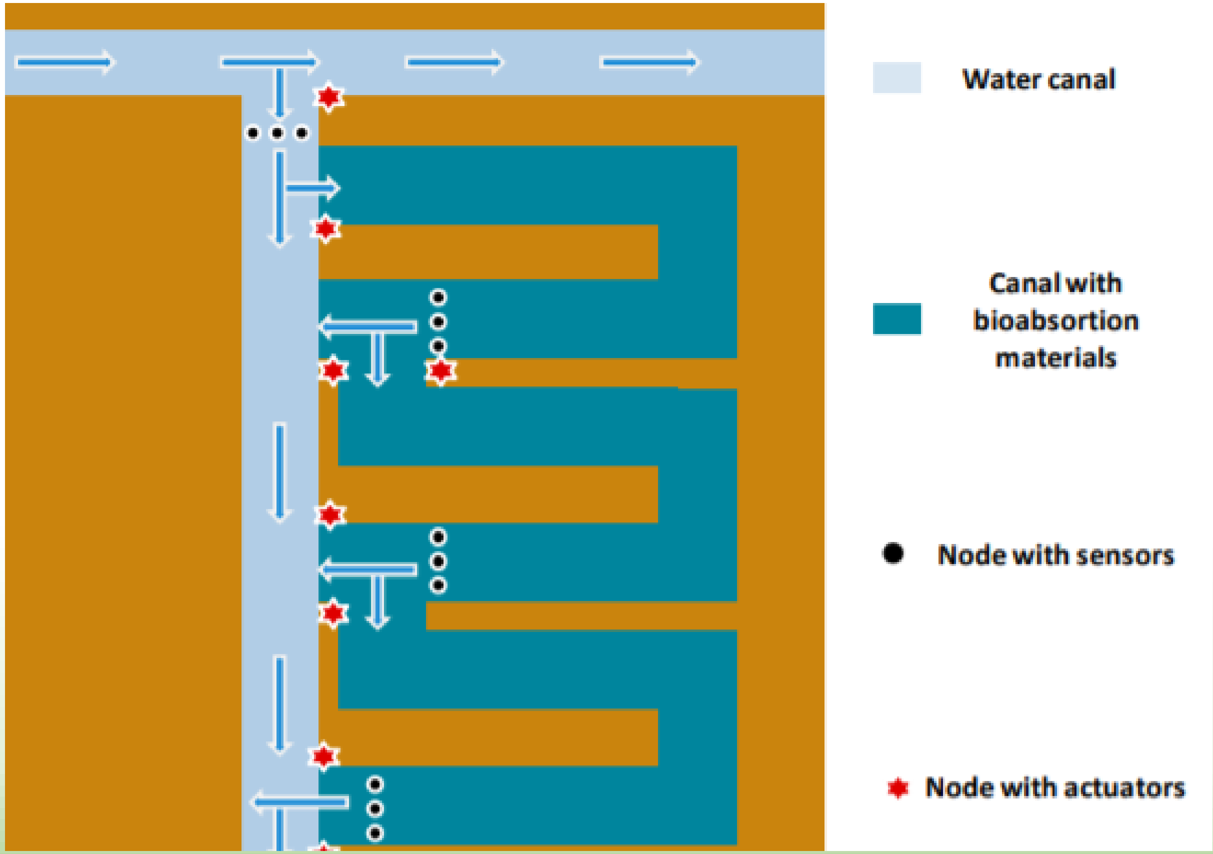
Network that combines LoRa + Wifi

Measure water quality

3 Nodes per measuring point

detect (false positives/negatives)





Jimenez, J. M., Parra, L., García, L., Lloret, J., Mauri, P. V., & Lorenz, P. (2021). New Protocol and Architecture for a Wastewater Treatment System Intended for Irrigation. Applied Sciences, 11(8), 3648.

