Efficient Bayesian Communication Approach For Smart Agriculture Applications

Cristanel Razafimandimby, Valeria Loscrì, Anna Maria Vegni, Alessandro Neri

▶ **To cite this version:**

HAL Id: hal-01524731

https://hal.science/hal-01524731

Submitted on 9 Oct 2017

Efficient Bayesian Communication Approach For Smart Agriculture Applications

Cristanel Razafimandimby*, Valeria Loscrí*, Anna Maria Vegni†, and Alessandro Neri†

*Inria Lille - Nord Europe, Lille, France
Emails: {*jean.razafimandimby_anjalalaina, valeria.loscri*}@inria.fr
†Department of Engineering, Roma Tre University
COMLAB Telecommunications Laboratory, Rome, Italy
Emails: {*annamaria.vegni, alessandro.neri*}@uniroma3.it

Abstract—To meet the food demand of the future, farmers are turning to the Internet of Things (IoT) for advanced analytics. In this case, data generated by sensor nodes and collected by farmers on the field provide a wealth of information about soil, seeds, crops, plant diseases, etc. Therefore, the use of high tech farming techniques and IoT technology offer insights on how to optimize and increase yield. However, one major challenge that should be addressed is the huge amount of data generated by the sensing devices, which make the control of sending useless data very important.

To face this challenge, we present a Bayesian Inference Approach (BIA), which allows avoiding the transmission of high spatio-temporal correlated data. In this paper, BIA is based on the PEACH project, which aims to predict frost events in peach orchards by means of dense monitoring using low-power wireless mesh networking technology. Belief Propagation algorithm has been chosen for performing an approximate inference on our model in order to reconstruct the missing sensing data. According to different scenarios, BIA is evaluated based on the data collected from real sensors deployed on the peach orchard. The results show that our proposed approach reduces drastically the number of transmitted data and the energy consumption, while maintaining an acceptable level of data prediction accuracy.

Keywords—Smart Agriculture, Belief Propagation, Markov Random Fields, Bayesian Communication.

I. INTRODUCTION

According to the UN Food and Agriculture Organization, in order to feed the growing population of the Earth, the world will need to produce 70% more food in 2050 than it did in 2006 [1]. Thanks to the benefits of IoT technology, such as advanced data collection and analysis, this food demand can be reached. Technological innovation in agriculture is already known but IoT was set to make agriculture smarter. Smart agriculture, sometimes known as *precision agriculture*, is an integrated information- and production-based farming system that is designed to increase long term, site-specific and whole farm production efficiency, productivity and profitability, while minimizing unintended impacts on wildlife and the environment [14]. Nowadays, most of farmers use smart agriculture, and high-tech farming has rapidly become a standard thanks to agricultural drones and sensors [7].

A concrete example of smart agriculture is the PEACH project [13], whose aim is to dramatically increase the predictability of frost events [1] in peach orchards by means of dense monitoring using low-power wireless mesh networking technology. In three months of operation, the network produced more than four million data just for the temperature data collection [4]. If the acquisition of such a huge amount of data is essential for avoiding frost damage on time, the data transmission to the network may affect the energy consumption of sensing devices, and can also cause network congestion issues. A strong reduction of massive amount of data generated by these sensing devices is therefore necessary.

For this aim, and hence to increase the network lifetime, we proposed in [8] an efficient Bayesian Inference Approach (BIA) in the IoT context. However, the results and discussions were restricted to indoor environments. In this paper, we go a step further by applying BIA to outdoor environments. The goal is to demonstrate that our approach is efficient also in outdoor environments, and then we will apply BIA to the PEACH project mentioned above. Experimental results will assess the benefits of BIA, mainly expressed in terms of reduction of transmitted data, and energy consumption.