Real cases of application of data management algorithms

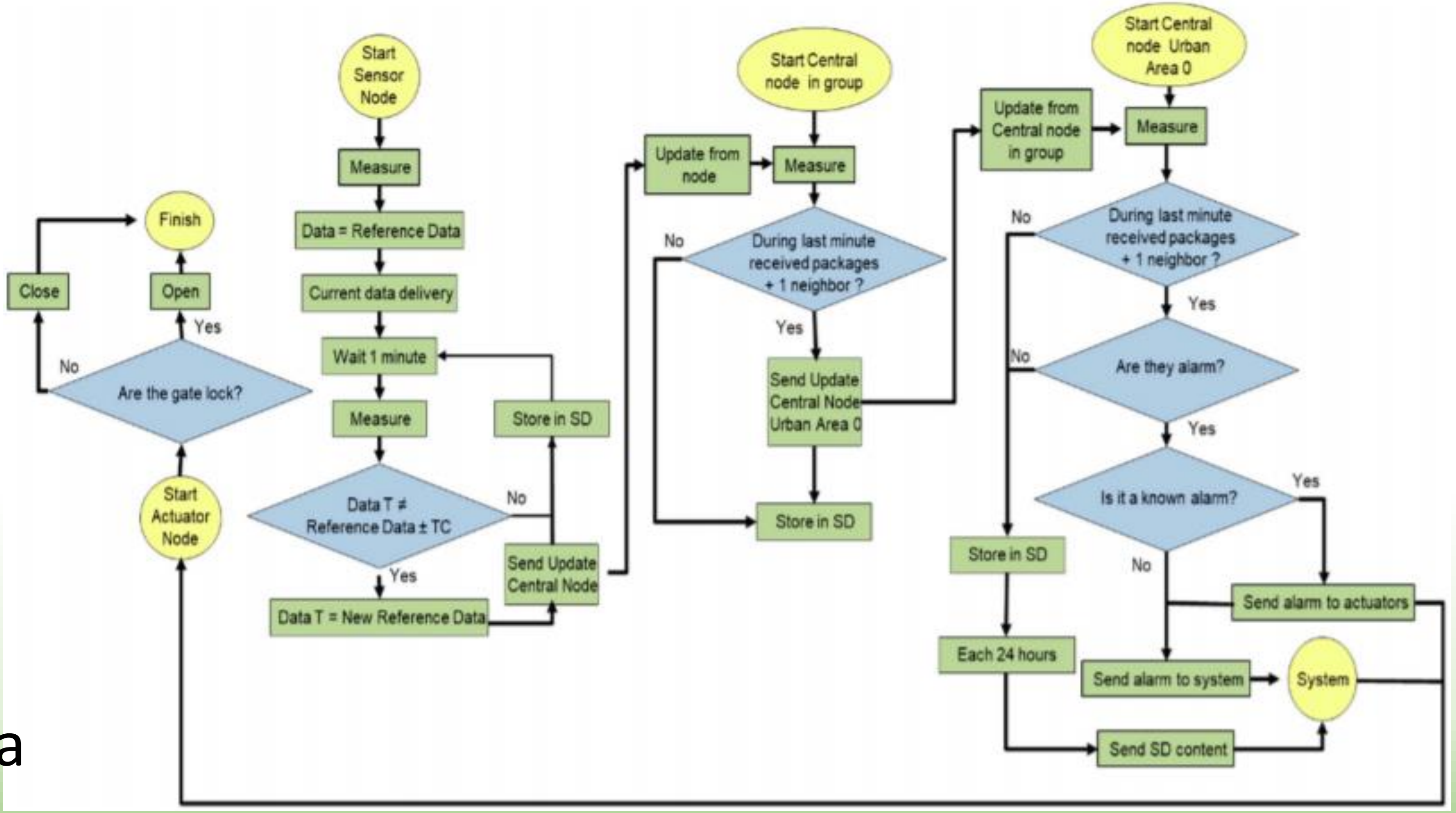
Second example: Outdoor monitoring – multi-hop network + energy harvesting

Algorithm for node groups

1ºEvery time data is gathered it is stored or send.

2ºWhen data should be sent the node sends all the stored data.

3ºAt certain time send all data



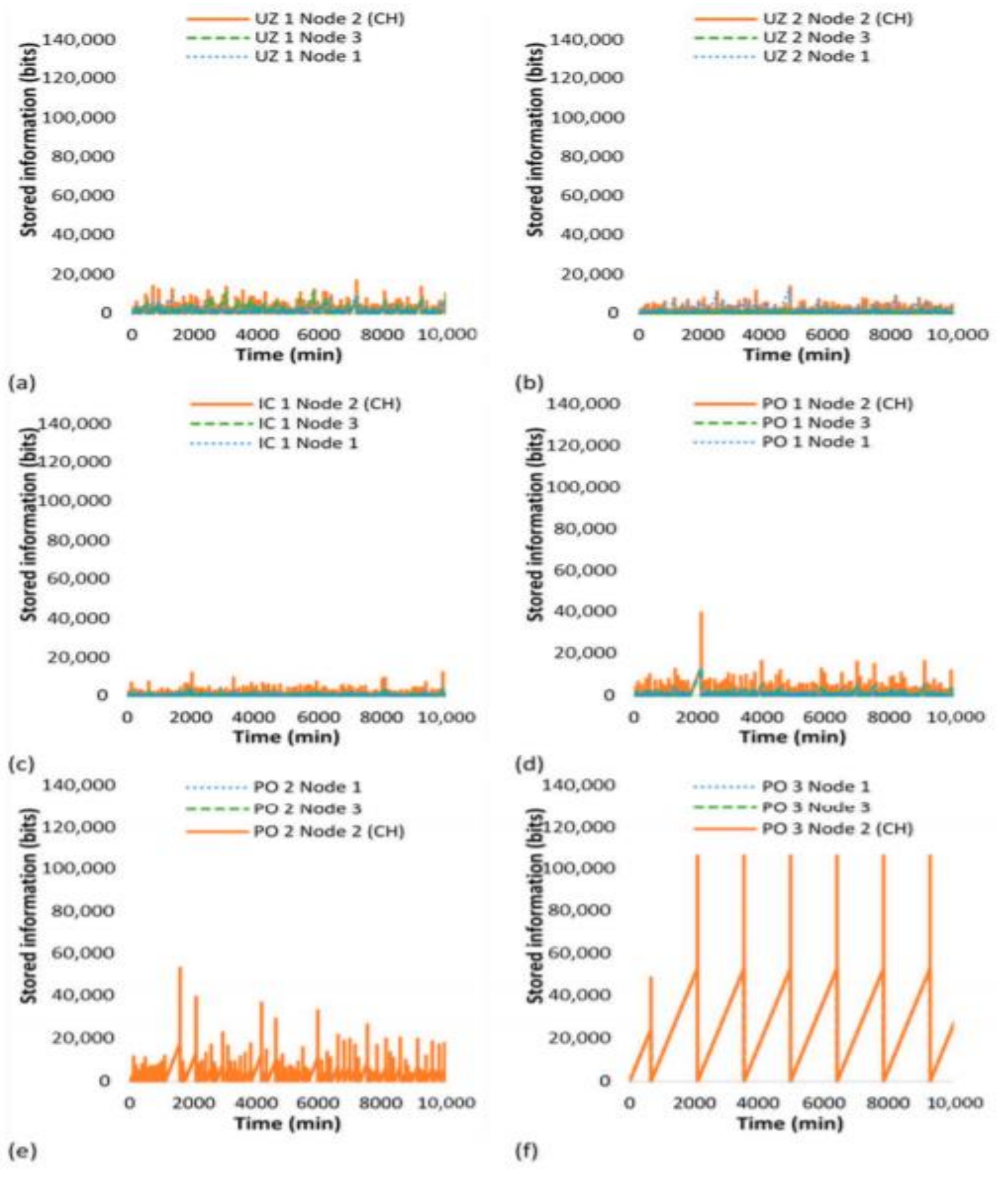
Jimenez, J. M., Parra, L., García, L., Lloret, J., Mauri, P. V., & Lorenz, P. (2021). New Protocol and Architecture for a Wastewater Treatment System Intended for Irrigation. Applied Sciences, 11(8), 3648.