It should be recalled that BIA approach allows to remove a great amount of spatio-temporal correlation data and is based on Pearl's Belief Propagation (BP) algorithm. BP is an iterative technique mostly used for solving inference problems [15]. In the IoT context, the belief of a device is related to the physical quantity measured by the device sensors. BP infers the measurements of other neighboring nodes, especially in cases where the data are missing. In BP-based approaches, each node determines its belief by merging its local measurement with the beliefs of its neighboring nodes, and its beliefs obtained in the past run. BP provides therefore a spatio-temporal cooperation among several devices. A good correlation between data is important in such inference problems since it dictates the accuracy of data inference, and hence reduces the estimation error of the global information. The estimation error is then compared to a predefined threshold, in order to decide if and how much data information to transmit back.

The main contributions of this paper can be summarized as follows:

- The adoption of BIA approach that allows avoiding the transmission of useless data in smart

---

[1] A frost event occurs when ice forms inside the plant tissue and injures the plant cells plant.

agriculture networks. A BP algorithm has been chosen to infer the missing data;

- The use of efficient and smart communication in order to decrease the estimation error in the remote server. A dual prediction scheme will be proposed for this aim;

- Performance assessment based on data collected from sensors installed around the peach orchard.

This paper is organized as follows. Section II briefly describes our reference PEACH network. Section III provides some related works on data propagation mechanisms in the IoT context. In Section IV, we present the BIA approach based on BP algorithm for data sharing in an outdoor scenario related to the PEACH network. Section V provides the experimental results for the assessment of the proposed BIA technique in different real scenarios. Finally, conclusions are drawn at the end of the paper.

II. PEACH Network Architecture

In this paper we take into account the network model proposed in the PEACH project [13]. The PEACH network consists of 21 low-power wireless motes deployed in a 110×50 m area (peach orchard) in Junin, in the West of Argentina (see Figure 1). A hierarchy architecture of the mote devices is used for different connectivity ranges (*i.e.*, long, medium and low range motes). The temperature and humidity sensors built into the motes publish their value every 30 s. Then, all the measurements are sent to the network gateway, which is composed of a Raspberry Pi single-board computer and a DC2274 SmartMesh IP manager. These two components are connected to each other via an USB port. The gateway is connected to the Internet thanks to the solmanager application, which runs on the Raspberry Pi. All the information received by the SmartMesh IP Manager are then transmitted by the solmanager to a remote server located in the Inria-Paris research center. The server in Paris in turn stores the data into a database by means of the solserver application.

What we described above is the basic model proposed in the PEACH project. However, in order to reduce the huge amount of data generated by the sensor devices, we need to modify this network model. More in detail, we have to implement our BIA approach on the remote server and on the gateway. Figure 2 illustrates the modified version of the network model, where the gateway (*i.e.*, SmartMeshIP Manager) has become "smarter" and able to send only the useful data. Finally, by using the inference algorithm, which will be detailed in the next section, the server application (*i.e.*, solserver) will be able to reconstitute the missing data.

III. Related work

In order to clarify the difference about our approach with those existing in the literature, we classify different related works according to the fields where they were treated *i.e.*, (*i*) in Wireless Sensor Networks (WSNs), and (*ii*) IoT domain.

Energy saving is considered as the most important part in WSNs, since sensors have to operate on extremely limited energy budget. A good survey of energy saving in WSNs context was proposed in [2]. Data reduction is
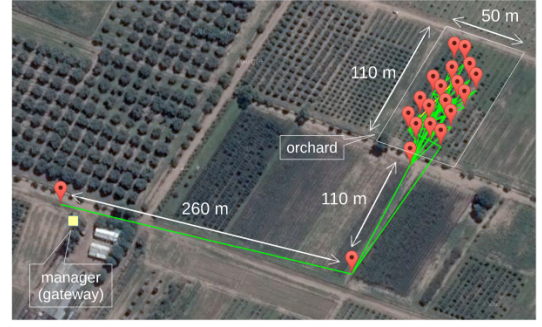

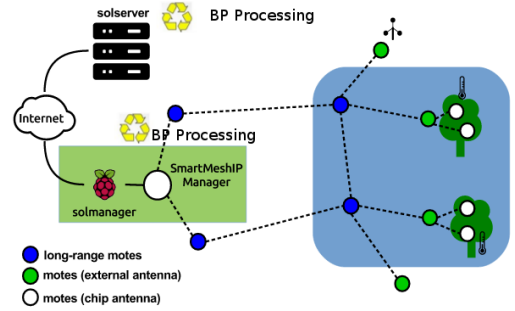Figure 1: The PEACH network [4].