Real cases of application of data management algorithms

Second example: Outdoor monitoring

– multi-hop network + energy harvesting

Stored data in the SDs

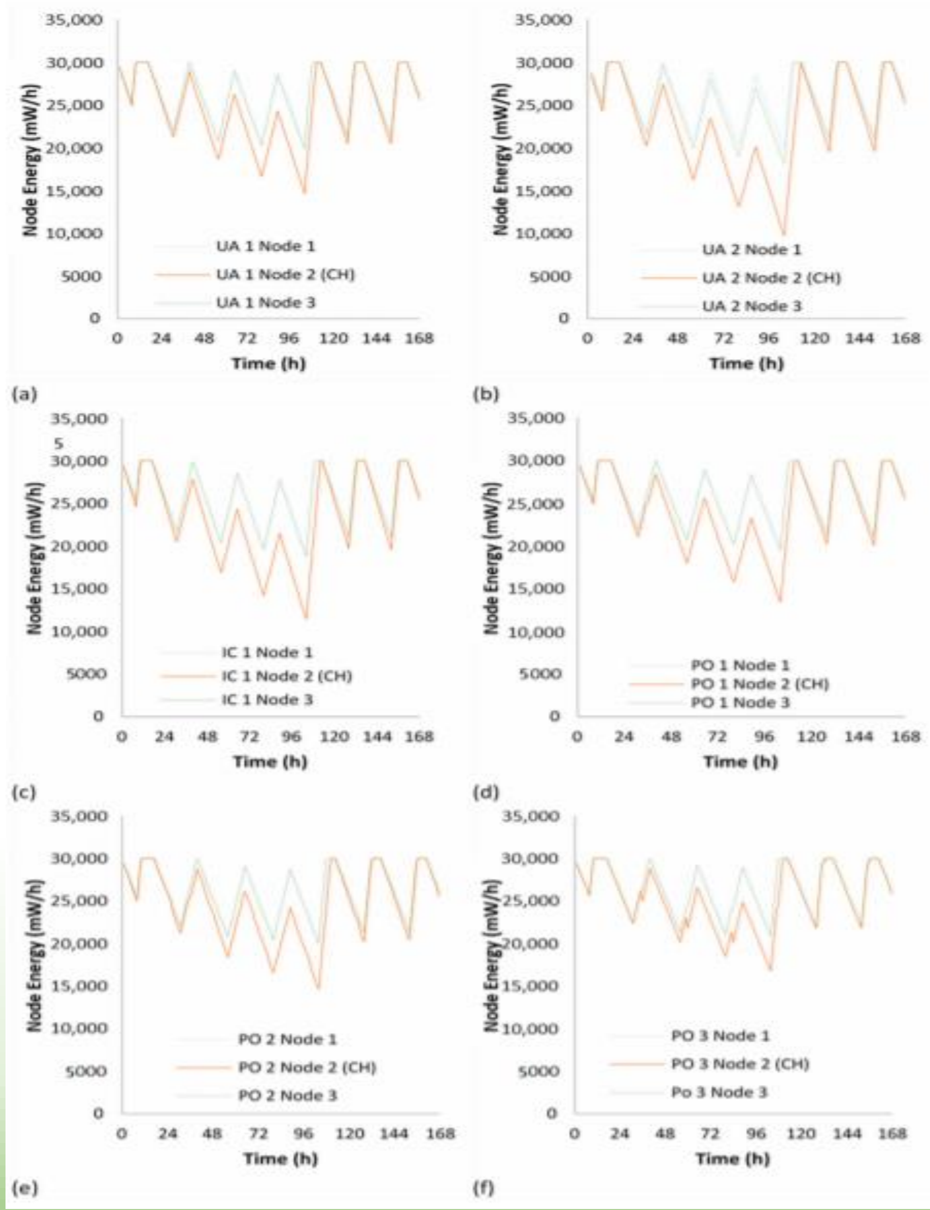


Jimenez, J. M., Parra, L., García, L., Lloret, J., Mauri, P. V., & Lorenz, P. (2021). New Protocol and Architecture for a Wastewater Treatment System Intended for Irrigation. Applied Sciences, 11(8), 3648.

Real cases of application of data management algorithms

Second example: Outdoor monitoring
– multi-hop network + energy harvesting

Remaining energy in the nodes of each areas in the worst scenario for energy harvesting (winter + 2 rainy days)

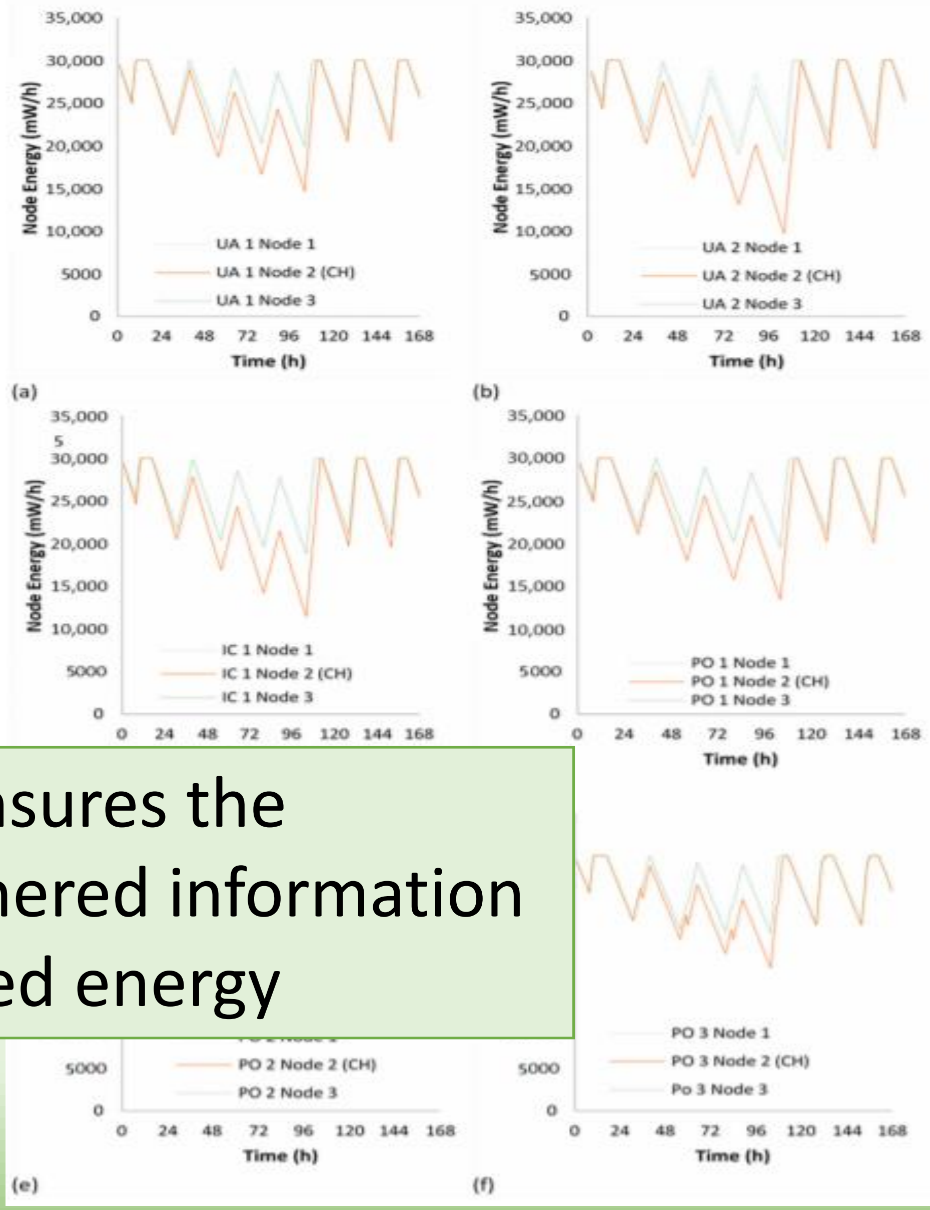


Jimenez, J. M., Parra, L., García, L., Lloret, J., Mauri, P. V., & Lorenz, P. (2021). New Protocol and Architecture for a Wastewater Treatment System Intended for Irrigation. Applied Sciences, 11(8), 3648.