Figure 2: Modified version of the PEACH network architecture.

one of the best approaches to achieve energy-efficiency in data acquisition. Its main goal is the reduction of the data to be transmitted over the network, such that the energy spent for communications decreases. Many data prediction approaches have been proposed in several works [6], [9], [10].

Even if many approaches have been developed for data prediction in the WSNs context, all the proposed approaches deal only with homogeneous data. Therefore, they neglect the huge amount of heterogeneous data typical of IoT scenarios. However, the heterogeneity of IoT devices requires that the employed prediction model has to be applicable also to a variety of applications with heterogeneous data. An example of concrete application is the smart agriculture, where the various agricultural inputs are collected with different sensor devices that can collect a variety of data correlated to each others. The existing methods in WSNs are therefore unsuitable for the IoT context and for smart agriculture.

One of the most distinguishing aspects of the IoT is the fact that the data are acquired from a variety of sources. As far as we know, there exist only three works that take in account this heterogeneous aspect of IoT [3], [8] and [11]. In [3], the authors presented a cloud-based Adaptive Sensing Belief Propagation protocol with energy-efficient data fusion for IoT applications. Our work differs from [3] in many aspects. Firstly, we assume that the correlations between different kind of data (*e.g.*, temperature, humidity, light and voltage) are known a priori, and consequently, we do not need to compute the correlation matrix each time. Secondly, the knowledge of the correlation is used only to decide which data have to be sent or inferred in the remote server. Therefore, in this

work we do not operate the sensor selection optimization. Finally, we propose a smart gateway concept in order to decrease the inference error in the remote server. Hence, we provide a dual prediction scheme *i.e.*, one in the gateway, and the other in the remote server, based on the BP algorithm.

Finally, the authors of [11] presented a theoretical bayesian based approach for data sharing in the IoT context, while in [8] we go a step further by proposing an extensive experimentation scenario for indoor environments. In this paper, we will present additional experiments and discussions in a realistic outdoor environment.

IV. BAYESIAN INFERENCE APPROACH

In this section, we describe our BIA technique. As mentioned before, our main goal is to avoid sending useless data, while keeping an acceptable level of data content accuracy. For this aim, BIA is based on Pearl's BP algorithm that will be described below.

As a starting point before any inference procedure, the design of a graphical model should be provided. Graphical models are schematic representations of probability distributions. They consist of nodes connected by either directed or undirected edges. Each node represents a random variable, and the edges represent probabilistic relationships among variables. Models which are comprised of directed edges are known as *Bayesian networks*, whilst models that are composed of undirected edges are known as *Markov Random Fields* (MRF) [12]. In this paper, we present an inference approach under the hypothesis of MRF, modeled by means of Factor Graphs. It follows that our goal is to estimate the state X of the sensed environment starting from the sets of data collected by each sensor node. Based on the remarkable Hammersley-Clifford theorem, the joint distribution $P_X(x)$ of an MRF model is given by the product of all the potential functions *i.e.*,

$$P_X(x) = \frac{1}{Z} \prod_i \psi_i(x_i) \prod_{i,j \in E} \psi_{ij}(x_i, x_j), \quad (1)$$

where Z is the normalization factor, $\psi_i(x_i)$ is the evidence function, E is the set of edges encoding the statistical dependencies between two nodes i and j, and $\psi_{ij}(\cdot)$ represents the potential function. Note that the graphical model parameters (*i.e.*, ψ_i and ψ_{ij}) can be estimated from the observed data by using a learning algorithm like in [5].

Figure 3 illustrates an example of the MRF model. The filled-in circles represent the observation nodes (*i.e.*, N_{ε_i}) and the empty circles represent the hidden nodes (*i.e.*, x_i). The potential functions are associated with the links between x_i whilst the evidence functions are associated with the links between N_{ε_i} and x_i. For simplicity, in this paper, we consider widely used pairwise MRF, *i.e.*, MRF with the maximum clique [2] of two nodes.

One of the main goals when dealing with graphical models is the marginal distribution computation, as shown in Eq. (2). They are used to predict the most probable assignment for a variable node. For notation convenience, let us assume that X and Y are two distinct multivariate random variables with assignments $x \in \mathcal{X}^m$ and $y \in \mathcal{Y}^n$.
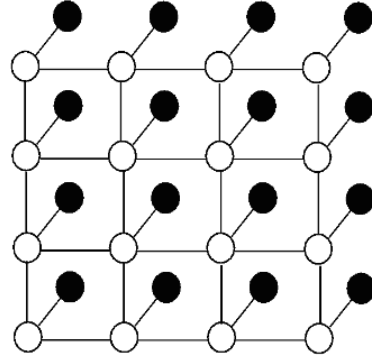

Figure 3: An example of MRF model.

The nodes in Y are called hidden nodes and those in X are the observed ones. So, given the i-th device in our network, x_i will be the observation of the phenomenon we intend to share (*e.g.*, temperature) and y_i will be associate to the phenomenon we want to infer, (*e.g.*, humidity)

$$p(y_v|x) = \sum_{y_1} \sum_{y_2} ... \sum_{y_n} p(y_1, y_2, y_3, ..., y_n|x). \quad (2)$$

Obviously, using (2), the complexity of a complete enumeration of all possible assignments to the whole graph is $O(|\mathcal{Y}|^{n-1})$, which is intractable for most choices of n. Therefore, we need a faster algorithm like Belief Propagation [3] for computing the marginal probability. BP is a well known algorithm for performing inference on graphical models [15].

Let $p(y_i)$ represents the marginal distribution of i-th node, and BP allows the computation of $p(y_i)$ at each node i by means of a message passing algorithm. The message from the i-th to the j-th node related to the local information y_i is defined as:

$$m_{ji}(y_i) \propto \int \psi_{ji}(y_j, y_i)\psi_j(y_j) \prod_{u \in \Gamma(j), u \neq i} m_{uj}(y_j) dy_j,$$

$$(3)$$

where $\Gamma(j)$ denotes the neighbors of node j and the incoming messages from previous iteration are represented by m_{uj}. Notice that (3) will be performed between all nodes in the model until the convergence or if a maximum number of iterations I_{max} will be reached. Thus, the prediction *i.e.*, the belief at the i-th node, is computed through all the incoming messages from the neighboring nodes and the local belief, *i.e.*:

$$\hat{y_i} = belief(y_i) = k \cdot \psi_i(y_i) \prod_{u \in \Gamma(i)} m_{ui}(y_i), \quad (4)$$

where k is a normalization constant. Finally, it is worth to mentioning that the BP is able to compute the exact marginalization in the case of tree-structured graphical models.

V. EXPERIMENTAL RESULTS

In this section we provide the experimental results of our BIA approach. Real data (*i.e.*, temperature and humidity) collected from sensors deployed in the peach orchard have been used. Each data collection has been performed every 30 seconds. In this paper, we focus on

---

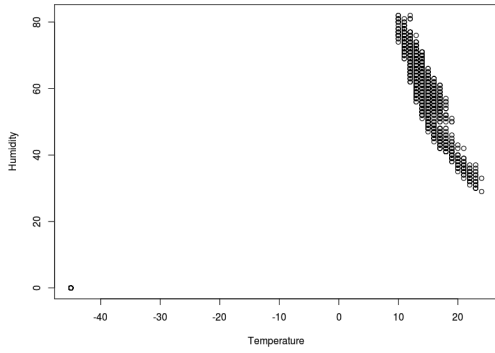[2] A clique is defined as a fully connected subset of nodes in the graph.

[3] Only take linear time.

Figure 4: Relationship between humidity and temperature data.

Scenario	#Transmitted data [Byte]	EC [kJ]	MSE	ER
s_1	8408	1634.5152	-	-
s_2	4204	817.2576	0.62	0.295
s_3	4260	828.144	0.022	0.0066

Table I: Results obtained during eighteen hours of readings for different scenarios.

the data collected between 16 and 17 October 2016. The collection started at 10pm on October 16th and ended at 4pm on October 17th (i.e. eighteen hours). Figure 4 illustrates the relationship between data during the sensors reading. We can notice that there is a good correlation between temperature and humidity data, so that we can easily infer the humidity from temperature data, and vice versa.

In this paper, we infer humidity from temperature. The temperature is in degrees Celsius, whilst the humidity is a value ranging from 0 to 100%. We assess our approach w.r.t. (*i*) the number of transmitted data, (*ii*) the energy consumption, (*iii*) the average value of the distortion level, and (*iv*) the average value of the estimation error.

The number of transmitted data represents the total number of data transmission performed by all the sensors during the readings. The inference error is an important metric for any inference procedure. The goal is to have an errorless inference approach, *i.e.* an approach that is able to estimate the true value of data during all the inference procedures. However, this is almost never the case but we want that this error is as low as possible. In addition to the inference error, computing the distortion level is also important. This allows to determine the difference between the real and the estimated value, and can be expressed using the Mean Squared Error (MSE) metric, defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \qquad (5)$$

where $\hat{y}_i$ and y_i are respectively the predicted and true value during the n-th reading.

All of our assessments are based on three different scenarios (*i.e.*, s_1, s_2, and s_3), based on the inference level provided. In scenario s_1, the gateway sends to the remote server all the temperature and humidity data it receives. This means that the server application does not perform

any inference (*i.e.*, no inference). In the second scenario s_2, the gateway sends only the temperature data to the remote server, and the cloud in turn infers the corresponding humidity data by using the BP algorithm (*i.e.*, inference level via BP). Finally, in the scenario s_3, we consider that the gateways are "smart" devices, meaning that before sending their data to the remote server, they first compute the probability $\Pr(e|T, h)$ of making an inference error e on the remote server given the temperature data T, and the humidity data h. In this scenario, the inference level is based on a probabilistic approach. If there is a strong chance that the error magnitude *i.e.*, $|e|$, exceeds a predefined threshold *i.e.*, $|e|_{Max}$, the gateway sends both humidity and temperature data to the remote server, else the gateway sends only the temperature data, and the humidity value will be inferred in the remote server using the BP algorithm. This can be expressed mathematically as the inference error probability higher than a maximum allowed value $|e|_{Max}$, and conditioned to the temperature and humidity measurements *i.e.*, T and h, is lower or at least equal to a given threshold P_e^{Max}, that is:

$$\Pr\{|e| > |e|_{Max}|T, h\} \leqslant P_e^{Max}, \qquad (6)$$

where the computation of $\Pr(e|T, h)$ is done by means of the BP algorithm. It should be noted that this computation requires the knowledge of the a priori probability of inference error *i.e.*, $\Pr(e)$. Also, the value of the threshold $|e|_{Max}$ strictly depends on the application context. In our case, we set this value equal to 1.5. A similar consideration can be applied to the probability threshold P_e^{Max}, which has been set to 0.5.

As you may have noticed, with the proposed scenarios above, removing the redundancy data at the gateway has not yet a positive effect on the sensing nodes since the data filter has been made only on the gateway. However, it is interesting to study the possibility of doing the raw data filtering in the sensing nodes. By doing that, not only the large computation and the single point of failure at the gateways will be avoided but also the energy consumption of sensing nodes will probably decreased. The energy costs reported in Table I is therefore the estimated energy costs assuming that data filtering will be done on the sensing nodes and the used model is exactly the same as on the gateways. In our energy consumption evaluations, we assume that the power consumption for sending each temperature and humidity value is 3mW. This cost has been obtained on the SHT-31 used mote.

A. Obtained results

As mentioned before, we evaluate the performance of our approach in terms of number of transmitted data, average value of the estimation error (*i.e.*, ER), average value of the distortion level (*i.e.*, MSE), and energy consumption (*i.e.*, EC). Our approach has been implemented in C++, and the assessments have been performed with respect to the ground truth collected on the peach orchard.

Table I illustrates the obtained results during eighteen hours of readings, for different simulated scenarios. We can notice that our Bayesian inference approach drastically reduces the number of transmitted data and the energy consumption, while maintaining an acceptable level of prediction accuracy and information quality. We can notice also that we decrease considerably the estimation error by using the scenario s_3. Indeed, the gateways are
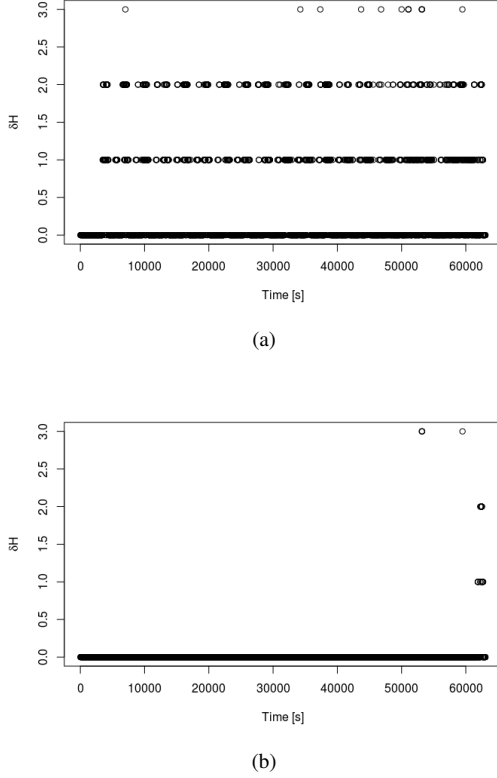
(a)

(b)

Figure 5: Variation of δH in scenario (a) s_2, and (b) s_3 versus eighteen hours collection time.

smarter in this case *i.e.*, by computing the a posteriori probability of the inference error, the gateways will be able to estimate the right moment and the data type to send in the remote server. However, this increases the number of transmitted data (and hence the energy consumption), as compared to scenario s_2. This is due to the fact that in s_2, the gateways send only the temperature data without worrying of the risk of inference error in the remote server.

Figure 5 shows the variation of δH during eighteen hours of reading using s_2 and s_3 [4], where δH is the difference between the true value and the inferred one of humidity data *i.e.*, $\delta H = \hat{y}_i - y_i$. This metric illustrates therefore the inference error of our BIA approach during all the readings. No inference error occurs for $\delta H = 0$, *i.e.*, when $\hat{y}_i = y_i$. In s_2, for the majority of time we notice no inference error *i.e.*, the probability of having a null inference error is $\Pr(\delta H = 0) = 70.44\%$, while we have $\Pr(\delta H = 1) = 19.21\%$, $\Pr(\delta H = 2) = 10\%$, and $\Pr(\delta H = 3) = 0.3\%$. Best performances are for scenario s_3, where we observe an error of 1.0 for the 99.31% of time, $\Pr(\delta H = 1) = 0.31\%$, $\Pr(\delta H = 2) = 0.26\%$, and $\Pr(\delta H = 3) = 0.095\%$.

VI. Conclusions

In this paper, we presented an inference-based approach –namely, BIA– applied to the PEACH network, with the aim of avoiding useless data transmission. The strong correlation between temperature and humidity data was taken into account for this study.

---

[4] Of course, in Figure 5 we did not consider the scenario s_1 since it does not use the proposed inference approach.

Through extensive simulations and by using the real data collected from sensors deployed in a peach orchard, we have showed that our BIA approach reduces considerably the number of transmitted data and the energy consumption, while keeping an acceptable level of estimation error and information quality. We have also shown that the use of smart gateway decreases significantly the inference error. Future works will explore the possibility of doing the prediction directly on the sensor nodes.

References

[1] Nikos Alexandratos, Jelle Bruinsma, et al. World agriculture towards 2030/2050: the 2012 revision. Technical report, ESA Working paper Rome, FAO, 2012.

[2] Giuseppe Anastasi, Marco Conti, Mario Di Francesco, and Andrea Passarella. Energy conservation in wireless sensor networks: A survey. *Ad hoc networks*, 7(3):537–568, 2009.

[3] Farshid Hassani Bijarbooneh, Wei Du, Edith C-H Ngai, Xiaoming Fu, and Jiangchuan Liu. Cloud-assisted data fusion and sensor selection for internet of things. *IEEE Internet of Things Journal*, 3.

[4] Keoma Brun-Laguna, Ana Laura Diedrichs, Diego Dujovne, Rémy Leone, Xavier Vilajosana, and Thomas Watteyne. (Not so) Intuitive Results from a Smart Agriculture Low-Power Wireless Mesh Deployment. In *CHANTS'16*, New York City, United States, September 2016.

[5] Zoubin Ghahramani. Graphical models: parameter learning. *Handbook of brain theory and neural networks*, 2:486–490, 2002.

[6] Caner Komurlu and Mustafa Bilgic. Active inference and dynamic gaussian bayesian networks for battery optimization in wireless sensor networks. In *Proceedings of AAAI workshop on artificial intelligence for smart grids and smart buildings*, 2016.

[7] Andrew Meola. Why iot, big data & smart farming are the future of agriculture. http://www.businessinsider.com/internet-of-things-smart-agriculture-2016-10?r=UK&IR=T. Accessed March 2, 2017.

[8] Cristanel Razafimandimby, Valeria Loscri, Anna Maria Vegni, and Alessandro Neri. A bayesian and smart gateway based communication for noisy iot scenario. In *International Conference on Computing, Networking and Communications*, 2017.

[9] Surender Kumar Soni, Narottam Chand, and Dhirendra Pratap Singh. Reducing the data transmission in wsns using time series prediction model. In *Signal Processing, Computing and Control (ISPCC), 2012 IEEE International Conference on*, pages 1–5. IEEE, 2012.

[10] Liansheng Tan and Mou Wu. Data reduction in wireless sensor networks: A hierarchical lms prediction approach. *IEEE Sensors Journal*, 16(6):1708–1715, 2016.

[11] Anna Maria Vegni, Valeria Loscri, Alessandro Neri, and Marco Leo. A bayesian packet sharing approach for noisy iot scenarios. In *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 305–308. IEEE, 2016.

[12] Chaohui Wang, Nikos Komodakis, and Nikos Paragios. Markov random field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding*, 117(11):1610–1627, 2013.

[13] Thomas Watteyne, Ana Laura Diedrichs, Keoma Brun-Laguna, Javier Emilio Chaar, Diego Dujovne, Juan Carlos Taffernaberry, and Gustavo Mercado. Peach: Predicting frost events in peach orchards using iot technology. *EAI Endorsed Transactions on the Internet of Things*, 2016.

[14] Brett Whelan and James Taylor. *Precision agriculture for grain production systems*. Csiro publishing, 2013.

[15] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.