Real cases of application of data management algorithms

Second example: Outdoor monitoring
– multi-hop network + energy harvesting

Remaining energy in the nodes of each areas in the worst scenario for energy harvesting (winter + 2 rainy days)



The use of this algorithm ensures the obtention of the whole gathered information while reducing the consumed energy

Jimenez, J. M., Parra, L., García, L., Lloret, J., Mauri, P. V., & Lorenz, P. (2021). New Protocol and Architecture for a Wastewater Treatment System Intended for Irrigation. Applied Sciences, 11(8), 3648.

Summary:

-We have analyzed the relevance of data to demonstrate that not all the data es equally important, might be redundant and cannot be considered as information.

-We have seen how the excess of data supposes a management problems in terms of network performance, storage requirements, and energy consumption.

-We have proof how the correct management of data locally in the nodes with algorithms can be done even without AI.

-Examples of application of these algorithms in networks for indoor and outdoor monitoring have been explained.

Intended for Irrigation. Applied Sciences, 11(8), 3648.

Future research lines:

-More powerful nodes will come with strongest capacity for data processing and data analysis, in which we can implement statistical tools to have better decision in the nodes.

-Adaptable event-triggered algorithms: Variable measuring periodicity or with variable data resolution after the event.

-Inclusion of similar algorithms in other network elements (not only in the nodes).

-Sending the variation vs. sending the data.

ACKNOWLEDGMENT

The research presented in this keynote was funded by:

Project PDR18-XEROCESPED, funded under the PDR-CM 2014-2020 by the EU (European Agricultural Fund for Rural Development, EAFRD), Spanish Ministry of Agriculture, Fisheries and Food (MAPA) and Comunidad de Madrid regional government through IMIDRA,

iMiDRA
Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario

GOBIERNO DE ESPAÑA
MINISTERIO DE AGRICULTURA, PESCA Y ALIMENTACIÓN

UNIÓN EUROPEA
Fondo Europeo Agrícola de Desarrollo Rural
Europa invierte en las zonas rurales

PROGRAMA DE DESARROLLO RURAL DE LA COMUNIDAD DE MADRID 2014-2020

ACKNOWLEDGMENT

GO TecnoGAR
Tecnificación del garbanzo

The research presented in this keynote was funded by:

"Proyectos de innovación de interés general por grupos operativos de la Asociación Europea para la Innovación en materia de productividad y sostenibilidad agrícolas (AEI-Agri)" in the framework "Programa Nacional de Desarrollo Rural 2014-2020", GO TECNOGAR,

Proyecto financiado por Programa Nacional de Desarrollo Rural (2014 - 2020): FEADER y MAPA

Unión Europea
Fondo Europeo Agrícola de Desarrollo Rural
Europa invierte en las zonas rurales

GOBIERNO DE ESPAÑA

MINISTERIO DE AGRICULTURA, PESCA Y ALIMENTACIÓN

PNDR
Programa Nacional de Desarrollo Rural 2014-2020

ACKNOWLEDGMENT

The research presented in this keynote was funded by:

Conselleria de Educación, Cultura y Deporte with the Subvenciones para la contratación de personal investigador en fase postdoctoral, grant number APOSTD/2019/04.

GENERALITAT VALENCIANA

Conselleria d'Educació, Cultura i Esport

The relevance of data:

Is all the data equally important?

Lorena Parra

Researcher at: Universitat Politecnica de Valencia & Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario,

Spain

loparbo@doctor.upv.es

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

iMiDRA Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario